SURVEY OF THE THEORY OF THE FIRM

by

Fernand Martin, M. A.

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Department of Economics
McGill University,
Montreal.

November 1961

PREFACE

The contributions of this thesis may be summarized as follows: By introducing the new theory of the firm its purpose is to provide a new perspective to the study of the firm. The survey is, then, the instrument to provide this perspective. We hope that after reading this thesis, some readers, previously unfamiliar with the subject, will be able to classify and relate the various components of the theory of the firm. Furthermore, very often we had to make contributions on minor points. For example we show that Cournot has been wrongly represented as "following an unbroken line, or, reasoning...without varying anything except the number of competitors" (Schumpeter, 185, p. 981). On the same topic, in our appendix D we modified the mathematical formula of Fellner. In chapter IV we make opportunity costs an integral part of the analysis of the firm. In chapter V on the Laws of Returns we show that many textbooks and other people have false conceptions about the shape of the product curves. We also give a formal proof to the Edgeworth's statement and show that it is true only in a particular case. In chapter VI it is shown that Marshall is not a traditionalist,

but on the contrary is the originator of a different theory of the firm. In chapter VII we try to prove that the analysis of Mr. Archibald is fallacious and that there is a difference between Chamberlin and Cournot. The contribution of the last chapter is to provide a summary of the vast literature on the contemporary theory of the firm; there is then an attempt to integrate new managerial tools such as linear programming, game theory, location theory, marketing theory, and other theories into the formal economic model of the firm.

When our views differed from the received knowledge, we tried to introduce them as unconspicuously as possible. In the historical sections our approach was to read the authors (those that were available either in English or in French), list our remarks and go back to the various competent historical works such as those of Schumpeter, Stigler, and Hutchison, and try to credit them with any statement they already made on the topic; this was usually done by using their own words. Yet we apologize in advance for any of their remarks that might have been overlooked through ignorance or oversight on the part of the present author. In many places they were used even as prime sources.

A more spectacular approach would have been to 'advertize' our contributions or present them as controversies, but the unhappy experiences of others who proceeded along these lines made us change our attitude. To link our minor

contributions we had to provide the reader with connecting material which has only an expository value and we of course make no claim for this. All this makes for a long thesis whose length may make us overshoot our original goal (of building bridges between authors instead of digging rivers between them) to the point where the content of the thesis is so diluted that it might leave the reader untouched. This is the risk we choose to run.

Similarly, not knowing the type of audience, we did not know what to include and what could be excluded on the count that 'everybody knows that'. That is why to some people some portions will look overworked while to others they will look too sketchy or evasive. If we would have been sure that our audience would be equally familiar with spatial analysis, homogeneous functions, linear programming and the work of Cournot, the thesis probably could have been brought down to one or two hundred pages.

On the other hand, attempting to survey the whole field of the theory of the firm in five hundred pages has forcefully to be done in a rather superficial way and it is only accompanied with the notation sub spe futuri studii (as Viner put it) that we present it.

In writing this thesis we were helped in many ways. We would like to thank the Canada Council for providing us with a fellowship to study at the London School of Economics. Chapter four benefited from the stimulating discussions that

were going on in the Seminar conducted by Professors Robbins and Philipps. In addition, the general attitude of the people who were at the London School at that time (1958-1959) being generally to minimize the importance of Chamberlinan analysis, forced us to rework our 'positions' in this matter. This produced chapters three and seven. We would like also to thank Dr. Carman Miller of the Department of Mathematics of the University of Saskatchewan, who helped us by writing appendix E and providing advice at certain points on spatial models. Yet since he was often kept in the dark about the economic implications, he cannot be held responsible for any misuse of his tools.

We finally apologize for the many mistreatments that the English language suffers each time we have to write in this language.

CONTENTS

PREFACE

Chapter 1 INTRODUCTION 1

Chapter 2 THE FIRM (1776-1870) 8

- Division of Labor 10
- Laws of Return 23
- Production Theory 32
- The Revenue Side 38
- External and Internal Equilibrium 39
- The Entrepreneur 44

Chapter 3 COURNOT 57

- A) Demand Curves 62
  - Cournot's firm's demand curves in various market morphologies 65
    - a) Monopoly 65
    - b) Duopoly 67
    - c) Oligopoly 76
  - Cournot's contribution to the types of modes of behaviour of the firm 83
  - Unlimited competition 95
- B) Cost Curves 100

Chapter 4 THE FIRM (1870-1920). THE NOTION OF OPPORTUNITY COSTS 104

- Division of Labor 105
- Factors of Production 106
- Theory of Production 108
- Costs 110
  - Real costs 111
  - Comparative costs 112
  - Money costs 115
    - Single Product, Long-Run 116
    - Single Product, Short-Run 122
    - Multiproduct Firm 129
      - No Interdependence, Long-Run
        - Fixed Proportions 130
        - Variable Proportions 132
      - No Interdependence, Short-Run
        - Fixed Proportions 143
        - Variable Proportions 144

Chapter 4 (continued)

Interdependence, Long-Run
Cost 147
Sales 152
Simultaneous interdependence in sales and in production 156
Interdependence, Short-Run 160
Limitations 161
Historical Section 176

Chapter 5 THE FIRM (1870-1920) THE LAWS OF RETURN 183

I. Production Function 183
II. Substitution Principle 193
III. Laws of Returns 197
IV. Laws of Returns: First Branch 197
Area of Factor Utilization, Long-Run 233
Area of Factor Utilization,Short-Run 237
V. Laws of Returns: Second Branch 239
Expansion Path 243
What governs marginal productivity? 244
VI. Laws of Returns: First Branch, History 246
VII. Laws of Returns: Second Branch History 269

Chapter 6 THE FIRM (1870-1920). MARGINAL PRODUCTIVITY, COSTS CURVES, AND SUPPLY CURVES 275

I. Marginal Productivity Theory 275
II. Cost Curves 285
III. The Supply Curve 291
Free or Open Competition 292
General Markets 295
Special Markets 299
Monopoly 304
Characteristics of the Marshallian Firm 306
IV. The Theory of the Firm in Other Writers than Marshall 311
V. Methodological Evaluation of the Theory of the Firm, 1870-1920. 319

Chapter 7 THE PERIOD 1920-1939 323

I. Monopolistic Competition 323
Analytical tools 325
Equilibrium conditions 337
Modes of behaviour 339
II. The State of the Traditional Theory of the Firm 343
Demand model 344
Supply model 347

Chapter 8 THE CONTEMPORARY THEORY OF THE FIRM 359

I. Methodology 359
II. Legitimate Extensions of the Traditional Theory of the Firm (1939-1960) 380
1. Price Variable and Oligopoly Problems 382
2. Extensions of the Theory of Product Variation and of Selling Costs 389
3. Extensions of the Theory of the Firm with Respect to Size and Growth of Firms 402
III. Efficiency Methods and the Traditional Theory of the Firm 406
Operations Research 408
Game Theory 410
Decision Theory 414
Simulation 416
Theory of Inventory Management 419
Location Theory 421
Marketing Theory 427
IV. Extensions of the Firm Through Unorthodox Channels 430
A) Introduction 430
B) The Organization Factor in the Theory of the Firm 434
C) Personal Goals of the Entrepreneur. The Problem of Multiple Goals 452
D) Integration of the Financial Factor 458
E) Size and Growth 463
F) Behavioural Theories 467
G) Methodological Considerations 474
V. Conclusion 482

APPENDIX A 483
APPENDIX AA 489
APPENDIX B 493
APPENDIX C 494
APPENDIX D 498
APPENDIX E 508
APPENDIX F 518
APPENDIX G 521

BIBLIOGRAPHY 524

CHAPTER I

INTRODUCTION

Economics started by being political economy. No wonder then, except for Marshall, that the firm as a subject of study has been somewhat neglected. This political bias remained prevalent in the profession until quite recently. In the beginning, the main efforts were spent on the study of aggregates, for example, industries, and what is regarded today as macro-economics and economic development. The excursions into micro-economics were brief and were oriented towards the making of special tools to meet particular situations. The micro-theories of the 1870's are still the backbone of our body of economic knowledge in that area, although again and again they have been criticized as false and based on outmoded psychology.

The economists did not spend all their efforts remodeling this edifice because after 1925 their preoccupations shifted to a large extent to problems of macro-economics. And especially at that time, macro-theories were established almost independently; that is they were not simple aggregations of micro-behaviours.[1] One example is the consumption

---

[1]Even today "macro-economic models can be set up in their own right..." (Allen, 11, p. 694).

function whose connections with the micro theory of the consumer behavior is quite loose and remote. Another reason is that most economists, for instance, Pigou, did not feel it was the task of the economist to advise the businessmen, although the economists have for a long time been advisers to the government.

Being an inhabited house, the state of the theory of the firm never worried the economists much. Recently, however there has been a renewed interest in micro-theories. And this for many reasons. First, macro-economists began to realize that independent model building "cannot be satisfactory"[2] so that now the economists should "in every case, develop a theory of micro economics and then show exactly how (they) construct aggregates in order to arrive at a theory of macro-economics," (116a, p. 13). Secondly, some economists either went into business as consultants or as companies' economists, or even wrote books on the subject, sometimes taking great care (and sometimes not) to tell the audience that in doing so they were putting on new hats. In this last instance, the experiences showed that the economic theory of the firm is lamentably wrong, irrelevant, and insufficient. Criticisms of the theory on these grounds or

---

[2]Allen, (11, p. 694). The full quotation is: "this cannot be satisfactory to the economist conscious that relations between aggregates are the resultant of many decisions by consumers and firms."

attempts at repairing the theory did not go unnoticed and brought severe scoldings from the elders who were quite satisfied with the traditional theory of the firm. The effect of the controversies that ensued was to divide those interested into two camps: the traditionalists and the revisionists.[3]

Yet, it seems when inspecting the literature that many discussions or controversies are useless because some of the participants do not pay enough attention to what is actually involved or what is meant exactly by the theory of the firm, or for what purpose it was devised. In other words, there seems to be a need to spell out in detail what are the characteristics, postulates, assumptions, and predictions of the various theories of the firm.[4] We say "various theories" because the first thing to be noted is that the theory of the firm differs among economists.

One encounters very often in the literature references to the 'traditional' theory of the firm. Controversies are sometimes engaged under its banner. Yet, there is no book or article entitled: The Traditional Theory of the Firm, An Exposition. Apparently its content is obvious for any of the participants. For some, i.e. Pigou or P. W. S. Andrews, its content is almost identical to Marshall's Principles of

---

[3]These terms will be defined in Chapter VIII.

[4]We may add to that the fact that, in general, economists don't read. See the admission of this point by Machlup and Weintraub in the American Economic Review, Proceedings, May, 1955. Vol. 45, pp. 482-483.

Economics; while for others, i.e. Downie (68, p. 19) it amounts to what can be found in Chamberlin, Mrs. Joan Robinson, and Machlup's The Economics of Sellers' Competition.

It can be noted that this form of speaking is rather confusing because, as will be shown in the following chapters, all these formulations are far from equivalent. Marshall is different from the reformulations of the 1930s which in many aspects are logical continuations of Cournot's work and not of Marshall's.

Our first task will, therefore, be the construction of a traditional theory of the firm. This will occupy us for the next six chapters, wherein the implications of its components will be brought in. Finally, we will show that after 1939 there is a conscious trend at expanding the theory of the firm along lines not reconciliable with the traditional ones.

The exposition should be sufficiently detailed so as to enable the classification of students of the firm into at least the two groups referred to above. The program of the thesis is partly to give an account of the origins of the theory of the firm, and to trace the course of its evolution, and partly methodological.

The theory of the firm, as in the case of many other topics in Economic Theory, lends itself to a treatment within an historical framework. It so happens that the way the theory was developed historically corresponds more or less to the logical processes of constructing such a theory. This is a

case where both approaches, the historical and the logical, help each other. This is why we have chosen to proceed in this fashion which we hope will provide a wider perspective to evaluate the current controversies.

For instance, an historical survey is necessary to determine the nature of the theory of the firm and its evolution. This means discovering why the fiction called the firm was invented at all. This means also listing its constitutive elements and achievements. In this case a chapter on Cournot is necessary because it appears that the present generation of economists is well trained in Chamberlin but apparently does not read the original. Furthermore it will be shown, that the traditional theory reached a summit in Cournot, and thereafter developed along two main branches, one consisting in introducing more and more realism into this fiction to a point where today the original concept is no more a fiction but is something akin to a managerial theory of the firm. The other consists of developing the theory along the traditional Cournotian postulates of profit maximization, marginal analysis and automatic entrepreneur. This last branch confines itself to the framework of price theory.

For most people, Marshall stands as the originator of the theory of the firm, (185, p. 1150). However, without introducing prematurely a controversy on this subject, it may be agreed that the constituent elements originate much further back, possibly with the classical economists. Marshall did

not invent the entrepreneur; furthermore, he invented neither the notions of costs and revenues which are used by the firms to set up their operating plans, nor the method used by the firm to reach optimum output.

A small excursion into the past suggests that the theory of the firm, like the Waltz of Ravel, has a confused beginning. Furthermore, time added not only frills to the theory but even components without which it would now be difficult to do. Since apparently nobody ever said "Here is the theory of the firm," it might be useful to ask, "who formulated it and what is its place and its relevance to the general body of economic theory?"

Before attempting this task, we should say a few words of caution. The decision as to when the theory was formulated properly surely depends upon the criteria used, and they vary according to whether they are formulated under one school or another. Secondly, the survey is limited to the well-known writers. Only the obvious contributions of these authors will be used. The material used is that which can be collected by an ordinary reading of the authors and their commentators. This should be enough to determine how the theory of the firm has been formulated by economic theory.

The formative period extends to and includes Marshall and is treated in detail in this thesis because there does not seem to be available anything especially oriented towards the particular topic of the firm. On the other hand, the

period 1920-1939 which saw the development and refining of monopolistic and imperfect competition theories is more well-known and can thus be dealt with more succintly. Furthermore, our aim is not only historical but mainly methodological. This is why any concept brought in will be evaluated for its methodological value as a component of the theory of the firm.

We would have preferred to adopt a simpler theme, but for many reasons this seems impossible at the moment. One reason is the confused state of the issues. In the development of a theory there are sometimes two phases: (1) the clarification of the issues, (2) the comparison of the various issues and their empirical testing. Our attitude is that in the theory of the firm there is still some spade work to do in the first phase. Another is that a historical exposition and a methodological evaluation of the theories of the firm seem to be the only types of study possible. A simpler theme could have been one of deciding which school is more successful in predicting and explaining items such as price or outputs, in a given situation, or for the economic theory in general. Unfortunately, this question cannot be answered satisfactorily by this thesis because the theories of the firm, especially the new ones, are not complete enough yet to permit empirical testing.[5]

---

[5]Yet the revisionists hope to be ready within the next decade. See: (59, p. 81)

CHAPTER II

THE FIRM (1776-1870)

To perform its activities the firm uses some basic tools such as the division of labor; the division of labor leads to diminishing returns or to increasing returns etc. These are among the basic elements of the theory and they are provided (as is well known) by the classicists. Consequently by common consensus (136, p. 200), it would be useless to go much further back than Adam Smith. This is not to say that the theory of the firm starts with Smith; but it is to say that some of the indispensable elements of the theory originate in the classical theory.[1] Credit cannot be given to the classicists for having invented the theory of the firm because they had no organized theory of the individual business unit. They had no concept of the entrepreneur, no satisfactory profit theory (they did not separate the rewards of the management from the rewards of bearing risks, providing capital, and carrying innovation), no mechanism of the operations of the firm; they made no attempt to find laws

---

[1]A theory is an organized knowledge of facts. Even if some facets of the theory can be traced to somebody, this does not make of him the originator of the theory, for example, the fact that Richardo used the incremental approach while studying rent on the marginal land does not make of him a marginalist.

that govern the size,[2] output, pricing policies, of the firms; in other words, they had no theory of an optimum for the firm. In fact the classicists were mainly preoccupied with the explanation of the value of a commodity and the relative shares of the factors of production, and with what we call today: economic development. Moreover, and this is what disqualifies them, their frame of reference was predominently a macro-economic type of economic analysis; that is, a type of analysis concerned with great collections of economic units, for example classes of factors of production, like labor and capital. Only in a few tenets of Ricardo's theory of value, do some types of economic reasoning follow a micro-economic approach, especially when one economic unit is taken as "representative" of a whole class, e.g. the farmer, (21, pp. 15-17, 24-25, 40). This method of analysis, (the one of using a "representative unit" to account for a whole class) is not completely micro because obviously this representative unit is not individualized but something approaching the "average". It is also not completely macro because the analysis is worked in terms of a single unit.

Having other problems in mind, it is only indirectly

---

[2]Smith is an exception because his treatment of the division of labor may have an 'implicit' connection with the determination (or limit) of the size of the firm. This point is dealt with later p.11ff below.

that, at various points, the classicists were forced to develop tools that are in use today as components of the theory of the firm.[3] However, as they provided the master theory of value and price, we should expect to obtain in their exposition an embryo of the theory of the firm, especially in that section of economics called production. Yet curiously "production and distribution do not seem however to have been used in England before 1821 as titles of division of political economy," (33, p. 26).

DIVISION OF LABOR

Notwithstanding this lacuna, Smith furnished the corner stone for a theory of production, and consequently for the theory of firm by formulating the principle of division of labor. The division of labor is indeed the most important concept for the theory of the firm. It is hard to visualize a firm without division of labor and vice versa, although the second alternative is more plausible.[4]

Smith not only provided the all important element of the division of labor but he also ventured into the determin-

---

[3]Knight, (118, p. 68): "The general conception of productive organization in relation to distribution is worked out by the classical writers only in connection with agriculture and with that only in a primitive form or aspect."

[4]See appendix A.

ation of the size of the firm[5] and into the extent to which the market mechanism alone can carry on the division of labor. Both problems are solved by his famous statement that the division of labor is limited by the extent of the market, (198, Bk. I, Chap. III).

Remembering that presently the production of final goods is, except for a few exceptions, carried on by firms,[6] let us classify these firms, according to their cost situation, into decreasing costs firms and into increasing costs firms, (205, pp. 192-193). If the firms are experiencing decreasing costs, this situation leads to monopoly, and oligopoly and the size of the firm here is clearly limited by the extent of the market (i.e. by demand). Also in this situation no production can be carried on through the market mechanism exclusively because, by hypothesis, all the demand is satisfied by a monopolistic firm. Here only the case of specialization within a firm is a possibility.

If the firms are eventually experiencing increasing

---

[5]The division of labor is also today the 'first principle' to take into account when "considering the various factors which determine the technical optimum scale of production" of the optimum technical unit. Smith did not explicitly make the claim. But since in this essay we are interested in the nature of the contributions, we must show the possibilities of the concept into the theory of the firm. See also Robinson, (176, pp. 13, 17).

[6]The statement of Smith obviously applies only in the case of a market economy. Firms did not exist in pre-market economies.

costs, the limit on the size of the firm will not be directly (although it will be indirectly) the size of the market for this commodity, but will be the point where the MC of the firm meets its own particular MR. Now, since the particular MR of a firm may be a function of the number of firms already present in this market, the limit or size of the firm is not directly connected with the size of the general market, but with the presence of other firms. This will be true for any market structure, competitive or monopolistic competition. This is because the firm determines its own size by using its particular demand curve.

This is well illustrated in the case of perfect competition where the limit on the expansion of the individual firm is not directly the extent of the general market (the industry demand curve) but the presence of competitors who share the general market, and may influence the cost situation of the firm and thus restrict the size of this particular firm. In the situation of a sudden increase in the market demand, the price will rise in the short-run period. In the short-run, the present firms will increase their output and will make plans to increase their size. However, in the long run, new firms will come in and this will affect the size of the firms, only if the industry is either an increasing cost or a decreasing cost one. In this last instance, the long run size of the firm will depend upon which types of factors will have their costs decreased. If factors such as machinery, or plants, experience a reduction in price, the size of the firm will have a

tendency to increase. Under a constant cost industry, in the long-run the size of the firm will remain the same.

In market structures such as duopoly, oligopoly, monopolistic competition, the size of the firms seems limited also by the output policies of the other firms. However, the notion of limitation of the size of the firm by the market can be reintroduced if we use the analytical device that consists in stating that each firm has, so to speak, a special market, (137, p. 182). For instance, the cases of monopolistic competition and chain oligopoly can easily be represented by a model of spatial competition where each firm has the monopoly of its own area (127, p. 105 ff.) which becomes its market area, and the corresponding demand curve determines (with the costs of the firm) the size of the firm.

Only one case now remains irreconciliable with the extent of the market as the sole limit to the size of the firm: it is the case of perfect competition mentioned previously. However, Stigler[7] seems to find one way out of this difficulty and reestablishes the symmetry for all possible market situations.

Here is the scheme: Granted that the various functions of the firm in perfect competition can be separated without changing the efficiency, the original firm should concentrate

---

[7]Stigler, (205). I warn the reader that I do not follow Stigler closely.

on those functions where it is experiencing decreasing costs (provided that at least one of these functions exists), and abandon the other functions to outside firms.[8] The initial firm will become a monopoly and its size will then be clearly limited by the extent of the market for the final commodity in question. The secondary firms, having taken over some increasing costs processes, analyze their functions and may (presumably) be successful in segregating one function in which they can experience decreasing costs. This is possible because when they took over the job of producing the intermediate goods they devised a new production function probably different from the one used by the original firms; this is why it may still be possible to find some function that behaves in a decreasing cost manner. Under these circumstances they will immediately abandon all the other functions and sub-contract them to new firms. If carried to its limits, the process leads to a world of monopolies.[9] In such a context

---

[8]Stigler proceeds in the opposite direction, but both methods lead to the same place. More serious is the objection that it is illegitimate to postulate that the functions when considered independently are likely to experience varying returns. The tendency today, for example, in linear programming and input-output analysis, is to postulate that each process yields constant returns. In such a perspective Stigler's point vanishes.

[9]Stigler (205) does not draw this conclusion but prefers to develop a theory of vertical integration. While I think that such a theory is more at home in a framework which would focus the attention on:
(a) the costs of using the market mechanism, (b) the interrelations between processes or in sales, (c) stability in the source of supply of intermediate products or factors of production.

(at least for all market structures, except the purely competitive case) the division of labor, or the size of the firms is a function of the extent of the particular markets of these firms.[10] We must now turn to the second branch of the argument which is to evaluate the prospects of having commodities produced exclusively through the market mechanism. The answer depends upon two sets of data:

(1) the extent of the market for the finished commodity,

(2) the costs situation with respect to the integration of the various processes to produce the finished commodity.

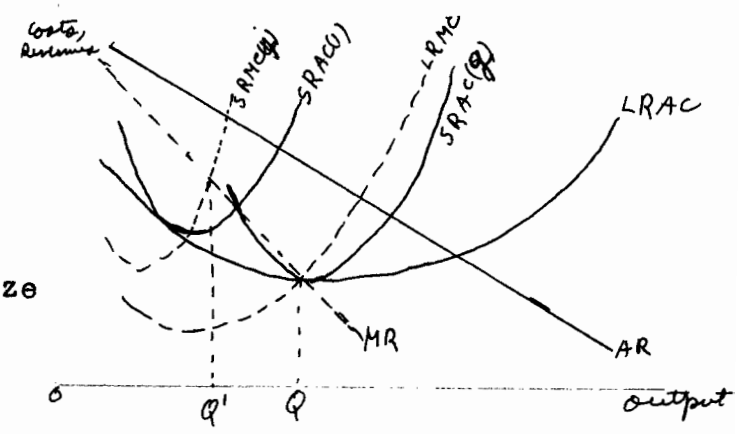
If the market for the finished goods is small it might not pay a firm to integrate and undertake the complete production. In this case, the intermediate products will be made by different firms and their coordination will be made through the market mechanism, especially if there are decreasing costs for the various functions (intermediate products) or if the market for intermediate products is big. (It may happen that an intermediate product has various uses in many different finished commodities). In such a situation the market surely

---

[10]We do not have in this context to take care of the converse situation where combining processes would lower aggregate costs, because if this was the case, the subdivision of processes should have stopped short of that point.

limits the size of the integrated firm since its size might be reduced to zero. If the market mechanism is effective and cheap, and if the commodity is produced at all, it will be produced by a firm only in the last stages of the production of the commodity, thus reducing the size of the firm. Too, it might be produced exclusively by households as described in appendix A.

Adam Smith not only provided one limit to the size of the firm in most market situations and to the division of labor, by using the demand approach, but he worked another limit from the point of view of an internal variable of the firm: the capital previously accumulated by the firm. In his own words: "As the number of workmen that can be kept in employment by any particular person must bear a certain proportion to his capital..." (198, p. 421). Regarding the firm, this simply means that its size is also a function of the ability of its owner to furnish it with capital. In this respect the size of the firm may even fall short of the previous limit, the extent of the market. In such a situation, the firm has a greater demand than it can satisfy. In the diagram at the right, the equilibrium condition requires output OQ, that is a firm of size No. q, while the firm may command capital only for size No. 1. The firm will produce



only OQ' for lack of ability to get more. Another example would be the situation where the firms could branch out into a new product, but this new product requires a new equipment and the firm does not have the capital necessary for this new venture. These are then two cases of underequilibrium situations.

Summing up, we have now two limits on the size of the firm in Smith (taking the size of the firm as synonymous with the degree of division of labor). Expressed differently, we have two measures of the size of the firm, whichever is smaller:

(1) the extent of the market; demand side,

(2) the amount of capital that its entrepreneur can secure.

These propositions[11] are very realistic, and are still used extensively today into the "new" theory of the firm.

Smith also reworked the whole analysis from another point of view, the one of capital looking for possible employment opportunities, "the quantity of stock (capital) which

---

[11]Mill was to follow Smith closely on this point by his first of four fundamental propositions respecting capital: "Industry is limited by capital." However he went into error when he stated that the converse proposition that "industry being limited by capital implies that every increase in capital...is capable of giving addition employment to industry; and this without any assignable limit," (Schumpter, 185, p.640). We shall see below that Smith did not agree to that when he reworked the original proposition from this point of view, and showed there that industry can limit the employment of capital.

can be employed in any branch of business depending very much upon that of the labour which can be employed in it," (198, p. 135). In other words, industry, or the degree of division of labour; limits the amount of capital that can be invested in a particular type of employment. Noting that the firm in this case has no difficulty in obtaining capital, we may study this statement according to the two senses that can be attached to it:

If by capital we mean "money capital," this is clearly in accord with our former exposition on the extent of the market limiting the size of the firm. If the aggregate demand is such that only a small firm (or a few firms) are enough to satisfy this demand, it would be illogical to use a size of the firm greater than the equilibrium one, that is, to try to use more capital than is needed. In the diagram of page 16, if $SRAC_1$ is enough to satisfy the given demand (different from the one pictured), then $SRAC_4$ will not be used. If the capitalist wants to invest one million dollars and size No. 1 of the firm needs only $100,000, the balance, $900,000 will have to be invested elsewhere.

If we mean by capital "physical capital", for example machinery plant, this statement means either that:

(1) the state of the technique is such, (e.g. not enough advanced) not to require a lot of costly machinery. For instance, a lawyer does not need much "capital" to operate.

(2) Labor is so abundant and cheap that it is not economical to use much capital, although the technique would permit the use of a lot of capital at certain relative prices. For instance, road or dam building can be undertaken in India almost exclusively with labor. A country experiencing a lot of unemployment would also fit the case.

(3) The technique and the relative prices of labor and capital (machinery) would use a lot of capital, but the demand is too small to warrant a big firm or a great number of firms. We here refer back to our previous case where the reasoning was in terms of money capital.

Up to now, we have established four propositions about the concept of the division of labor:

(1) It is essential to the existence of a firm;

(2) it is limited by the extent of the market;

(3) it is limited by the capital available;

(4) one of its effects is to limit the use of capital.

It remains to explore the "causes" of the division of labor if one dares to use such a dangerous word.

In Smith, the "illustration of division of labor in the pin manufactory, the advantages of division of labor derived from increased dexterity rather than from improved machinery. It was invention which flowed from division of

labor, not division of labor which flowed from invention."[12] More exactly, the division of labour "is the consequence of a certain propensity in human nature; propensity to truck, barter, and exchange one thing for another," (198, p. 13). This proposition is in exact opposition to modern thought. The modern theory of organization (122a, Chap. 8) holds that, on the contrary, the degree of division of labor depends upon the invention of machinery. We will not explore for the moment this second trend of thought, that is modern organizational theory, of which Smith seems to have had an inkling, when he ventured to find the causes of division of labor. We can now suggest a general picture of the developments to which this element (division of labor) gave birth. Diagrammatically:

Division of labor ——→ Modern concept of organization.
↳ Laws of Return → Production Function → Law of Variable Proportions → Cost of Production → Costs Curves, → ····· etc....[13]

Two main developments are distinguishable. The first one, modern concept of organizational theory, is already referred to above. The other development leads to a complete analysis

---

[12] Lekatchman, (123, p. 72). Also A. Smith (198, p. 9): "I shall only observe therefore that the invention of all those machines...seems to have been originally owing to the division of labor..."

[13] A more elaborate scheme is provided in appendix B.

of the cost aspect of the model of the firm. Smith provided mainly the corner stone here; the other items, production function, cost curves, etc. were provided by the writers who came after him, but they all undoubtedly built upon his foundation. A short explanation of the links between items is à propos. For instance, it may be asked, "In what respect is the division of labor a prerequisite to the laws of return, the production function?" The answer is that to the extent that the division of labor is meant to mean specialization, specialization can be connected with efficiency, (176, p. 4ff.). Changes in the degree of specialization lead to changes in efficiency and consequently to changes into the cost situation (average costs, marginal costs, etc.). At this point it is easy to see that applications of further doses of a factor have an effect on the degree of specialization. For instance, if the proportion between capital and labor is changed, this has an effect on efficiency and consequently on the state of costs or returns.[14] This intimate connection between division of labor and the laws of return may explain why Smith was brought "though loosely to state a law of increasing returns for manufactures" (185, p. 259) (in Book I, Chap. I). Also on another occasion, "in the digression on

---

[14]This is true only if we speak generally. See Schumpeter, (185, p. 1045); and Lerner, (124, p. 164).

the 'Effects of the Progress of Improvement upon the Real Price of Manufactures'" (185, p. 259) he explained increasing returns by the use of "better machinery, greater dexterity, and a more proper division of work" (185, p. 259).

This long catalogue accounts for the main contributions of Smith. A few minor ones beside should be noted. We may mention here the distinction between fixed and circulating capital (Book II, Chap. I). His concept is much different from the modern counterpart where circulating capital is more or less replaced by the concept of working capital. In the same chapter he also brings in the notion of depreciation to keep the capital intact and thus reduce gross income to net income. In this connection he draws (p. 48) an imperfect profit and loss statement for a firm when he tries to compute the rate of profit. Finally, although this is not a specific contribution to the theory of the firm, his market mechanism (p. 56) is an important tool that will be used extensively by the whole body of economic theories.[15]

---

[15]Almost contradictory opinions on this point are expressed by two eminent scholars:

(a) Robbins (174, p. 234): After having praised the classicists for their theories of division of labor and market, he says: "given law and order strong spontaneous forces working through the market were ceaselessly in operation tending to adapt the organization of production to the incluence of pecuniary demand."

(b) While Prof. Knight (118, p. 41) says quite the contrary: "The classical economists give no picture of a system of prices and practically no hint of a system of economic organization worked out and directed by price forces."

It seems that the classicists had a sense of economic forces but had not yet reached the price system stage.

LAWS OF RETURN

The classical economists contributed one more major element toward building a theory of the firm; it consists in what is called today the laws of return. The importance of this element[16] can be appreciated by looking at its place in the scheme of appendix B. This scheme illustrates the hypothetical chronological development of the theory of the firm in its aspect of costs. This diagram shows the logical sequence of how one concept could have led to another if the contributors had not missed any step.[17]

As will be realized during the exposition, the classicists and their followers did not adhere completely to this "logical order". The particular being always more easily grasped than the general, the development proceeded by jumps.[18] For instance, the law of diminishing returns in the use of one factor, (which they discovered much before the production function was formulated), is in fact a special case (the first

---

[16]This importance has been recognized in contemporary literature; see: Stigler (203, Chap. 8): "The Laws of Return and Cost Curve". Dorfman, (64, p. 1): "Ricardo and his contemporary Malthus share credit for the development of the concept of decreasing returns the progenitor of the modern U shaped cost curve."

[17]This is not a pure logical scheme because it is using one constraint: the choice of the elements is limited to those that had been historically used.

[18]"This is revelatory of the 'ways of the human mind', which rarely discovers the obvious and fundamental first. More often it gets hold of some particular aspect of an idea and then works back to the conception that hold priority in logic." Schumpeter, (185, p. 261n).

branch) of the general Law of Returns, which itself is a description of the production function. To appreciate exactly this contribution, we must match their contribution with the present formulation of the law which reads as follows:

Presently "the laws of Return describe the characteristics of the production function." They comprise two parts: "the first branch known as the law of variable proportions, is concerned with the behavior of total product when only one productive service is varied in quantity," (203, p. 116). In this branch of the law, increasing and diminishing returns "are...due to changes in the proportion of different factors used in combination..." (124, p. 69). Generally speaking it refers to either decrease or an increase in the average product, such an average figure being obtained by dividing total product by the amount of particular factor in use (124, pp. 152, 155). The law is also expressed in terms of the marginal productivity of a factor;[19] the marginal value is obtained by dividing the increment in product by the increment in factor.

The second branch of the law is labeled: "the law of return to scale of plant" and it "considers the general problem of the effect on the product of variations in the

---

[19]"The principle of diminishing mp (marginal product) is sometimes called the law of diminishing returns, but that name is more often--and more conveniently--applied to the diminishing average product" (Lerner, 124, p. 151). Later when studying the period 1870-1920 we will discuss the merits of both formulations.

quantities of all the productive services" (203, p. 116). The two branches of the law can be contrasted with the utilization of a production diagram[20] involving two factors of production and one output. In such a diagram, moving along parallel to either of the axes illustrates the workings of the first branch of the laws of returns, while moving in a northeast direction would illustrate the working of the second branch of the laws of returns (in this case returns to scale).

Since it can be shown that with a certain production function (homogeneous with degree one) we can simultaneously have constant returns in one branch and diminishing returns in the other, mentioning only diminishing returns, without specifying the branch to which it applies is to talk loosely to say the least. Furthermore expressing the first branch in the form of diminishing returns is not sufficient because it is not known whether the statement refers to diminishing average product or diminishing marginal product; and we will see later that these two formulations are not equivalent.

This distinction and others, were not made by the classicists. Actually, these things were not completely settled before 1936.[21] The Laws of Returns were the progenitors

---

[20]See appendix C for an exposition.

[21]Schumpeter, (185, pp. 587, 1037). Karl Menger, (143) is the one who settled the question.

of all the other developments because they forced the economists who wanted to explore their implications, as far as the production process is concerned, to get acquainted with the notions of production, the connection between inputs and outputs (production function), and how particular output varies when only one factor is varied (marginal productivity). The necessity of these explorations implies that they themselves had not mastered these notions. Furthermore their contributions to the laws of returns occur rather late. Smith himself had not much to say on the laws.[22] Curiously here we have to go back to Turgot who is credited with the best formulation.[23] Long before the classicists,[24] economists had the feeling that in the case of manufacturing industry there was a law of increasing returns manifesting itself in the form of decreasing unit cost as output increases. This is not

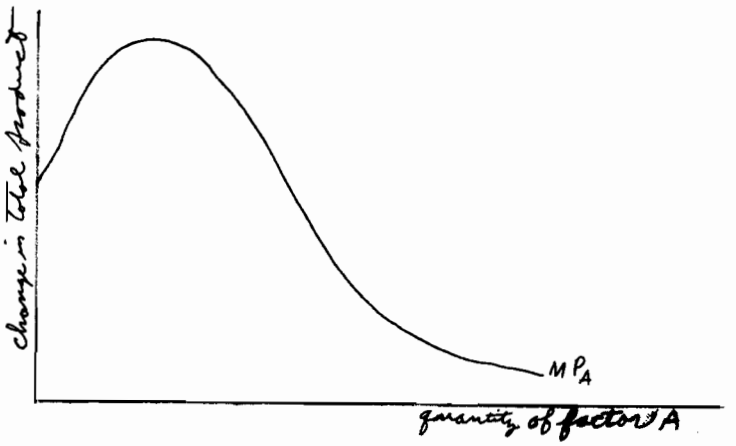
---

[22]"Nowhere did he state a law of decreasing returns though he repeatedly brushed against it" (Schumpeter, 185, p. 259). About the law of increasing returns, ibid. See also Knight (118, p. 74): "Smith has no statement or clear implication of the principle of diminishing returns in any form."

[23]"Turgot's formulation was indeed not surpassed until Edgeworth took the matter in hand" (Schumpeter, 185, p. 260). Also "Few writers offered a statement of the principle of varying returns clearer and more concise than that of Turgot." (Whittaker, 225, p. 385).

[24]Schumpeter (185) mentions pp. 258-259 that "a general law of increasing returns in manufacturing industry, also in the form of a law of decreasing unit cost, had been stated explicitly in full awareness of its importance by Antonio Serra (1613) much as it was to be stated in the nineteenth-century textbook."

very precise and does not differentiate between the first and the second branch of the law. Moreover, immediately before Smith[25] both the extensive and the intensive margins of use of a factor had been explored, thus presenting the first branch of the law (various amounts of labor are successively applied to a given piece of land). The next refinement was introduced by Turgot[26] who noticed that there may be an initial phase of increasing returns (to further units of a factor) before decreasing returns set in. This is the basis of the peculiar shape of the modern MP curve. Furthermore, it must be noted that the law is formulated by Turgot "in terms of successive increments and not in terms of average product (per unit of variable factor)" (185, p. 261), thus paving the way to marginal analysis. To obtain stability in the MP curve, we have to make sure to eliminate changes in the state of the technique which might induce shifts in the curve. The distinction between increasing



---

[25]Steuart (1767) had presented the case of extensive margin (See Schumpeter (185, p. 259) and Turgot the case of intensive margin.

[26]See pages 384-385 of Whittaker (225) for a sample of the writing of Turgot.

returns due to the working of the law and increasing returns due to changes in the state of the technique was made around 1800 by Anderson (185, p. 263). However Turgot's work was not brought to the attention of the English economists, so that in a way they had to rediscover the laws.[27] At the time of Ricardo[28] the laws of return had already been formulated in various degrees of exactitude, and comprehensiveness. With Ricardo's coming, the first branch of the law was remarkably well stated in micro-economic terms; both sides of the law (increasing and decreasing phase) were mastered. A synthesis of both sides remained to be made in order to have a law of proportionality.[29] The first writer to do this synthesis was West who showed that in agriculture, for instance, the two tendencies were simultaneously at work: (1) one towards diminishing returns per unit of effort due to the fact that we must resort to working poorer soil or working more

---

[27]"As a matter of fact the law remained practically unknown till near the close of the great War. Malthus may, perhaps, display some inkling of it here and there...but to imagine that the Essay on the Principle of Population was ever based on the law of diminishing returns is to confuse Malthusinanism as expounded by J. S. Mill with Malthusianism as expounded by Malthus." (Cannan, 33, pp. 113-114).

[28]"He certainly had much to do with the addition to nineteenth-century political economy of the 'law of diminishing returns', but he and Malthus and West seem always to have been more concerned with the effects of that law on distribution than with its effects on production" (Cannan, 33, p.31).

[29]Petty and Smith and several other writers touched on both sides of the law, but none except Turgot dealt extensively and in a unified manner with both.

intensively good land, (2) another one in the direction of increasing returns due to the division of labor" (225, p. 391).

This "synthesis" does not fail to cause some problems as far as the static formulation of the Law is concerned. If the alternative is working more intensively the same land, the first branch of the law (the size of the firm is not changed) remains in effect. If the alternative is to "resort to working poorer land," then both factors (land and Labor) might[30] be varied and we should notice a change in size and thus, the second branch of the law. Moreover we are not even sure that we can legitimately do that because the homogeneity of the factor units (land) is not preserved in the latter case of West. Consequently, even with this effort, we have not yet a clear distinction between both branches of the law. This imprecise way of thinking was continued by Malthus who reverted to a dynamic point of view: he added that large scale production besides inducing increasing returns through a greater division of labor, may breed improvements in technique. This is reminiscent of Smith where the division of labor causes "inventions"; but this is also confusing because the shifts of the curve are hopelessly mixed up with movements along the curve.

---

[30]We say "might" because it is possible that we resort to "poorer" land while keeping the same labor force; in this case it is compatible with the first branch of the law, yet it must be noted that the homogeneity of the units of the variable factor is not preserved.

The most concise formulation of the law in those times (not the more precise, because he did not distinguish between branches of the law) was made by Cournot.[31] He states that in the case of manufacturing articles, the cost of production "becomes proportionally less as production increases, or in other words, where D increases, $\phi'D$ is a decreasing function. This comes from better organization of the work, from discount on the price of the raw materials for large purchases, and finally from the reduction of what is known to producers as general expense. It may happen however (that) ...when the exploitation is carried beyond certain limits, it induces higher prices for raw materials and labour to the point where $\phi'D$ again begins to increase with D."

This formulation is deficient on many points. Among other things it is probably limited to the second branch. Furthermore increasing costs are not brought about by diminishing productivity (physical) but by a rise in the costs of the factors. However it was chosen to be quoted as an expression of how the people of the time looked at the matter. Its elegance given by its mathematical flavor is worth noting.

Cournot being in his time a completely ignored writer, the English authors had to rediscover or polish their theories

---

[31]Cournot (54, pp. 59-60). Among the many quotations available to show how the law was expressed in those days, we have, chosen this one; others are available in the references listed before.

alone. After the West-Malthus-Ricardo formulation, clearer statements are encountered, although from then on the substance did not change. For instance, Senior reemphasized the necessity of keeping the state of the arts constant. He restated the law (obviously limiting himself to the first branch) in terms of a fixed factor (land) upon which an increasing amount of labour is bestowed, the postulate being that the increase in product is not in proportion to the increase in labor.[32] Furthermore he did not spare us the ever-recurring statement that decreasing returns are felt in agriculture and increasing returns in industry.

However, these formulations of Senior and West were not mathematically precise. Consequently their followers were confused, e.g. formulating and using the law in terms of average returns.[33] This error had two important consequences:

---

[32]See Senior's formulation as cited by Whittaker (225, p. 393) "That agricultural skill remaining the same, additional Labour employed on the land within a given district produces in general a less proportionate return, or, in other words, that though, with every increase of the labour bestowed, the aggregate return is increased, the increase of the return is not in proportion to the increase of the labour."

[33]This makes some people judge the classicists severely on this matter. Knight (118, p. 72), goes as far as to say: "a more technical but quite fundamental criticism is that Ricardo and the other writers had no clear or accurate understanding of the principle of diminishing returns..." Their failure to use marginal analysis is also caused by the use of "increments (that) are enormously large instead of negligibly small as a sound analysis requires" (Knight p. 70). Knight concludes that there are many "indications that Ricardo did not really understand the principle of variable proportions." (p. 70).

(1) Since the average is not equivalent to the marginal, it barred classical economics from attempting any problems of maximization; in the theory of the firm, "it is the marginal concept which is needed in all maximum problems" (185, p. 587). We had to wait till 1911 for Edgeworth to point that out.

(2) Their failure to use marginal analysis is the reason why "the conquest of the idea of diminishing physical returns did not ipso facto and immediately lead to a marginal productivity theory and why the latter had a separate history at all" (185, p. 587).

It should be noted that we have not seen, except in Cournot, any trace of an appropriate formulation of the second branch of the law. Being obsessed with the fixity of the aggregate supply of land, they apparently never graduated to a case where all the factors would be plentiful and could be varied simultaneously at will from the point of view of the single firm. Yet most baffling is their persistent lack of clear distinction between the first branch of the law and the second. For instance, later with Mill, Carey, Bastiat, and Sidgwick, the interest in the laws of returns lies mainly in evaluating the probability, in the case of economic progress, of one side (diminishing returns) overcoming the other side (increasing returns), principally in agriculture.

PRODUCTION THEORY

Having seen the classicists manipulate the employment of factors on the marginal farm, we should be inclined to think that with such a vision of the economic process it would automatically lead them to establish the fundamentals of a theory of production. Nothing is further from the truth. "Before the middle of the eighteenth century a theory of production can scarcely be said to have existed. Durable objects being looked upon as the sole or chief kind of wealth, the function of industry and trade seemed to be the 'circulation of wealth'" (33, p. 128). Furthermore, the classicists on this topic had an aberration which marred their theory of production, when they have any, for our purposes; when they talk of production theory, they mean exclusively the production of wealth, not the production of goods and services as we mean now (118, p. 39). Ricardo and some of his immediate followers had no theory of production.[34] "It was Torrens who set the example of writing considerable quantity about production" (33, p. 31). And even there, it took some time for

---

[34]"If Ricardo had been asked where his theory of production was to be looked for in his Principles of Political Economy and Taxation, he would have answered with perfect justice, that in spite of the generality of its title, his work did not profess to deal with the production of wealth. It was merely an attempt to offer a solution of the 'principal problem in political economy', which is, he thought, 'to determine the laws which regulate' the distribution of the produce of a country between rent, profit, and wages." Cannan: (33, p. 31).

such a theory to take form. We had to wait for the second half of the nineteenth century to find textbooks "ascribing production to the cooperation or concurrence...of agents" (33, p. 32). "Even in 1848 the triad of requisites of production was not quite firmly established" (33, p. 33). In France, this step was made initially by Say at the beginning of the nineteenth century, and in England by Senior in the middle of this century.[35] With at least three factors, production takes the shape of a process of cooperation and coordination. This element of cooperation is crucial for the future of the theory of the firm.[36]

The establishment of the "triad" was in one way unimportant because it is not popular with modern economists, except Marshall, who do not see the point of restricting or classifying contributions to production in three classes.[37] Yet at the time it was a definite improvement over the Physiocratic doctrine with its emphasis on a single factor as pro-

---

[35]See Schumpeter (185, p. 557). Let us remark that it was not primarily a production setup, but a tool for distribution analysis.

[36]The firm is the nucleus of cooperating factors; see Liebenstein, (122a, p. 121). "A firm is more than an administrative unit; it is also a collection of productive resources ..." Penrose (160, p. 24).

[37]Yet, "it so happens that the triad also makes economic sense, because it presents a complete list of the requisite of physical production, the items that are both non overlapping and distinguished by economically relevant characteristics. Therefore it makes a useful basis to start from." Schumpeter (185, p. 561).

ductive. Smith spoke about three factors but could not admit that every kind of labor was productive.[38] Starting with Senior, little by little the model of production became a collection of productive resources, all conceptually of the same importance.[39] This is as far as the English classicists went. At the same time in France progress was made with great strides. Already Say was moving from the "triad" to a quartet of types of factors of production by adding the 'entrepreneur'. He was adding what would become in Marshall's hand the fourth factor of production. Meanwhile, in England, until Mill, nobody took notice of the fact. "Ricardo and the Ricardians accomplished what is described as an almost impossible feat, namely, the exclusion of the figure of the entrepreneur completely" (185, p. 561). Finally the classicists did not go as far as to make the next step which is to use

---

[38]"La théorie des trois facteurs de la production ne se trouve pas chez Smith. C'est cependant lui qui a donné indirectement naissance a cette théorie en distinguant, dans la distribution des richesses trois sources de revenu: le travail, le capital, et la terre. Cette distinction faite, il devait paraître tout naturel de considérer chaque source de revenu comme facteur de production. C'est ce que fit J. B. Say dans son traité." Gide et Rist: (90a, p. 62n).

[39]In France the factors were already on the same footing since Say (1803). Say, "possibly following Turgot's lead" began by establishing "the triad scheme and the practice of dealing both in the theory of production and in the theory of distribution with the services of three factors (land, labor, capital) on the same footing." Schumpeter, (185, p. 560).

the substitution principle among these factors.[40]

Concurrently with the establishment of the "triad", other useful elucidations were made. Mill gives the first clear distinction between two types of industries, agriculture and manufacturing, (123, pp. 185-186) and correctly states the law of diminishing returns with respect to agriculture (1st branch). He also distinguishes between the laws of production and the laws of distribution. He said about the laws of production, for example laws of returns, "...there is nothing optimal or arbitrary in them" (145, pp. 199-200). Of distribution he said, "That is a matter of human institution solely; the things once there, mankind, individually or collectively, can do with them as they like" (145, pp. 199-200). At times, the analysis leaves the physical world and graduates to the cost formulation. The classicists tended to enunciate various notions of costs according to the type of problems they faced.

Since much of their analysis had a welfare connotation, they devised a notion of cost that would serve as a "measure" in this domain. This notion of cost is known today as real costs. "The real price of everything, what everything really

---

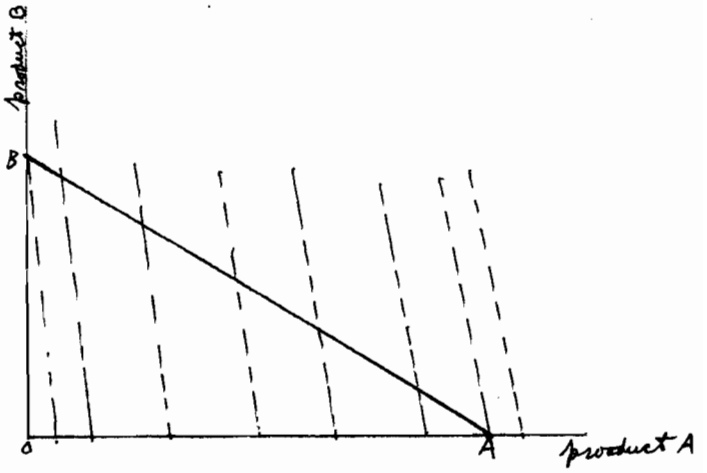
[40](185, pp. 590, 680n). Yet Schumpeter does not take into account the substitution between labor and capital in Ricardo's chapter on "Machinery". This was later to become famous under the name of the Ricardo Effect, see Allen (1, p. 638 ff)

costs to the man, who wants to acquire it, is the toil and trouble of acquiring it" (198, p. 30). This is not revolutionary, because this concept can be traced back to Thomas Aquinas' method of pricing, and even to Aristotle (225, pp. 407-418). During the classical period, labor became the irreductible element of costs. Later with Senior, abstinence, or the cost of capital, is added. The dichotomy of the classics theory is between real and money costs. They usually preferred to work their analysis in real terms because of the unreliability of money costs to measure "toil and trouble". The real costs are thus physical quantities and in the end reducible to a sum of subjective disutilities: efforts and pain. Although this notion is very interesting, it was never used directly in the analysis of the firm. Its importance lies in the fact that it was later used in another useful concept, the theory of comparative costs which they brought about to solve international trade problems.

The co-discoverers of comparative costs doctrine were Ricardo and Torrens (217, p. 441 ff., and 174, pp. 31-35). The theory stipulates that what is important in terms of economic efficiency is not the absolute real costs, but the ratio of real costs of various commodities. The "economy in cost of obtaining a given income was the criterion of gain" (217, p. 437). This served as the foundation for their plea for free trade. However we must note that their formulation was almost exclusively in terms of real costs and that these

costs were constant. This does not mean that they were not aware of money costs (217, pp. 470, 483). They often conceded that trade is actually conducted in money terms (174, p. 25). But as noted before, their goal was to solve a welfare problem and that is why any problem has to be translated in real terms.

As far as the theory of the firm is concerned, this apparatus can help solve simple problems. Suppose a firm in short run is endowed with a fixed amount of resources (thus we are working in real terms) or in long run the firm is endowed with a certain of money capital (thus working in money terms). Suppose also that it can produce two products, that their production is independent, and can be carried on under constant costs. The classical comparative cost doctrine can solve the product-mix problem of the firm. Geometrically, (see chart), if all the resources are devoted to product A, OA of A can be produced; similarly, devoting the resources to B, OB can be produced. Since we postulated constant costs, the straight line AB represents the combinations of A and B that can be produced with the given resources. Superposing a given set of iso-revenue curves, (as shown in chart), shows that the firm should



specialize in product A. Similarly, the same problem could have been solved through linear programming which in this elementary problem uses an approach akin to comparative costs.[41] This is about all that the classical economists had to say on costs. It is by far the most useful part of their analysis that can be incorporated into a theory of the firm.

THE REVENUE SIDE

Turning to the revenue side, it is commonplace to state that the apparatus is very insufficient and practically nonexistent, (185, p. 602m), at least up to Mill who "was to develop the supply and demand analysis so fully that, as Marshall himself was to indicate, there remained not so very much to do beyond removing loose ends and adding rigor..." (185, p. 603). Yet we should not be carried too far away by this classical contribution. Not being able to picture the demand as a schedule of quantities and prices, it prevents them from contributing in any direct manner to the theory of the firm on this topic. There was apparently no attempt to connect a particular demand, as a schedule, with a particular firm. Their contribution, as far as Mill is concerned, is

---

[41]Dorfman, Samuelson, Solow: (67, pp. 31-32). "The classical theory of international trade illustrates the concept of linear programming in an entirely different context." This is all we will say for the moment on this topic; later in Chapter IV, in the section on Costs, we will present extensive developments of these points.

limited to elaborating the basic concepts upon which a demand schedule can be built.[42]

EXTERNAL AND INTERNAL EQUILIBRIUM

The operations of the firm are carried on in two 'milieus' or surroundings, the external and the internal.

The external 'milieu' is the network of interactions of the firm with the outside world, that is with other firms, consumers, and resource holders. Consequently, this is where its data such as the demand curve, are built and where cost data are collected.

For instance, the external influences on the demand curve of a firm are:

(1) the industry demand,

(2) the morphology of the market, e.g. the presence of other firms and their policies,

(3) the competitive activities of the firm in question other than those involving adaptative manoeuvres in price or output. These competitive activities can take the form of selling activities, product differentiation, etc.

The firm can alter its market position, that is, influence

---

[42]However Mill, while working on International Trade problems, devised one tool of analysis which was, in the hands of Marshall and Edgeworth, to be quite useful in solving problems of bilateral monopoly. Modern analysis uses it in the form of the box diagram. The specific tool of Mill is the 'reciprocal demand'.

the shape and position of its demand and cost curves,[43] either by incurring selling costs, or production variation costs, or by embarking into tactical manoeuvres in terms of price and output. A tactical manoeuvre is any deliberate change, as differentiated from adaptative manoeuvres, in price or output which is not made to achieve internal equilibrium but as a part of a strategy to vary the market position of the firm. This means for instance, indulging in cut-throat competition, cutting prices below the one corresponding to MC = MR, and any competitive manoeuvre having the purpose of intimidating, ruining or bluffing rival producers. External 'milieu' is then synonymous with market position, and external equilibrium is achieved when the firm has no tendency to reshape the "contours" of its particular market area through the use of any of the weapons mentioned above.

The internal 'milieu' corresponds to a set of data, revenue and cost curves, as they finally appear on the drawing board of the firm. We suppose that, being accomplished facts, nothing, as far as their position and shape are concerned, can be done about them, except to adapt the output or price of the firm as best as possible with regard to these data. In other words, we hold that whether the firm tried

---

[43]I include cost curves in the schema because similar activities can be undertaken into the securing of advantageous conditions in the factor market.

or not to influence the shape and position of these curves, it is now confronted with the results and must do its best in such a situation. Only one variable remains in the power of the firm: the price or output.

Achieving 'internal' equilibrium consists, for the firm, in adjusting the price (or output) in such a way as to decrease the inducement to change any of these to better its position. This is exclusively an adaptative movement. The process of attaining this 'internal' equilibrium is either profit maximization or the attainment of other goals as determined by the entrepreneur who runs the firm. Seen in this perspective, each change in the data in the external 'milieu' will entail a disequilibrium which will have to be reestablished by altering the variable output, or price. These disturbances have an effect not only on the profit and loss statement, where the simple maximum principle MC = MR operates exclusively, but also on the balance sheet. This balance sheet, which is another sector of 'internal' equilibrium, may have to reestablish its particular internal equilibrium in the manner of the principle of homeostasis (25, p. 26).

Total or complete equilibrium, the subject of which is called the parameter theory, will be achieved when the firm has no inducement to vary any of the three variables at its disposition: output or price, selling costs and product variation.

The notion of market classification, for example,

competition and monopoly, was known before the period of classical economics.[44] Monopoly as a form of a market had been recognized as far back as Aristotle. However, this is not important, since we are interested in theory and not in a nomenclature. Smith does not score high in this matter.[45]

Aside from a description of the morphology of a few markets, namely monopoly and free competition, there is, in classical economics, no analysis of the operating mechanism of these markets.

They have two standard cases, free competition and monopoly, to which they refer without ever showing the logical content of this market framework. "In fact the concept (of free competition) was usually not even defined" (185, p. 557). Furthermore, this free competition framework is taken for granted as the "normal case" when they dare to bother about the market situation.[46]

---

[44]See Schumpeter: (185), pp. 305-306, 60-61, 150-155 and also Whittaker: (225), p. 466 ff.

[45]In Smith "there is no theory of monopoly price beyond the meaningless (or even false) statement that the price of monopoly is upon every occasion the highest which can be got whereas 'the price of free competition... is the lowest which can be taken in the long run'." (Schumpeter, 185, p.189).

[46]"Ricardo failed completely to specify his institutional assumptions" (Schumpeter 185, p. 544) and thus this market framework. However we may think that he works on the hypothesis of a Free Competition system which for them, the classicists..."was an institutional assumption rather than the result of certain market conditions" (ibid., p. 545).

The market form is for them an institutional setup, rather than an 'external milieu' subject to manipulations and having a definite influence on the determination of the behavior of the firm. They were, then, not tempted or obliged to explore extensively the connections between the morphology of the market and the modes of behavior of the firm. It is then impossible to be informed by the classics on the source or explanation of the behavior of the firms. They restricted their descriptions to one type of market setup; the owner-manager firm, operating in a competitive situation, whatever this means, and without realizing what this term implies in regard to the equilibrium analysis. For instance, Ricardo's theory of rent "is based on a highly abstract model of a single type of farming" (64, p. 1). They did not even show, in a competitive market, the interactions of the firms and "ignored the essential phenomena of price competition and economic organization through price competition" (118, p. 66).

Having no "conception of the nature of economic organization, ... and ... of the system or order best called enterprise economy," (118, pp. 41, 63-64), they are unable to interest us in such a concept as industry and the mechanism of interaction within an industry. The farmer in Ricardo's concept was not an entrepreneur taking initiatives of any kind. "For them...(the Ricardians)...as well as for Marx...the business process runs substantially by itself, the one thing needful to make it run being an adequate supply of

capital" (185, p. 556).

The internal 'milieu' is not better described. The firm is institutionalized instead of individualized; the internal process being automatic, there is no need to describe or analyse it. Their main lacuna here was that their firm had no maximizing mechanism.[47] One thing that did not help them in this matter was their failure to analyze thoroughly the revenue side. Obviously the maximizing process involves maximizing the difference between two variables, revenue and costs; since one was practically nonexistent, they could not build a complete picture.[48]

THE ENTREPRENEUR

Since its introduction before the classical period, the figure of the entrepreneur has experienced various fortunes in economics. After having attained a summit with Schumpeter (183), it came to be progressively neglected, so that today it is almost completely absent from economic theory (50, p. 29). Economic theory being the main source

---

[47]"Ignorant of mathematical concepts, the classical writers had no clear or definite conception of the meaning of economy as a process of maximizing a value return from resources of whatever kind" Knight (118, p. 41).

[48]This is not to say that they were not able to picture an over all view of the operations of the firm. Quite the contrary, Malthus (see p. 80 of Cannan, 33) draws a quasi-perfect profit and loss statement for an individual firm. This profit and loss statement differs and rightly so from those of the accountants because it includes as it should for economic analysis, the imputed interest on invested capital.

of the theory of the firm, a few comments might be needed.

Cole (50, p. 29) furnishes a plausible reason for this present neglect: economic analysis being based on the "economic man" which is rational and thus maximizes profits, there is, in such a model, no room for the entrepreneur whatever the definition used, and there are many rival ones, except if the entrepreneurship is completely assimilated to management.[49] In the present traditional theory the equation MC = MR rules out any special or distinctive influence of the entrepreneur per se, since he must follow the rule. Such that even today in the traditional theory, the entrepreneur is "automatic" or unnecessary.

Nevertheless, to decide whether or not the exclusion of the entrepreneur from the model of the firm has been wise, one must determine first what is meant by the entrepreneur. Schumpeter's definition for a long time was quite popular. For him, what distinguishes the entrepreneur from the capitalist and other kinds of labor is that he acts "outside the pale of routine" (184, p. 253). "The distinction between adaptative (routine) and creative response (entrepreneurial) to given conditions may or may not be felicitous, but it conveys the essential point; it conveys the essential difference" (184, p. 253), presumably to differentiate between

---

[49]We shall see in Chapter VIII that the Cole's position does not account for the present tentative of the revisionists to reintegrate an unautomatic entrepreneur in the theory of the firm.

the entrepreneur and the manager. In short, entrepreneurship is essentially innovation. However, this condition of innovation besides eliminating a lot of businessmen from the picture, is open to a few criticisms.

One defect may be that we are left without a precise guide to decide on particular cases. For instance, Schumpeter talks solely of a break in the routine as involving "entrepreneurial" labor; yet sometimes the decision to stick to a routine, as opposed to changing it by introducing a novelty which might be warranted by new circumstances, would involve, in the eyes of many people, a decision involving no "ordinary" managerial skill. Moreover, seen from the outside, the fact that a firm pursues the same policy during a certain period is no indication that "entrepreneurial decision" were not involved. Lastly, who can decide whether a situation involves "an entrepreneurial" or a "managerial", decision? The difficulty in managing a concern is in being able to discern among the problems that occur everyday, those that are significant and warrant entreneurial decision, even if the decision is <u>not</u> to change a routine, and those that involve "routine" managerial responses. Agreeing to this mode of thinking is to agree that every fact, routine or not, having a bearing on the firm involves "entrepreneurial" labor. Because obviously only an entrepreneur can decide whether a fact is worth considering at the entrepreneurial level or can be sent to the managerial decision center. In such a

case, the "entrepreneurial" function engulfs the managerial function.

Secondly, it is obvious that not every innovation should be worth attributing to an entrepreneur. Only the major ones should be so. This leads to two difficulties:

(1) sometimes innovation is carried by non-business groups, for example, engineers, or sometimes by a simple worker who deposits an idea in the suggestion box located in his foreman's office;

(2) secondly, what is a major innovation? In the beginning few people thought that the airplane would never be anything but a weapon of war or a showpiece.

Today many people, for example, Cole, (49, p. 89; 76, p. 50 ff.), have abandoned the innovating feature as the sole criterion. Entrepreneurship is now defined as "the purposeful activity, including an integrated sequence of decisions, of an individual or group of associated individuals, undertaken to initiate, maintain, or aggrandize a profit-oriented business unit for the production or distribution of economic goods and services" (50, p. 7). Consequently the emphasis today is toward identifying decision-making as the core of entrepreneurship (50, p. 10). Here, only top level decisions should be included, and this is once more an occasion to distinguish between entrepreneurship and routine management. Actually it is not correct to oppose entrepreneur-

ship and management as mutually exclusive categories. It is preferable (49, pp. 103-106) to use a spectrum as a frame of reference. In such a spectrum, minor decisions, for example, the hiring or firing of a night watchman would be a management decision, while gradually, as the decision involves more important matters, we move toward entrepreneurship. This method is still not completely satisfactory because there is no clear-cut demarcation between the two concepts. Recently Evans (76, pp. 50-51) has ventured to restrict the top level decisions involving entrepreneurship to those involving the determination of "the kinds of goods and service to be offered, the amounts of these to be supplied, and the clientèle to be served." This is probably insufficient, and the choice of the production function to be used, the mode of securing capital (fixed and working), decisions about launching, merging, and winding-up the firm should be added to the definition.

Under the production function heading, one can infer that the entrepreneur has also the function of coordinating, at the top level, the division of labor within the firm.[50]

---

[50]The coordinating function does not extend to the managerial field. The entrepreneur is not necessarily the one who decides who is right when there is a dispute between a foreman and one of its subordinates. Entrepreneurial function is exercised in ratifying or choosing the key men to hold various posts such as the one of production manager or sales manager.

A firm can only be individualized when it is shown that a definite entrepreneur, single or collective, presides over its destiny. A firm being, among other things a collection of resources, there should be an agent which gives a personality to these otherwise unspecific factors. There is a difference between labor as a general type of productive factor and labor hired and making its services available to a definite firm. The difference is that labor in general cannot be part of the economic plan of the firm, while specific labor can. We can then identify a firm only if we can recognize its own factors in the mass of factor services currently at work in the economic system. Only an entrepreneur can induce this identification of factors and make them cooperate in a definite way. The entrepreneur is the catalysor in the third method of achieving the division of labor.

In no case today do we reckon as an entrepreneurial function the providing of capital and the risk-bearing (49, p. 66). In such a framework not all entrepreneurs are innovators. Already Danhof (49, p. 23) has introduced the "imitative" entrepreneur, etc.

Who is the entrepreneur?

According to Schumpeter it was generally a tycoon. Today it is any individual or group of individuals who takes the type of decisions dealt with above. Yet, we exclude the economic units which are primarily households. It is for this

reason that the professional man, the individual peddler, the rentier playing the stock market, and the serf[51] are not entrepreneurs. Although, as said previously, parts of the theory of the firm can be applied to them.

This leaves the owner-manager of small businesses, the agriculturist if he must take "action in which the results are dependent upon his analysis of production, the market, and other relevant factors, and he is under no contractual or political obligation to accept advice or direction from anyone" (49, p. 23), and those who satisfy the requirements listed above for the big firms. Most people agree that in this perspective the stockholders and sometimes figure heads, for example, the president and other senior executives are excluded.

Schumpeter (49, p. 71) goes as far as to include the Department of Commerce as an entrepreneur and Granick[52] says that in some fields, the Soviet director of a plant can also be viewed as an entrepreneur.

Relation between entrepreneurship and the theory of the firm:

Some people have sweeping ideas on the subject and go

---

[51]Some people, for example Cole (50, p. 152 ff.) mention the serf-entrepreneur; but this serf corresponds to an artisan which is a firm in my definition; see appendix AA.

[52]Granick (91, p. 285). Also ibid., p. 5, he says "This study should provide strong evidence that the businessman or entrepreneur has survived, although in considerably altered from, in the highly centralized socialist economy of the Soviet Union."

as far as to identify the firm with the entrepreneur (209). In the conventional theory the economists' entrepreneur has not been cumbersome since he was identified with profit maximization. But if the entrepreneur is introduced as a function, with its idiosyncracies, for example, problems of motivation, information and communication, new theories such as decision theory and organization theories, must be integrated into the conventional one. This brings us to the crucial question: is it possible and useful to extend our conventional theory of the firm? Surely a theory of the firm without a "functional entrepreneur" cannot explain the birth, the permanence and the disappearance of the firms.

Historically, of all the deficiencies of the classicists, none was more damaging than the absence of an integrating force such as the entrepreneur. Nevertheless, the concept was not new in their times, even if it did not contain what we require of it today in a definition. For instance, we have the initial formulation in Cantillon. In Cantillon, an individual, i.e., a farmer, is seen buying services of factors of production and providing the necessary funds by selling the product. Besides being described as an "organizer" of production he is also seen taking "risks" in the process (185, p. 22).

Let us remark also that at that stage, and for some time afterward, the capitalist function was not differentiated from

the entrepreneurial function. Moreover, we digress from the "organizer approach" when we come to Smith. There the undertaker, as he called it, is primarily a provider of capital or an investor.[53] "The directing activity as a distinctive function played a surprisingly small role in his analytic scheme of the economic process," as mentioned previously, for the whole of classical economics, the "reader is bound to get an impression to the effect that this process runs on by itself."[54] In Cantillon, we had partially a functional approach: risk-taking. In classical economics, the function was simply capital providing. No wonder there were no attempts to separate the rewards, interests and profits, which were not differentiated in their analysis. In the meantime, in France the development of the theory was continuing. In the Cantillonian tradition, Say took over the concept and gave it a rejuvenation that will help keep it forever in the French exposition of economics. Say "was the first to assign to the entrepreneur--per se and as distinct from the capitalist--a definite position in the schema of economic process.

---

[53]"What the businessman does in the system of Adam Smith is therefore, to provide real capital and nothing else; the identification of the capitalist's and the entrepreneur's function was then accomplished." Schumpeter (184, p. 250).

[54]Schumpeter (184, pp. 249-50). Also "With Ricardo and Marx the process of production and commerce are still more automatic. The designing, directing, leading, coordinating functions has practically no place at all in their analytic schemata" ibid., p. 250.

His contribution is summed up in the pithy statement that the entrepreneur's function is to combine the factors of production into a producing organism" (185, p. 555). With Say, the entrepreneur became the important personage in economic theory of the firm in the sense that he was the bridge between production theory and distribution theory. He coordinated exchange theory with production theory by using supply and demand in hiring the factors and selling their product. He coordinated distribution theory within the firm by evaluating the factors and apportioning to each his share of the total production. The entrepreneur became the intermediary between factors of production and their product.[55]

Say was in a better position to give a more active and important role to the entrepreneur because he studied the economic process in the context of manufacturing industry[56] instead of doing it in agriculture as the classicists did.

---

[55]Gide et Rist: (90a, p. 124). Also: "L'entrepreneur sert the pivot à toute la répartition. Par la loi de l'offre et de la demande les entrepreneurs combinent les facteurs en vue de la production. Grâce à l'entrepreneur la valeur des produits se répartissent entre les 'divers services productifs', et les divers services se repartissent suivant les besoins des industries. La théorie de la distribution se coordonne harmonieusement à la théorie de l'échange et de la production ...Ce qui exerce la plus notable influence sur la distribution des richesses dit Say, c'est la capacité des entrepreneurs d'industries." Ibid., pp. 131-135.

[56]He was himself an industrialist from 1804 till 1813. He was the proprietor and creator of a cotton spinning factory at Auchy-les-Headins in the Pas the Calais region. He was also the first to give a course in 'Industrial Economics' at the Conservatory of Arts and Trade Paris. (see 90a, pp. 131-135).

He then pictured the entrepreneur as the main agent of economic progress, because he is the only one "actively" undertaking to combine the possible factors of production; capitalists, landlords, and labourers. Although Say appreciated the contributions of the entrepreneur either as the soul of the firm or as a fourth factor of production, however, it is not sure (185, p. 555) whether he saw all the analytical implications of such a concept. For instance, he probably did not differentiate between the managerial labour needed to run a successful concern, and the 'special labour' needed to start a new firm or modify the course of one already in business. The first type of labour is of the routine nature and is probably not much different in essence from any other kind of specialized labour; as such it is not a fourth factor of production but simply a variety of the third one: labour. In a stationary state, Say's entrepreneur has nothing very special to do and is similar to that of J. S. Mill "who brought the term entrepreneur into general use among English economists, and, in analyzing the entrepreneurial function, went from 'superintendence' to 'control' and even to 'directing' which he admitted, required 'often no ordinary skill'. But this defines the function of management and not anything distinct from mere administration" (185, p. 556).

Say, Cantillon and Mill also introduced risk-bearing as another function of the entrepreneur and as such there might be a justification for including it in a fourth category

of factors. This contention has been rejected by most people today, namely by Schumpeter and by Knight. Risk-bearing must be associated with the function of the capitalist, not with the entrepreneur.

After Say, the main body of economic doctrine settled for Mill's conception of the entrepreneurial function which was to prevail throughout the century (185, p. 557). His doctrine is discernable in his theory of distribution, especially in the share of national income received by his businessman. At the end of the classical period, the businessman received four type of rewards (185, p. 646):

(1) wages of management,

(2) a premium for risk-taking,

(3) an interest on the owned portion of the capital employed, and

(4) a fourth type (not in Mill but sometimes in Ricardo and Marx) of return of an essentially temporary nature, the one that is derived for a time from the first introduction into the economic process of a novel improvement such as a new machine, etc..

The first type is actually labour; the two others are capitalist returns, and only the fourth could be accepted as an entrepreneurial return. But for Mill and for most classical economists, the main income of the entrepreneur is an income from capital (151, p. 167), part of it made in interest and

the other part a profit to compensate risk-taking, the rate of profit being commensurate with the risk taken.

Summing up, in the classical system the firm has no entrepreneur, or it is an automatic one. Put in the terms above, it has only a manager for routine decisions. This type of firm is viable only in a static model and even there it is not explained why and how firms appear and disappear. However, the classical system having no business cycle and no drastic change in the production function, we may say that it did not need the entrepreneur of Schumpeter.

CHAPTER III

COURNOT

Chronologically speaking, Cournot belongs to the period 1776-1870. Yet he was noticed only in the following period.[1] However, even if his knowledge was not "used" by his contemporaries this does not preclude the fact that in 1837 the theory of the firm had attained quite a high degree of development, as will be shown in this chapter.

No doubt part of this neglect is due to the mathematical garb in which he clothed his reasoning. This accounts for the momentary neglect, but this does not explain why his "translation" into ordinary language (Principes de la théorie des richesses, 1863) met also with little success. Actually his highly abstract method of approach rebuked many of his readers. For instance Marshall said: "My confidence in Cournot as an economist was shaken when I found that his mathematics re I.R. led inevitably to things which do not exist and have no near relation to reality," (164, pp. 406-407). Similarly, his unrealistic case of duopoly awakened

---

[1] Irving Fisher in his preface to Cournot (54, p. xi) comments in these terms: "For nearly forty years the Principes Mathématiques was completely ignored by economists. The writers chiefly instrumental in reviving the work were Jevons, Walras, and Boccardo."

apparently no interest and only passing remarks for many years to come; for example, Bertrand devoted only one paragraph to Cournot some forty-five years later (185, p. 958).

Another feature of Cournot is that he built upon Say and other previous writers. Apparently enough elements were then available to permit the construction of a complete theory. The time was then ripe for a summation, implying that we can do without Mill. This spared him the trouble of giving long pioneering exposés on topics such as the entrepreneur and factors of production. In other words he assumes that Say has been read. Consequently, there is no separate section on the division of labor and laws of returns. All this is assumed. By not repeating others, but by building upon them, Cournot limits his exposition to his own contributions.

Finally, there are indications in the work of Cournot that the theory of the firm proper originates with him because he is the first to give a complete model, accounting for the vital aspects of the firm. The striking feature of Cournot's contribution is the representation of the firm as an economic model, that is as a "set of relations among a set of economic variables" (19, p. 19). These variables comprise demand and cost data. Furthermore these data are completely segregated from the general model of the whole economic system and thus gives to the analysis of the firm an existence of its own. Cournot's firm is completely

individualized. Not only can demand and costs be pinned down to a particular concern, but particular models of behaviour are definitely identified. This is true also for the maximizing process, which was totally absent or assumed as a matter of fact, but never described by the previous writers. Regarding the internal equilibrium of the firm, as an analytical model, many people (214, p.6) think that the analysis of Cournot has not as yet been surpassed.

Not only did he pioneer the model describing the internal equilibrium of the firm, but he also made a breakthrough in the field of external equilibrium. He was the first to set up in an analytical form the basic and most important morphological situations of the firm: monopoly, duopoly, oligopoly, and unlimited competition. Further, he first described three modes of behaviour: polypolistic price-fixing type A, polypolistic price-fixing type C, and polypolistic quantity adjuster.[2]

In this, he shows an interest which goes far beyond the restricted field of perfect competition contemplated by the classical model. Moreover, his approach is completely opposed to that of the classicists. For them, the normal way to attack the problem of external equilibrium was to situate the single firm among great numbers of others with negligible differentiation of their products. In this way

---

[2]The technical terms are explained below, pp. 85 ff.

the case could easily be handled by a model of quasi-perfect competition. They sometimes mentioned, but merely as a complement, a monopolistic case. Until Cournot, then, the poles of the analysis were known, but the attention was focussed upon the one of competition.

Cournot's approach is just the opposite. He begins by first studying the case of monopoly and then the one of duopoly, which he widens by increasing the number of firms. He ends with the analysis of pure competition which he calls unlimited competition. As will be seen below, it should not be said that Cournot reaches the case of pure competition simply by multiplying the number of firms. The truth is very different. His analysis has a break in the sense that there is no continuous passage from the case of oligopoly with a large number of firms and pure competition. He does not attain pure competition by indefinitely increasing the number of firms, but rather by changing the mode of behaviour of the firms. This approach (starting from monopoly) was to be neglected by most of the economists who followed him, and the method had to be rediscovered a hundred years later by Sraffra, Chamberlin, Mrs. Robinson and others. By choosing to begin his analysis with the monopolistic firm he reoriented the whole approach to the study of market morphologies. He was also successful in establishing the internal equilibrium of the firm, whatever its market morphology, through a single maximizing process: MC = MR. If Mrs. Robinson's dictum: the analysis of monopoly immediately swallows the

analysis of competition" (178, p. 5), is correct, then the theory of the firm has made quite a long detour through the competitive land of Marshall.

Cournot was also the first to introduce mathematics, on a large scale and with rigourousness and clarity, into the theory of the firm. He pioneered formal micro-analysis, and, to a certain extent, partial analysis when he studied demand.[3] Finally, he was aware of the effect of time.[4]

The following elements of the theory were established and related to each other by Cournot:

(1) Demand curves and cost data; all functionally related to the output of the firm,

(2) a behaviour hypothesis pertaining to the maximization process,

(3) the maximization mechanism of the firm, and

(4) a market framework for the operations of the firm.

In Cournot, this job is done simultaneously. Consequently, we will not try to put into different sections his contributions to market structures, and to the firm's internal

---

[3]Cournot (54, p. 52), where he is aware that during a certain period of time (one year) "the country may experience a movement of progress or decadence"; consequently a correction should be made for these movements to isolate the law of demand. He corrects the situation by using averages.

[4]Cournot (54, p. 45), he specifies that if we want to compare two prices, and their corresponding quantities, we must be careful to do this "in the same lapse of time and under the same circumstances."

maximization mechanism.

One way to illustrate most of the points we have alluded to above, and to a certain extent to prove our assertions regarding Cournot, is to study the demand curve for the firm established by Cournot.

Proceeding in a slightly different manner from that used in the preceding chapter, we will first establish a particular contribution of Cournot; and then, using this contribution, we will translate his formulation into a modern example. Consequently, two things are accomplished in a parallel way here:

(1) the differentiation of Cournot's contributions, and

(2) their translation into a model.

While doing this, we will simultaneously bring in the relevant material concerning market morphologies and modes of behaviour, and the maximization process of the firm.

A. DEMAND CURVES

Before drawing the demand curve for a firm, Cournot established the Law of Demand itself. This was a field badly neglected by his predecessors, so that he had to start practically anew. Briefly, he began by showing that the demand curve, either for a consumer or for a monopolist, slopes downward to the right, at least for ordinary goods. Yet he is aware of the upward sloping demand curve for "some objects of whim and luxury which are only desirable on account of

their rarity and of the high price which is the consequence thereof.... In this case a great fall in price would almost annihilate the demand." (54, p. 46)

Contrary to many other writers of the period he could not remain vague on the concept of demand because he needed precise relationships between price and quantity to use in his model of the firm. Actually, he created the theory of the demand (185, p. 960). This obliged him to distinguish between demand and quantity demanded, that is between a "vague desire of possession of the article" and quantity demanded at a "limited price". His next step was to consider the demand elasticity, a word he did not use but outlined in concept.[5]

However he does not "prove" that the demand curve for an ordinary good has to slope downward. This had to be established by Marshall through utility analysis. Cournot simply relies on common experience.[6] He conveniently expresses the relationship between quantity demand and the price in the form of function $D = F(p)$. (where D is quantity demanded; p is the price, and F "depends on the kind of utility of the

---

[5]"the price of violins or of astronomical telescopes might fall one half and yet probably the demand would not double..." Cournot (54, p. 46).

[6]"the sales or the demand generally, we say, increases when the prices decreases" (p. 46)..."Observation must therefore be depended on for furnishing the means of drawing up between proper limits a table or corresponding values of D and p".... (p. 47).

article, on the nature of services it can render or the enjoyment it can procure, on the habits and customs of the people, and the average wealth and scale on which wealth is distributed" (54, p. 47). This function is continuous if a great number of people are involved. Market demand curves are bound to be continuous because the idiosyncracies of different people cancel each other in this case (54, pp. 49-50).

In due course (p. 52 ff.) he introduces the total revenue function: $R = p.F(p)$.[7] He then shows that the value of the function is zero when p is zero and increases for a time as p increases. It reaches a maximum when the differential coefficient of the function F is equal to zero. In Lagrange's notation this is when:

$$F(p) + pF'(p) = 0.$$

All of us have recognized in this expression the marginal revenue function which will be rediscovered later.[8] We should say, however, that he does not draw the marginal revenue curve. In his figure 1 in his appendix, he shows geometrically only that if a straight line demand curve is involved, or a tangent to a curved demand curve, the maximum revenue is attained on the

---

[7]The letter R is not used by Cournot, but it is supplied here to take the place of the words he uses.

[8]"This reasoning in marginal terms will be completely forgotten (Marshall reasoned in total net revenue curve) to the extent that when Chamberlin and particularly Mrs. Robinson will introduce it, it struck economists as something of a novelty" (185, p. 977n).

demand curve at a point where a bissector from the origin meets this curve, a property which was to be widely used by Mrs. Robinson later. Finally (pp. 53-54) he puts forward the proposition that when the elasticity is greater than one, R increases and vice versa.[9]

As a result he provided the theory of the firm with three pieces of analysis: a total revenue curve which is bell shaped, an average revenue function whose curve slopes downward, and a marginal revenue function which reaches zero when R attains a maximum.

Cournot's firm's demand curves in various market morphologies:

After having established the shape of the market demand for a certain commodity, Cournot sets out to establish the various demand relationships that would be experienced by a firm, depending on whether it is a monopoly, duopoly, oligopoly, or unlimited competition prevails in the market.

(a) Monopoly case:

He was the first to give a complete analytical model of a monopoly. He distinguished between various types. For example the natural monopoly which comes from a natural

---

[9]His notation is a bit clumsy. For him in the case of a price increase if $\frac{\Delta p}{\Delta f} < \frac{D}{f}$, the R increases; if $\frac{\Delta D}{\Delta p} > \frac{D}{p}$ then R decreases. This is different from our present notation which writes the formula of elasticity as

$$e = \frac{p}{D} \cdot \frac{dD}{dp}$$

advantage (166, p. 16). This was important because apparently economists before him and for a long time after him had "very loose ideas of what a monopoly really is" (185, p. 602).

When there was only one firm in a particular market, Cournot did not need to make any modification on the market demand curve in order to use it as the monopolist's demand curve. In the case of a monopolist, Cournot uses, then, the following functions (p. 56): Demand is written as $D = F(p)$. The corresponding total revenue functions is: $p.F(p)$ and the marginal revenue function is $F(p) + pF'(p)$.

Revenue is maximum when $F(p) + p.F'(p) = 0$. Consequently, the optimum price is, where there are no cost of production: $D + p.\frac{dD}{dp} = 0$ (p. 82). With variable costs of production the monopolist maximizes profits by maximizing the function $pF(p) - \phi D$. (where ϕD is the cost function). With costs, then, the optimum price is

$$D + \frac{dD}{dp}\left[p - \frac{d\phi D}{dD}\right] = 0 \qquad \text{(p. 57)}$$

If the costs are constant the optimum price is

$$D + \frac{dD}{dp}\left[p - g\right] = 0 \qquad \text{(p. 60).}$$

(where g = constant costs).

In all cases $p > \phi'(D)$ (p. 59).

The above is the functional notation of Cournot. To show the degree of usefulness, we will translate Cournot's apparatus into a concrete example by using a demand function

such as: $p = a - bx$. This will provide us with our first model of the firm. Others will be provided in the course of this chapter as our tools of analysis become more complete.[10] Here, total revenue is: $xp = ax - bx^2$; the marginal revenue curve is: $\frac{d(xp)}{dx} = a - 2bx$. The optimum output is then $x = \frac{a}{2b}$ and the optimum price $a/2$. In the case of costs of production or, to simplify, constant variable costs, c, $x = \frac{a - c}{2b}$, and $p = \frac{a + c}{2}$.[11] (AD, 1)

The elasticity of the monopolist's demand curve is then $e = \frac{a - bx}{bx}$ (AD, 2).

(b) Duopoly Case:

Next Cournot undertook to study the case of duopoly, two firms in the same market. The commodity dealt with being homogeneous, the same price must be charged by both producers. If the market demand is $D = F(p)$, this can be written as $p = f(D)$ (p. 80). Furthermore, since there are only two producers, their respective outputs D_1 and D_2 must add up to the total output. In other words $D_1 + D_2 = D = F(p)$, (p. 79). From this we can obtain:

---

[10] We must note here that by using a straight line demand curve, we are not following Cournot closely because he used curved demand curves. However for the purpose of simplicity it is easier to use a linear equation.

[11] Hereafter all references in the form of (AD, ?) refer to a particular paragraph in the appendix D.

$D_1 = F(p) - D_2$ and $D_2 = F(p) - D_1$. Conversely:

$D_1 = f(D_1 + D_2) - D_2$ and $D_2 = f(D_1 + D_2) - D_1$

Their revenue functions are pD_1 and pD_2 (p. 79).

After it has been said that $D = F(p)$ can be written as $p = f(D)$ or $p = f(D_1 + D_2)$ the revenue functions become:

$$R_1 = D_1 . f(D_1 + D_2)$$

$$R_2 = D_2 . f(D_1 + D_2) \qquad \text{(p. 81)}$$

and their marginal revenue functions are:

$$f(D_1 + D_2) + D_1 f'(D_1 + D_2)$$

$$f(D_1 + D_2) + D_2 f'(D_1 + D_2) \qquad \text{(p. 81)}$$

Then Cournot postulated that proprietor No. 1 determines his optimum price or output by taking the output of proprietor No.2 as given because "proprietor No. 1 can have no direct influence on the determination of D_2"(p. 80).

In other words, proprietor No. 1 must take into consideration his direct influence on the ruling price but not the fact that he may have an indirect influence on the ruling price by altering the industry conditions in such a way that proprietor No. 2 is likely to readjust his output.[12]

---

[12]There might be some people who may not agree with this approach which stipulates that the firm in Cournot is a price adjuster. For instance Machlup (131, p. 372) says that in Cournot "the two sellers do not set or announce any prices." Also Shubik (194, p. 63) says: "Cournot assumes that the producers...compete through quantity adjustment." Also a similar statement by Fellner, (80, pp. 57-58). All this, while Cournot himself textually says: "by properly adjusting his price..." (54, p. 80). See also below pp. 89 ff. for citations from Cournot where price-fixing is clearly postulated. It is then a surprise to find that contemporary writings always picture Cournot's entrepreneur exclusively as a quantity adjuster; this comes probably from his reaction curves which utilize output axes.

At this point Cournot introduces a real innovation: in writing the marginal revenue functions as above and equating each to zero, he actually pictures how each firm will operate, that is, how it will change its price, or output, until its marginal revenue is equal to zero. By varying the output data of the other firm in each equation he shows how each firm will react to these changes. By exploring various values for D_1 and D_2, he can obtain a reaction curve for each firm (see figure 2 in his appendix). These reaction curves show how, for each value of D_2, the subjective demand curve for duopolist No. 1 can vary. However, Cournot shows that the system tends toward equilibrium. That is, there is a set of values, output of duopolist No. 1 and output of duopolist No. 2, where both producers will name prices that will bring compatible values which will bring a state of equilibrium such that "if either of the producers, mislead as to his true interest, leaves it temporarily, he will be brought back by a series of reactions, constantly declining in amplitude, and of which the dotted lines of the figure give a representation by their arrangement in steps" (p. 81).

The market equilibrium price is, then, where neither proprietor has any tendency to vary his action parameters further. This is obtained when the price is: $D - 2\,p.\frac{dD}{dp} = 0$ if there are no costs (p. 82). When variable costs are involved, each duopolist is in equilibrium when:

for No. 1 $f(D) + D_1 f'(D) - \phi_1'(D_1) = 0$

for No. 2 $f(D) + D_2 f'(D) - \phi_2'(D_2) = 0$ (p. 85)

The optimum price is where $D + \frac{dD}{dp}\left[np - \Sigma \phi_n'(D_n)\right] = 0$ (p. 86)

where n = 2 (the number of firms involved).

A small amount of preparatory work remains to be done before commencing to build our second model of the duopolist firm.

In our own notation we must start using the same market demand curve: $p = a - bx$.

Consequently the subjective demands for each duopolist become:

$p_1 = a - bx_2 - bx_1$ and $p_2 = a - bx_1 - bx_2$,

where in the first equation x_2 is a constant, while x_1 is a constant in the second equation. If we put $k_1 = a - bx_2$ and $k_2 = a - bx_1$ (where x_2 and x_1 are constants) the demand curves of the two duopolists can be rewritten as such:

$p_1 = k_1 - bx_1$ (for the duopolist No. 1)

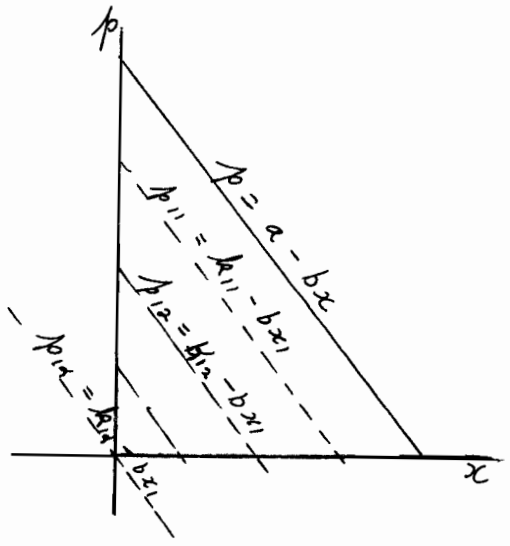
$p_2 = k_2 - bx_2$ (for the duopolist No. 2)

Introducing now the mode of behaviour postulated by Cournot, we can then infer that the individual demand curve for each duopolist has the same slope ($\underline{b}$) as the industry demand curve, whatever the output of his rival. However, the demand curve of a duopolist is not stable, in the sense that each different output of his rival will entail a different demand for himself. For instance, each increase in the output of duopolist No. 2 will shift to the left the demand curve of duopolist No. 1 and vice versa.

In the notation used above the expression: $p_1 = k_1 - bx_1$

is, in reality, a general formula for a whole family of demand curves for the firm No. 1, each demand curve being a parallel to another. In other words, the value of k_1 is likely to vary with each output (x_2)put on the market by firm No. 2. If the outputs of firm No. 2 are: x_{21}, x_{22}, x_{23}, etc...., then firm No. 1 will have the following demand curves: $p_{11} = k_{11} - bx_1$, $p_{12} = k_{12} - bx_1$, $p_{13} = k_{13} - bx_1$, etc.....

Because k takes the following values: $k_{11} = a - bx_{21}$; $k_{12} = a - bx_{22}$; $k_{13} - a - bx_{23}$, etc.... Now x_2 can take values only from 0 to a/b, consequently x_2 takes values from 0 to a/b, the value of k passes from a maximum (a) to move toward 0, (when $x_2 = a/b$). Graphically, the demand curve of firm No. 1 is identical with the market demand curve when $x_2 = 0$. Then as x_2 increases, it shifts parallel to itself to the left and finally passes by the origin when $x_2 = a/b$. In economic terms, this means that when $x_2 = a/b$, the demand curve of firm No. 1 disappears completely.



We are not considering the values in any quadrant except the first one, so that for $x_2 = a/b$ the demand curve for the firm No. 1 is reduced to the point of origin (0,0).

Yet one important feature must be noticed. As long

as his subjective demand curve exists, firm No. 1 feels, and is, the determinant of the ruling price. Let us see how:

When firm No. 2 puts x_{21} on the market ($0 < x_{21} < a/b$), the price is determined by firm No. 1 which equates: $k_{11} - 2bx_1 = 0$, or $x_1 = \frac{k_{11}}{2b}$; when this value of x_1 is put into the demand equation of firm No. 1 we obtain the price optimum for firm No. 1 (and ruling in the market) which is $p_1 = \frac{a - x_{21}}{2}$ (AD, 3). Actually, in the case of no costs, the firm No. 1 is changing its own output, with regard to changes in the output of firm No. 2, according to the equation of its reaction curve which is $x_1 = \frac{a - bx_2}{2b}$ (AD, 4). For instance it can be shown that if x_2 is continually increased by one unit, x_1 decreases steadily by 1/2 a unit and the price is also steadily decreased by 0.5b units (AD, 5).

The above formulas account for a no costs case. In the case of constant variables costs (c) the firm No. 1 will stop producing (although its demand curve continues to exist for a while) when the output put on the market by firm No. 2 is $x_2 = \frac{a - c}{b}$ (AD, 6).

The same reasoning of course applies to firm No. 2. The same set of relations, demand curves and reaction curves, can simultaneously be drawn for firm No. 2. This is also according to Cournot's behaviour hypothesis about any of the duopolists. Consequently, the demand curve of firm No. 2 would also disappear, in the case of no costs, when $x_1 = a/b$;

if firm No. 2 experiences constant marginal costs (c) it will stop production when $x_1 = \frac{a - c}{b}$. To sum up, each duopolist does not have one demand curve like $p_1 = k_1 - bx_1$, but a whole series of these parallel demand curves as the output of the other firm varies. This prompted Machlup (131, p. 335) to say that "there is no such thing as an oligopoly demand curve."

However, Cournot, through the use of his mode of behaviour hypothesis, is concerned with the subjective demand curve of a duopolist and this surely exists in the mind of the entrepreneur. We must remember that, in Cournot, the duopolist neglects his <u>indirect</u> influence on the price.[13] Consequently, each duopolist does not see his situation as a whole map of demand curves, but as a unique and stable demand curve, which he takes as given, without venturing to find a relation between shifts in his demand curve and his previous changes in output. However, each time he adjusts his price on his subjective demand curve, the only one he sees, this demand curve shifts.

---

[13]Showing that a duopolist takes into account his indirect influence on the price would amount to show that he incorporates in his calculations the eventual readjustment of output by duopolist No. 2; a readjustment occasioned by his own attempt to maximize profits. In other words his computations would include not only changes in quantity demanded, or vice versa changes in price on his subjective demand curve, but also the induced shifts in his demand curve. This latter type of behavior is ruled out by Cournot.

Yet, as implied before, he knows that he has a direct influence on the price, as pictured by any of his subjective sloping demand curves. When he assumes that the output of his competitor is constant at output x_2, he adjusts his price or output according to his marginal revenue curve and it is this price that will objectively prevail in the market. This is where Cournot's behaviour hypothesis leads to. Cournot was subsequently criticized for having ruled out "indirect effects", a type of behaviour which in the eyes of many people is the only logical one for a duopolist of the real world.[14] We do not have to discuss this point here, since we are interested in Cournot only as a model builder. As such we need only to test the internal logical consistency of his model, and realize that Cournot's analysis is simply an exercise in the theoretical implications of a mode of behaviour hypothesis. Whether or not it pictures the real world is beside the point in this context.

Continuing, we find that Cournot did not isolate himself from the objective data of the problem. Although he gratified each duopolist with a subjective stable demand, which shifts objectively, he did not assume that each duopolist always makes a correct guess of the rival's output.

---

[14]It might be interesting to take note of some contemporary appreciations of this hypothesis used by Cournot. Pirou (166, p. 184) labels it an historical curiosity, while Schumpeter (185, p. 959) says that his theories of oligopoly and bilateral monopoly will become the "whipping boys of a larger literature."

On the contrary, by the device of reaction curves, he showed that each duopolist is usually wrong. The reaction curves show the market effect of these moves and countermoves. Yet, by the same device Cournot is able to show that a stable and determinate price, and output, will eventually prevail. What to notice here is that seemingly uncoordinated data, subjective demands in themselves insufficient to picture adequately the objective conditions, yield, through their interactions, a determinate objective solution, a solution in terms of a stable and unique price and output.

In modern notation (using a market demand $p = a - bx$) the equilibrium price is $p = a/3$, the aggregate equilibrium output is $x = 2a/3b$, each firm's output being: $x_1 = x_2 = a/3b$ (AD, 7). The above is true in the case of no costs. When there are constant marginal costs the values become:

$$p = \frac{a - c}{3},\quad x = \frac{2(a - c)}{3b},\quad x_1 = x_2 = \frac{a - c}{3b},\ \text{(AD, 8)}.$$

The equilibrium demand curves of the firms are (AD, 8)

$$p_1 = \frac{2a + c}{3} - bx_1;\quad p_2 = \frac{2a + c}{3} - bx_2$$

(These demand curves are interdependent, but the above are compatible with equilibrium conditions; yet any experimentation with them will change them temporarily until they are eventually reestablished). The elasticity of the equilibrium demand curves are:

$$e_1 = \frac{2/3a - bx_1}{bx_1} \text{ and } e_2 = \frac{2/3a - bx_2}{bx_2} \quad \text{(AD, 9)}.$$

From this we can deduce the following proposition: when the market is in equilibrium, in the case of no costs, each firm is operating at the point of unit elasticity of its demand curve (AD, 10). In the case of constant marginal costs the firms would operate on their respective demand curves were

$$e_1 = e_2 = \frac{a + c}{a - c} \quad \text{(AD, 11)}$$

Since the subjective demand curves of the duopolists are straight lines sloping downward, they can experience elasticities of all values, zero to infinity, over their total length. It might now be interesting to notice that, although each duopolist operates at a point of unit elasticity on his subjective demand curve, in the case of no costs, at equilibrium, the ruling price and output do not warrant operating at the point of unit elasticity on the market demand curve (AD, 12). It should then be realized that when we go from a monopoly case to a duopoly case, with one more firm in the market, the equilibrium position on the market demand curve moves from a point of unit elasticity to one of a value of one half.[15]

(c) Oligopoly case:

Cournot moves directly from the duopoly case to the oligopoly case, small and great numbers, without changing the

---

[15]The reader may find this a little tedious, and may wonder if he is not wasting his time in reading these apparently too much detailed derivations. Yet we have a purpose, because these properties will be used extensively when we will study the oligopoly case. Especially, this apparatus will permit us to differ from some people, notably G. C. Archibald, (5) writing on the subject.

mode of behaviour and by simply saying: "if there were 3, 4,...n producers in competition, all their conditions being the same, equation (3) would be successively replaced by the following: $D + 3\,p.\frac{dD}{dp} = 0$, $D + 4\,\frac{p.dD}{dp} = 0$, ...

$D + np.\frac{dD}{dp} = 0$; and the value of p which results would diminish indefinitely with the indefinite increase in the number n" (p. 84). n here is the number of competitors. The marginal revenue functions are:

$$f(D) + D_1f'(D),\ f(D) + D_2f'(D),..f(D) + Dnf'(D),\ (p.\ 84).$$

Each firm will increase output till its marginal revenue equals zero which defines the equilibrium condition of the firm with no costs. When all the firms are in equilibrium, $D_1 = D_2 .... = D_n$ and $D_1 + D_2 + D_3 + ...D_n = D$, (p. 85). With variable costs of production equilibrium means that for each firm to be in equilibrium:

$$f(D) + D_1f'(D) - \phi_1'D_1 = 0 \quad ; \quad f(D) + D_2f'(D) - \phi_2'D_2 = 0;$$

$$..... \quad ..... \quad ....= 0 \quad ; \quad f(D) + D_nf'(D) - \phi_n'D_n = 0$$

In this case the market price toward which the system tends is: $D + \left[\frac{dD}{dp}\left[np - \sum \phi_n'(D_n)\right]\right] = 0$. (p. 86) where n tends toward infinity.[16] Cournot goes on after p. 87 ff. to prove that the price arrived at by the monopolist in the case where there are

---

[16]For any firm k, the equilibrium price is: $p - \phi_k'(D_k) = 0$; in other words the price equals the marginal costs, while in the case of the monopolist we found previously that $p > \phi'(D)$, the price is greater than the marginal costs. Cournot, (54, p. 59).

costs is always greater than the one that will prevail in the case of oligopoly.

It is now easy to translate Cournot's oligopoly model into modern notation. Since the same mode of behaviour is postulated as in the case of duopoly, it amounts merely to adding more equations to the model.

The market demand curve being $p = a - bx$, in the case of n firms it can be rewritten as: $p = a - b(x_1 + x_2 + x_3 + \ldots x_n)$. Consequently the demand curve of each firm is now:

$p_1 = a - b(x_2 + x_3 + \ldots\ldots x_n) - bx_1$

$p_2 = a - b(x_1 + x_3 + \ldots\ldots x_n) - bx_2$

. . . . . $p_n = a - b(x_1 + x_2 + \ldots. x_{n-1}) - bx_n$

As before we can write: $k_1 = a - b(x_2 + x_3 + \ldots x_n)$ for firm No. 1,

$k_2 = a - b(x_1 + x_3 \ldots\ldots x_n)$ for firm No. 2,

.. ..

$k_n = a - b(x_1 + x_2 \ldots\ldots x_{n-1})$ for firm No. n.

Consequently the subjective demand curves of the firms can be rewritten:

as: $p_1 = k_1 - bx_1$; $p_2 = k_2 - bx_2$: .....; $p_n = k_n - bx_n$

Now, objectively speaking, the demand curve of any firm (e.g. $p_1 = k_1 - bx_1$) is not unique since it is a function of all the other firms' outputs. Consequently there is here, as in the case of duopoly, a whole family of p_1 demand curves for firm No. 1 and, similarly, for all the other firms. Each demand curve is parallel to the others of the same family and

to the market demand curve. There is a demand curve for firm No. 1 for each value of k_1. If for instance $(x_2 + x_3 + \ldots x_n) = 0$, the demand curve of firm No. 1 becomes $p_1 = a - bx$ (the market demand curve). Yet this demand curve will shift to the left as k_1 decreases, that is, when the other firms increase their outputs. Yet the new demand curves for firm No. 1 will have the same slope, because b does not change, and k_1 is taken as constant by firm No.1. Therefore, the family of demand curves for firm No. 1 (oligopoly) looks like the one of page 71, (a case of duopoly). As before, the value of $(x_2 + x_3 \ldots x_4 + \ldots x_n)$ can take any value from o to a/b (when there are no costs of production). When $(x_2 + x_3 + \ldots x_n)$is zero, the demand curve of firm No. 1 is identical to the market demand; when $(x_2 + x_3 + \ldots x_n)$ is equal to a/b, the No. 1 firm's demand curve disappears completely from the first quadrant. In the case of costs of production (constant marginal costs c) the firm No. 1 drops out of the market when $(x_2 + x_3 + \ldots x_n) = \frac{a - c}{b}$. For obvious reasons, the proof of these statements is identical to the one supplied previously in the case of duopolists. Here, as in the case of duopoly, each oligopolist firm considers his subjective demand curve as stable, because, as a member of this group, he neglects his indirect influence on the price, although his demand curve will shift each time he will readjust his output to new conditions. However, as far as he is concerned, he believes that he has a definite

influence on the ruling price as depicted by his sloping demand curve. Furthermore, even with many firms, or more firms, in the market he does not feel that his influence on the market price has diminished. The same formula, as devised on page 72, can be used here. Each time his competitors as a group increase their output by one unit, in total, he feels that he will have to change the ruling price by .5b, and his output by one-half unit.

This covers the subjective field, but the question remains whether an objective (observable) solution can be realized. The answer is yes, because Cournot gratifies each oligopolist with the same mode of behaviour used in the case of duopoly. Under this premise, the system reaches a stable objective solution in terms of price and output, (market price, aggregate output); and each firm has a stable demand curve.

The various values obtained, in our notation, are as follows: In equilibrium the equilibrium output x put on the market by all the firms will be:

$$x = \frac{n(a)}{(n+1)b} \quad \text{(AD, 13)}$$

The equilibrium market price will be $p = \frac{a}{n+1}$ (AD, 14)

Here, as in the case of duopoly, subjective data through its interaction leads to an objective solution from which no oligopolist will have any inducement to move away. Each oligopolist will have only one subjective sloping demand curve

which will be stable and determinate, and where the value of k will be objectively determined. Even if he does not have any inducement to move from his position of equilibrium so determined, if he moves, his demand curve will immediately shift because of the readjustment of the others. Through a process of interaction between the firms, the whole system of demand curves will move back toward the former position so that he will end up with his original demand curve and will readjust his output accordingly.

(i) The no cost case:

In this case the equilibrium output of the firm will be $x_1 = \frac{a}{(n+1)b}$ (AD, 15).

The equilibrium demand curve for firm No. 1 is

$$p_1 = \frac{2a}{n+1} - bx_1 \quad \text{(AD, 16)},$$

its marginal revenue curve is consequently: $MR_1 = \frac{2a}{n+1} - 2bx_1$, (AD, 17), the elasticity of the demand curve of firm No. 1 is $e_1 = + \frac{2a - bx_1(n+1)}{(n+1)(+bx_1)}$, (AD, 18).

It should be noted here that the firm, when maximizing profits, will operate on its demand curve where the elasticity is unity (AD, 18). The elasticity of the market demand curve is $e = \frac{a - bx}{bx}$, (AD, 18) in equilibrium. The value of the market elasticity will be $\frac{1}{n}$, (AD, 18) or, in other words, the market elasticity tends toward zero when the number of firms

in the industry tends toward infinity.

(ii) The case of constant costs (marginal or variable) c:

In this case the critical values are as follows:

the equilibrium aggregate output will be

$$x = \frac{n(a - c)}{(n + 1)b}, \text{ (AD, 19)},$$

the market price (equilibrium) will be

$$p = \frac{a + cn}{n + 1}, \text{ (AD, 20)},$$

the equilibrium output of each firm will be

$$x_1 = \frac{a - c}{(n+1)b}, \text{ (AD, 21)},$$

the equilibrium demand curve of each firm is

$$p_1 = \frac{2a + c(n-1)}{n + 1} - bx_1, \text{(AD, 22)},$$

the equilibrium marginal revenue curve is

$$MR_1 = \frac{2a + c(n-1)}{n + 1} - 2bx_1, \text{ (AD, 23)}.$$

Unfortunately this display was necessary, in the case of no costs, to establish the following.

In the case of oligopoly, as the number of firms tends toward infinity, the aggregate output tends toward a finite quantity $\frac{a - c}{b}$, and the price tends toward c. The elasticity of the point on their demand curve at which these firms operate tends toward infinity, while the elasticity on the market demand curve tends toward c/bx. The elasticity of the demand curve of each firm is:

$$e_1 = \frac{2a + c(n-1)}{+ bx_1(n+1)} - 1, \text{(AD, 24)}.$$

The elasticity of the market demand curve is as always

$$e = \frac{a - bx}{bx} \quad \text{(AD, 24)}.$$

Cournot's contribution to the types of modes of behavior of the firm:

Before passing to Cournot's case of Unlimited Competition, we must make a small digression into the concepts of morphology of the market and modes of behaviour of the firm. Otherwise we might miss an important feature of Cournot's analysis.

By morphology of market we mean simply the distinction in terms of the number of firms and the homogeneity of the product. The morphology is a question of the structure of the market.

By mode of behaviour we mean the psychological attitudes or state of mind (131, p. 91) of the managers which determine what action parameters should be used when taking decisions pertaining to the economic plan of the firm.[17] Most people distinguish two main types and many sub-types of modes of behaviour (188, pp. 54-55). The two main types are polypolistic behaviour and oligopolistic behaviour.

---

[17] This section is inspired from Machlup (131, Chap. 4); Schneider, (188, pp. 51-67), and R. Frisch, (87, p. 31).

(1) Polypolistic behaviour

A firm behaves polypolistically when in formulating his plans "there is absence of rivalry or rival-consciousness" (131, p. 85). "The polypolist is a seller who is unconcerned about rivals' reactions..." (131, p. 86). A firm acts polypolistically when it takes the price, output, location, or any of the action-parameters of the other firms, as constant. It is simply a case of non-reaction on the part of the surrounding firms. Chamberlin would say that this is a case where interdependence, if it exists, is not recognized, that is, where any indirect influence on the market price or output is not recognized. This has, essentially, nothing to do with the question of numbers although most people (87; 131, pp. 85-86) associate that type of behaviour with the conditions that will be likely to prevail when a large number of competitors are in contact. However, the use of the term cannot be restricted to large groups. Schneider (188, p. 62), for example, talks about a monopolist behaving polypolistically.

(2) Oligopolistic behaviour

Here the seller is conscious of rival's reaction (131, p. 97). This is the case where the firms expect reactions of the part of the surrounding competitors. This is a conjectural mode of behaviour. In this case, interdependence is recognized in the sense that the oligopolist will try to appreciate not only his direct influence on the price in the whole market, but also his indirect influence. His

sales depends upon:

(a) his own action parameters,

(b) others' action parameters (which are expected to vary as he varies his).

If we restrict the action parameters to the variable price, his price-sales function is

$$x_1 = f(p_1, p_2, p_3, \ldots\ldots p_n) \text{ where}$$

$$(p_2, p_3, \ldots\ldots p_n) = g(p_1)$$

Yet, these general types of behaviour, polypolistic and oligopolistic, by the very fact of their general formulation, are not precise enough to be used analytically without being further differentiated. That is, in order to make them fit the various modes of behaviour of the firms of the world, we have to distinguish sub-types or species. For instance in the case of a polypolistic situation, we may distinguish the following modes of behaviour:

(1) Polypolistic quantity-adjuster

Here the firm deals in an homogeneous product and feels it is too small to have any influence whatsoever on the market price, the price of its rivals, or their outputs. It can then ignore any reaction. It consequently thinks it can sell any amount at the ruling price. The market price is then, of course, held as constant. Its price-sales-function is:

$p = k$. Consequently x_1 can take any value.

Its key indicator is[18] the market price.

Its action parameter: its own output.

(2) Polypolistic price-fixers

Type A. Besides having all the psychological dispositions pertaining to this general mode of behaviour, the market situation is such that the particular mode of behaviour of the firm in this situation is modified by the fact that his sales now depends upon:

(i) his own price,

(ii) the output of the others (which is held as constant).

His price-sales function is then:

$x_1 = f(p_1, q_2, q_3 \ldots q_n)$ where:

$(q_2, q_3, \ldots q_n)$ are a constant.

The key indicator is output of others.

The action parameter: his own price or the market price depending upon whether or not competition is homogeneous. If the commodity dealt with is undifferentiated, all the firms must charge the same price. Consequently the firm of Type A in such a situation realizes that it cannot change its price without simultaneously changing the market price. In other words the firm which is active acts on the market price.

---

[18] By key indicator we mean the main external data as far as the other firms are concerned, which is taken as given by the firm who has to take a decision as to which action-parameter should be used and to what extent.

Type B. Here the sales of the firm depends upon:

(1) his own price,

(2) the prices charged by others (which are held constant).

His price sales functions is then:

$$x_1 = f(p_1, p_2 \ldots p_n) \text{ where:}$$

$$(p_2, p_3 \ldots p_n) \text{ are constant.}$$

The key indicator is: the prices of the rival(s).

The action parameter is: his own f.o.b. price.

This type of behaviour makes sense only in heterogeneous competition. Otherwise, different prices could not last. The firm, then, does not recognize any direct or indirect influence on the market price.

Type C. This is the case of a polypolistic monopoly. This is not a contradiction of the terms as long as we restrict our considerations to the mode of behaviour.[19] The monopolist, morphologically speaking, is surely not conscious of rivalry; that is, he is unconcerned about rival reactions. It is in this sense that we say that a monopdist can behave polypolistically. By definition then his sales depend exclusively upon his own action parameters. His price sales function can either be written as:

---

[19]See Schneider, (188, p. 64) for the distinction between the morphology of the market and the mode of behaviour.

$x_1 = f(p_1)$ or as $p_1 = h(x_1)$

He does not use any key indicator and his action parameters is either the price or the output.

Some people might think that these distinctions are solely the subject of a semantic argument over whether or not type A is "analytically" different from type B. This is something the reader must decide for himself. It is agreed that the class of concept called "polypolistic" possesses as sub-classes both type A and type B, but it should be remembered that following this line of reasoning will lead us to say that there are no "analytical" differences between sub-classes of concepts. For instance, since man and flower are both members of the same class, living things, we would then be obliged to say that there is no analytical distinction between a man and a flower. I would imagine that the psychologist might find one.

We hold, then, that when a concept covers ground which another one cannot cover, these two concepts are different. For instance, in type A, the key indicator is the output, while in type B the main variable is price.

Summing up, we think that we may have four modes of polypolistic behaviour (polypolistic quantity adjuster, polypolistic price fixers, type A, type B, and type C) which are analytically different because they function differently and use different variables.

With this equipment we can then review the various

cases treated by Cournot and realize that, in fact, he used many modes of behaviour in various morphologies. The first form of market contemplated by Cournot was the case of one firm. Consequently, we have a monopolistic morphology, He then studied it by using a polypolistic mode of behaviour price-fixer type C. In Cournot there is no indication that the monopolist is in any way preoccupied with would-be competitors. He is not watching, neither has he the feeling of being watched. The monopolist is either the sole proprietor of "a mineral spring...(which)...possesses salutary properties possessed by no other" (p. 56) or he "possesses the secret of a medical preparation or an artificial mineral water" (p. 57), or of a "theatrical enterprise" or of a "bridge" (p. 61). In all these situations the proprietor is a price-fixer: "He could doubtless fix the price of a liter of this water at 100 francs" (p. 56).

The second morphology of the market considered by Cournot is the one of two or many firms selling an homogeneous product, the duopoly (oligopoly) case. The duopolist of Cournot is a polypolist price-fixer type A: Firstly, the product dealt with is perfectly homogeneous. Cournot talks about "two springs of which the qualities are identical... which on account of their similar position supply the same market in competition" (p. 79).

Secondly, price fixing polypolistic type A is postulated when he says that "each of them independently will seek

to make his income as large as possible" (p. 79)..."This he will be able to accomplish by properly adjusting his price ..." (p. 80). Thirdly, the output distinction comes when he says that: "D_2 (are) the sales from the spring (2), then $D_1 + D_2 = D$". He follows by: "Proprietor (1), can have no direct influence on the determination of D_2; all that he can do, when D_2 has been determined by proprietor (2), is to choose for D_1, the value which is best for him" (p. 80). This is done by "properly adjusting the price". He does this under the assumption that proprietor (2) will change his price immediately to match his own price: "proprietor (2)... seeing himself (is) forced to accept this price and this value of D_1..." (p. 80). Consequently duopolist (1) is a market price fixer in the sense that he really thinks he can influence the ruling price.

This behaviour is extended to a case of many firms: "If there were 3, 4, ...n producers in competition, all their conditions being the same, etc..." (p. 84). We must note in passing that Cournot does not link a particular type of behaviour (type A) with the question of numbers. Here two firms or n firms (n being able to stand for any value including infinity) can use the same mode of behaviour.

What is the effect of all this in terms of the models? The answer is that as long as Cournot used the same type of behaviour, for example polypolistic price fixer type A, the number of firms does not matter. In other words, in Cournot

there is no analytical difference between a small number case and a large number case. The only effect of increasing numbers is to increase the aggregate output put on the market and to reduce the market price.

In this respect, it seems that in the recent literature Cournot is a bit misrepresented:

Firstly, picturing Cournot's duopolist (or oligopolist) exclusively as a simple quantity adjuster, is false.[20] The initial change in Cournot (p. 80) by firm No. 1 is in terms of "adjusting his price". It is true also that firm No. 2 will react to "this price (new) and this value of D_1 (by) adopting a new value for D_2, more favourable to his interests" (p. 80). The sequence is as follows: Firm No. 1 adjusts the price, thus producing a new output D_1. Firm No. 2 reacts by drawing a new demand curve (taking the new D_1 as constant) and choosing a new price that will maximize its revenues. And the reactions go on until they both name compatible values. This shows that even if the output connotation is present, the Cournotian firm can be pictured as a price-fixer, instead of as a quantity-adjuster as the current literature does.

Secondly, nobody mentions his contribution to the catalogue of modes of behaviour of the firm. He is usually studied exclusively from the market structure angle.

---

[20]See for instance Shubik (194, p. 63) and all the people mentioned previously in footnote (12).

Thirdly, people[21] are of the impression that Cournot finally obtains the solution of pure and perfect competition simply by increasing the number of firms. For instance, G. C. Archibald, (5, p. 108), probably mislead by Mrs. Robinson's definition of a perfectly competitive case exclusively in terms of infinite elasticity, proposes to show that in the Cournot duopoly case, without changing the mode of behaviour, the elasticity of the demand curve "can be made to approach infinity" and that, consequently, the competitive case is attained. As will be seen below, we must object strongly to this view. For Cournot, attaining the competitive case is not simply making the firm operate at a point of infinite elasticity on its demand curve, because Cournot, as opposed to Mrs. Robinson (178, p. 18), maintains that infinite elasticity is not necessarily the case of perfect competition. But it is a question of changing the mode of behaviour of the firms.

---

[21] Pirou (166, p. 153): "L'originalité de la présentation de Cournot à cet égard c'est qu'il passe de l'hypothèse du monople à l'hypothèse de concurence en envisageant une situation intermédiaire qui est une sorte de dégradation du monopole." Also Hutchison, (110, p. 308) where competition in Cournot is pictured simply as an "extreme limiting case". Also Schumpeter (185, p. 981): "We know already why Cournot neglected all this: in his brief sketch of the theory of pricing he evidently wished, starting from pure monopoly to follow an unbroken line of reasoning that would lead up to the case of pure (illimited) competition without having to vary anything: except the number of competitors."

To realize how different Cournot is from what people think he is, let us study the argument of Archibald in this connection. This will be an occasion to use the apparatus built previously pp. 83 ff.

Mr. Archibald's goal is to show that in Cournot's model, by simply increasing the number of firms, without changing the mode of behaviour, we can pass from a case of finite elasticity to a case of infinite elasticity. For him, this means that we can move continuously from duopoly to oligopoly to perfect competition without a break.

Mathematically speaking this is true only in a particular case, the one of positive constant marginal costs. In other cases, for example, the case of no cost, this is not true. Taking the illustration used by Cournot, mineral springs operating at no cost, it can be shown that in such a case the firm will <u>always</u> operate under new entry (although its demand curve is shifting gradually to the left) at a point on its demand curve where the elasticity (e_a) is <u>one</u> (AD, 18). This is true as long as its demand curve is not wiped out of the first quadrant by the entry of new firms.

In the notation of Mr. Archibald (where n = elasticity), with no cost of production, sufficient entry will gradually reduce the market price to zero. Consequently, at the limit, the market price is zero, but so also is the elasticity on the market demand curve, that is $n = 0$. The same happens to the share (S_a)of the market of any firm which also tends to

zero. At the limit: $n_a = \frac{n}{S_a} = \frac{0}{0}$. Consequently, n_a (the elasticity of the demand curve of each firm) does not tend toward infinity, as Mr. Archibald proposes, but toward an indeterminate value. The result in the case of Cournot, is that new entry not only does not change the slope of the firm's demand curve, but also it might not (in the case of no cost) even change the elasticity at which the firm will operate, even if there are an infinity of firms in the market.

Secondly, in the case contemplated by Mr. Archibald (a case of short-run and constant MC), this type of reasoning leads to a curious situation: When there is a sufficient number of firms, the elasticity will increase until it becomes infinite. This is illegitimate on two grounds:

(1) We cannot speak of the elasticity being infinite at point k (in the diagram Fig. 1 below). Because, at that point q = 0 and "an attempt to evaluate the fraction P/Q at that point forces us to commit the sin of dividing by zero...this ...is immoral," (18, p. 144).

(2) Secondly, at that point one of the fundamental equations of Cournot is not fulfilled: Cournot held that: $D_1 + D_2 + D_3 + \ldots + D_n = D$ where D_1 is the sales of firm No. 1 etc. and D the total quantity that should be put on the market by all the firms as read on the market demand curve.

However, when the firms are operating at $n_a = \alpha$ they are in the situation depicted by Fig. 1. Here entry of new firms has driven the market price down to k. It is a no profit situation, but also a zero output situation for any firm. Consequently here $D_1 = D_2 = \ldots D_n = 0$. Yet the market demand curve, (see Fig. 2) warrants an aggregate output of D = OQ at price k. How then, can the sum of individual outputs equal the required aggregate output D = OQ, if $D_1 = D_2 = \ldots D_n = 0$? It is obvious that this type of reasoning leads to an absurdity. This is probably why Cournot never attempted to reach the solution of the perfectly competitive market simply by increasing the number of firms. Instead, as will be seen below he changed the mode of behaviour.

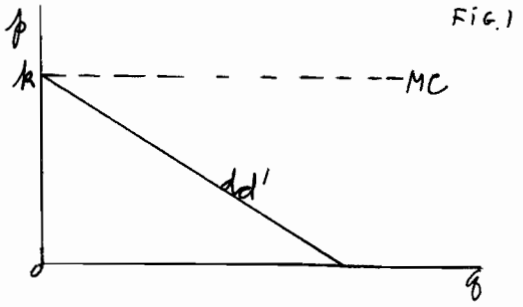


Fig. 1

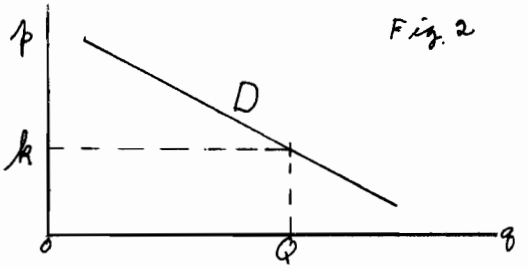


Fig. 2

Unlimited Competition

Finally, without changing the morphology of the market used in the previous case (oligopoly large numbers, dealing in an homogeneous commodity), Cournot attained another form of competition which he called unlimited competition, and which corresponds to our present pure and perfect competition

case. As said previously, Cournot attained this case not by changing the number of firms, moving from small numbers to large numbers, but by changing the mode of behaviour of the firms, from a price fixing type A to the quantity adjuster type. Here, the firms reason as if the price of the competitors, and also the ruling price in the market, is constant. That is, the firm does not think it has an influence on the price. "The effects of competition have reached their limit, when each of the partial productions D_k is inappreciable...so that the partial production D_k could be subtracted from D without any appreciable variation resulting in the price of the commodity" (p. 90).[22] The firm maximizes profits here when it equates P to MC or in Cournot's notation, where:

$$"p - \phi_k' D_k = 0" \text{ (p. 90)}$$

The implications of this method of approach are thus:

There is no need to have an infinity of firms to obtain a price zero in the case of no cost, because in this case, even if the price is zero the firms will continue to put their output on the market as they did previously. Let us take the example of the mineral spring of Cournot. In the case of the mode of behaviour type A (price-fixing), if the firm had a fixed production of j units (the mineral spring

---

[22] Schumpeter (185, p. 545) footnote (2) correctly says that the authors of the 1850s used the hypothesis of competition without ever defining it, the important exception being Cournot who did it adequately.

flows, whether it is used or not), it put only x_1 units on the market and diverted the difference between j and x_1 to waste. It would put on the market only the required amount necessary to maximize its profits. When the price went to zero, nothing was put on the market. But in the case of unlimited competition (and the mode of behaviour that brings it) the firm has no inducement to restrict the amounts put on the market. Even when the price is zero, there is no inducement to restrict output. It can be noted now, that using the proper mode of behaviour does not involve us in the problem which baffled Archibald. We may conclude this section by summarizing Cournot's position as follows: Cournot used two distinct modes of behaviour; price fixing type A and quantity adjuster, but he does not pass from one to the other simply by altering the number of firms. Many historians[23] were, then, quite wrong in assuming that Cournot started with one firm, (a case of monopoly) and then by gradually increasing the number of firms he could successively take care of duopoly, oligoply, and finally pure competition. Actually Cournot's evolution is quite different. He used two modes of behaviour so that there is an analytical break between his analysis of oligopoly (large numbers) and unlimited competition (also large numbers). To obtain the same numerical solution, for example, to have the market price reach

---

[23] See footnote (21).

zero (in the case of no cost of production) he needs an infinite number of firms in the type A analysis, and only a relatively great number of firms in the case of quantity adjuster. In other words Cournot does not need an infinite number of firms to attain a market price of zero in the case of unlimited competition but only a different mode of behaviour. His model is thus.

Competition between firms will reduce the market price to: $p - \phi_k'(D_k) = 0$ (in the case of positive marginal costs). (p. 90). Each firm will be in equilibrium when

$$p - \phi_1'(D_1) = 0,\ p - \phi_2'(D_2) = 0,\ \ldots\ p - \phi_n'(D_n) = 0,$$

in such a way that $D_1 + D_2 + \ldots D_n = D$ (p. 91)

A complete contribution should include not only the concept of a group, which Cournot undoubtedly has but also the mechanism of equilibrium of the group, or industry. In other words the equilibrium should not only be worked out in terms of market output and price, but also in terms of entry and exit of firms. Unfortunately there is not much of this in Cournot. The only trace of this mechanism that can be found in one of his works (166, pp. 158-160) is where he studies a case of colonization. Briefly he says that price movements will determine the level of rent received by landlords and this will determine whether or not they will practice afforestation or deforestation. In other words, landlords go in or out of the wood business according to the price of wood.

There is also practically no trace of the case of monopolistic competition except in his chapter X where he speaks about the possible communication between markets (spatial competition). He says clearly "that an article capable of transportation must flow from the market where its value is less to the market where its value is greater, until this difference in value, from one market to the other, represents no more than the cost of transportation" (p. 117). Symbolically to obtain equilibrium we need:

"$\phi a\ (pa) + \phi b\ (pa' + E) = F_a(p_{a'}) + F_b(p_{a'} + E)$

when pa' represents the price in the exporting market A and E the cost of transportation from A to B". (p. 117).

To complete the catalogue of morphologies contemplated by Cournot, one should also mention a case of bilateral oligopoly (in the case of isolated exchange (185, p. 983) where correctly, but not as elegantly as Edgeworth was to do later, he came to an indeterminate solution.

Yet he did not explore the practice of price discrimination that a monopolist might adopt in certain circumstances (185, p. 978). In the same order of ideas, he considered one strategy in monopoly behaviour: where a firm, being in a situation of lower costs than its rivals (because of increasing returns through increase in size etc.) takes the opportunity to eliminate its rivals simply by growing faster than they; and thus implement a monopoly. This is a striking anticipation of Sraffra who, ninety years later, was to

crystallize the proposition that competition is incompatible with increasing returns.

B. COST CURVES

Having established on solid grounds the revenue side, (in this he was a real revolutionary) he sets out to streamline the tools of his predecessors on the side of costs.

As seen in the various equations of the equilibrium of the firm, Cournot recognizes that each firm has costs functionally related to output. For example, he says, "ϕD denotes the cost of making a number of liters..." (p. 57). Marginal costs are also used in the form: "$\frac{d(\phi(D))}{dD}$" (p. 57) or as $\phi'(D)$ p. 59. As said previously, "the function $\phi'(D)$ is capable of increasing or decreasing as D increases, according to the nature of the producing forces and of the articles produced" (p. 59). In the next paragraph he states that, at least in the case of manufactured articles, there might be initially a range of increasing returns (diminishing average costs), then a range of increasing costs. Thus, he has a U shaped cost curve (185, p. 976; 166, p. 145). He also envisages the firm under the hypothesis of constant marginal costs (p. 60) and in chapter IX of joint costs.

With this formidable analytical equipment Cournot was ready to picture the theory of the firm in action. All he needed was an hypothesis as to the goal setting and method of achievement of the firm. Simply, he postulated the hypothesis that the firm strives for the greatest possible profit

(166, p. 146). This was all the more easy because the firm in Cournot, as in the classics, is still owned and managed by a single proprietor. The apparatus of maximization is the same as the one used in the traditional theory of the firm today. A revenue function (marginal revenue) is confronted with the marginal cost function and he derives (185, p. 977) "the theorem, now so familiar to every beginner, that (instantaneous) gain will be maximized if the monopolist sets a price for which marginal revenue equals marginal costs." This is called the Cournot point by Stackelberg (201, p. 175).

The above accounts for the contribution of Cournot to the main body of the theory of the firm. Besides this he contributed some frills like the study of the effect of taxation of a monopolistic firm. With his apparatus he easily discovered that a tax per unit (because it raises the marginal cost) will have the effect of raising the optimum price. He pursues further the analysis of the effect of a tax in the case of a continually sloping downward cost curve, in which case the price will rise for two reasons:

(1) partial shift of the tax to the consumer,

(2) reduced output means higher unit cost of production.

In the case of an ever increasing cost curve the final results is the algebraic sum of (1) and (2) where (2) works now in the opposite direction of (1). This is so because when output is reduced the unit cost of production is also reduced.

In the case of an income tax, he correctly says (p. 68) that the firm will have no inducement to change either its price or output.

As far as the theory of the firm is concerned, Cournot develops the first precise model of the firm. Yet this theory of the firm is only a sub-theory of a master price theory, although Cournot proceeds in the inverse direction by establishing a price theory through the firm and not the reverse. This can be verified by ascertaining that the first industry model of Cournot is one of a monopoly.

The firm of Cournot is also exclusively a mental construction and as such has only a remote resemblance to the real life firm, although Cournot studs his exposé with a quantity of references taken from real life situations. The real purpose of Cournot is to show how the market price and output are established.

Therefore Cournot's model has also no predictive value except if, by coincidence, a number of conditions happens to be fulfilled. This is easily understood. Cournot's model is a mechanistic construction involving (1) an internal behaviour hypothesis; profit maximization, and (2) an imposed mode of behaviour describing the adjustments of the firm to external stimuli. His firm is a logical construction which may or may not be logically consistent but certainly not necessarily relevant. Obviously, the outcome predicted by the model will be the one of an actual situation only if firms

do maximize profits, utilize marginal analysis and are endowed with that particular psychology described by the imposed mode of behaviour. Cournot did not establish the various postulates he uses in any empirical manner.[24] All of them are products from his imagination. In other words, Cournot did not predict that the duopolist will always make computations under the hypothesis that the rival will not change his output, but established a tendency for the market price to move toward a particular value if the participants behave according to the hypothesis.[25]

The predictive value of the Cournot model is also reduced even in the case of monopoly because of the great simplifications of his model which he does not account for. These are (1) price discrimination, (2) goals setting within the firm, (3) manipulation of the market demand curve through advertising, changes in location, and product differentiation. In general his theory is static, mainly because of the automatic entrepreneur which he used all along. Good theories require other criteria than good predictions, but we postpone until the last chapter (section I) the elaboration of that point.

---

[24]The method of reasoning in Cournot is always: "suppose this or that", then...such and such a thing will follow." See Cournot (54, pp. 56 ff.).

[25]See Cournot (54, pp. 79-80), where he introduces the duopolistic mode of behaviour (in our language: polypolistic price fixer type A) as a "restriction".

CHAPTER IV

THE FIRM 1870-1920

THE NOTION OF OPPORTUNITY COSTS

By 1870 most of the basic elements of the theory of the firm were present, and they are now part of a model. The elements were outlined but far from established on a solid basis. For instance, the laws of returns which are the basic explanation of the shape of the cost curves were not in a satisfactory state. An enormous amount of work remained to be done in the form of revitalizing the various elements, streamlining them, "proving" them, and, finally establishing their interconnections. The goal was to attain a stage of development where one terminal element, for example cost curves, could easily be retraced to more fundamental elements, such as the division of labor.

This is the reason why the authors of the period once more made a general assessment of most of the elements furnished to them by their predecessors. This is not to say that all the elements were scrutinized to the same extent because the amount of reformulation depended upon the needs of repairs of each. Let us see the first one, division of labor:

DIVISION OF LABOR

Adam Smith had dealt adequately with this topic. Consequently, few people spent a long time on this; they satisfied themselves either by reproducing Smith or by referring their readers to him. On the other hand, others such as Marshall spent chapters and even books, (Book IV) on it. In this line of thought there is no substantial contribution but the hope is now to find a way to integrate the division of labor with other elements. In the hands of Marshall it became the pillar in explaining the laws of returns. Marshall partly explained increasing returns to scale through two types of economies: internal and external.[1] These economies are ascribed mostly to the phenomenon related to the division of labor.

The various economies as seen by Marshall can be divided as follows (202, p. 169):

(1) External economies:

(a) those connected with using specialized skill and machinery due to the growth of industry,

(b) those improvements generally connected with the growth of knowledge, progress in the arts, etc..

---

[1]For the moment we accept this claim at its face value. However, the very existence of external economies has been challenged during the period 1927-1933.

(2) Internal economies:

They are defined as the possibility, through increasing the size of the plant, of using more productive methods of production, for example, a better division of labor.

Both classes of economies (either within the firm, or of the industry, or of the society), clearly involves the Smithian principle of the division of labor, especially if it is remembered that he included "inventions" as a consequence of the division of labor.

Marshall is not the only one to make explicit use of the division of labor as an explanation of the laws of returns. Edgeworth (73, pp. 81-84) enumerates causes which seem to come straight out of Smith's analysis.

FACTORS OF PRODUCTION

In the field of production, much remained to be done. One achievement of Walras, Menger and Wicksteed (202, p. 52) was to get rid of the classical "triad" or "holy trinity" of factors of production.[2] Yet this movement was not unanimous, because Marshall and Bohm-Bawerk remained faithful to the classical tradition, while Wieser and Wicksell retained it for convenience only. For Menger, a factor of production is simply a commodity of a higher order than a consumption good

---

[2]The expression "holy trinity" is borrowed from Stigler (202, p. 171).

(which by definition is of the first order).

Walras was the first to open the road in the new direction. With him the theory of production is thus freed from an unnecessary and cumbersome piece of machinery. If two units of factors of production are perfect substitutes, there is no need to separate them analytically. However, if they are not perfectly substitutable, then, we have good reasons to differentiate between them. This point had been blurred by classical analysis. It combined in one factor, for example, land, units of factor that are sometimes analytically different because of differences in productivity. This is similarly so for labor, and capital.

Walras simplified the analysis completely when he decided to use only one type of unit, a unit of services, thus separating the resource from the services it yields.

These two refinements are welcomed because what really matters for the theory of the firm is not whether it is hiring capital or land (whatever the names), but the contributions of these factors in terms of a quantity of services. With the new type of unit, heterogeneous units, in terms of Land, Capital, are reduced to a homogeneous type. Furthermore, the distinction between the source of the services and the services themselves help to separate clearly the two markets in which the firm can deal to obtain inputs. If the firm does not own the factors, it can rent their services by paying a wage, an interest, or an ordinary rent. Yet in the case of

a machine, land, it can buy these things outrightly, thus dealing in the capital goods market. Walras showed that the apparatus needed to solve the first type of problems is inadequate to deal with the second one. In the first case, we are talking of equating prices of the factor services with the marginal revenue product of a particular unit of factor, while in the second case, the purchase price of a capital good must be equated with the sum of a discounted stream of services. This is the problem which he solved in his Lesson 28 (220, p. 307 ff.).

As noted previously not everybody followed the path of Walras, For instance Marshall not only clung to the classical "triad" but enriched it with a fourth member: organization or entrepreneurial activity.[3]

THEORY OF PRODUCTION

Being in possession of a new outlook as far as factors of production are concerned, the production problem now consists in allocating these units of services of factors in such a way as to maximize output. Keeping in mind that the competitive conditions still form the normal framework in which the analysis is operating, the theory of production is simply an exposition of the mechanism of competition in its aspect of allocating different kinds and qualities of agents

---

[3]However Marshall (136, p. 283), when analyzing the costs of production, does not use this quaternary classification.

of production (185, p. 1009).

Another characteristic of the approach of the economists of the 1870-1914 period must be noted: they did not see the production problem in isolation but as part of a greater problem of distribution. For many of them, solving the distribution problem was equivalent to solving the production problem. Their ambition was to build a theory that would "tell us which quantities of which products each firm will decide to produce, and which quantities of which productive services it is going to buy in view of the given tastes of perspective consumers of its products and the given propensities of the same consumers considered as 'owners' of productive services."[4]

To determine how much to use of one factor and simultaneously to determine how much to pay this factor, the manager needs to operate with two concepts: costs and revenues, both imputable to this individual unit of the factors. Costs and revenues can also be formulated in relation to units of output. Both formulations are sometimes necessary to solve most of the problems of the firm.

---

[4](185, p. 1010). This is more or less done by the people of the period. For instance Marshall studies production in Book IV 'On the Agents of Production', and then separately in another Book VI, 'Distribution'. "This order does...involve a certain separation of the theory of production from the theory of distribution, and perhaps conceals somewhat the unity of the two theories as provided by the marginal productivity principle." (Hutchison, 110, p. 83).

In contrast with the previous period, most people lavished much space and effort in establishing a theory of production; Walras and especially Pareto overhauled the theory completely. The criticisms of Pareto on variable coefficients and marginal productivity analysis (pp. 605, 636 of his Manuel) because of the presence of limitational factors are well known (110, p. 83).

COSTS

For the theory of the firm it is obvious that costs are one of the data necessary to enable the managers to determine the best output; to determine the rewards of the factors etc.; in a word costs are necessary to solve the problems raised by production theory.

Yet, as will be shown below, not merely any notion of costs will do. Up to now three concepts of costs have been used by the economists of 1776-1870: (1) real costs, (2) comparative costs and (3) money costs or outlays (mainly used by Cournot).

The contribution of the present period was to devise the notion of opportunity costs. They are defined today as the value of "displaced alternatives; the cost of obtaining anything is what must be surrendered to get it.... In the theory of production, they reflect also the value of alternative uses of productive factors, that is, of products which do not come into existence because existing products are

preferred" (172, p. 2). Alternately, the doctrine of opportunity cost can be stated in two ways:

The opportunity cost of producting commodity A is the value of commodity B, C, D... not produced with the same amount of resources devoted to the production of commodity A.

OR

The opportunity cost of a unit of resource applied to commodity A is the foregone products B, C, D... this unit of resource could produce in its best alternative use (216, p. 469).

It must be shown now that the notion of opportunity cost is for the purpose of the theory of the firm not only superior to others but absolutely indispensable, and that the other notions of costs, notably money costs, may lead to wrong decisions in the conduct of the firm. Consequently, the concept of opportunity costs is a definite and outstanding contribution to the theory of the firm. Let us deal with the various notions of costs already devised:

(1) Real Costs

The manager has no use for real costs. He is interested solely in his monetary outcome, because he is accountable only in monetary terms. The accounting of the firm is burdening him with monetary costs, it does not matter to him, to know what the burden, in terms of effort, is to others. For instance, the wages determine his burden as far as the labor he hires is concerned; the fact that this wage measures

or does not measure exactly the burden of working to the laborer is beside the question for him.[5]

Secondly, for the firm, the use of real costs would involve a lack of common denominator and this might prevent the manager from forming an exact notion of costs.

Finally, as used by the classicists, the real costs had a connotation of being exclusively historical costs, that is costs that are, will, or had been experienced, as opposed to those which could have been experienced. Historical costs are those that will sooner or later be a matter of recorded history. Only historical costs are antithetical to opportunity costs which will never be a matter of recorded history. Paradoxically, as will be shown below, the opportunity costs are the only ones that matter, even if they are also those that will never show in the books of the firm. Consequently, there may be such a thing as real opportunity costs, real costs being used in this context simply as a form of accounting units. However, as noted above, as such they do not interest the firm; at most they might interest the welfare economist.

(2) Comparative Costs

This notion of costs also devised by the classicists to solve international trade problems was, moreover, formulated

---

[5] "The businessman is obviously not greatly concerned with real costs--with distinctions between burden costs to someone else and money cost to himself...he is interested solely in his monetary outcomes." Davenport, (61, p. 108).

in terms of real labor costs; as such, this notion of cost is complicated thusly:

(i) To be operative, a comparative cost doctrine, formulated in real labor costs, requires that the prices of the commodities within a country be roughly proportional to their real costs. This is unlikely because of differences in wage rates, occupations, variable proportions of the factors (211, pp. 493-508).

(ii) A labor cost approach is not realistic. Other factors of production are necessary for production. An attempt at patching up was made by Senior, when he introduced capital cost in the form of abstinence, a form of real cost. Later Marshall combined labor and capital in "units of productive power". But all these are insufficient to measure costs completely. (See iii)

(iii) One factor that was never taken into consideration in a real cost approach was Land. Land in the classical analysis received a rent which is not a real cost as far as the nation is concerned. This is probably the most important deficiency, and this error was pointed out by Pareto (217, p. 509).

(iv) Finally the Ricardian comparative cost doctrine was based on constant costs of production whatever

the scale of operations. This restricted its field of application for a time until a method of indifference curves could incorporate varying costs.

Besides these "technical" difficulties, there is a further drawback which consists in the fact that the firm does not need comparative costs if it can use opportunity costs. In other words, in simple cases (as seen in our classical chapter p. 37) of no interdependence between the products produced by the firm, the comparative method will yield the right solution, yet it will not permit the direct reading of the complete opportunity costs, of producing a particular commodity. The basic problem for which the comparative cost doctrine was devised was for a case where the factors of production are immobile between countries. In this context the mobility of commodities makes up for the lack of mobility of the factors of production. The comparative cost doctrine stipulates that this movement should go on as long as the ratio of the real costs of the commodities differ from one country to another. This problem is rarely present in a firm; usually there is mobility of the factors between departments, products, if not between processes.[6]

---

[6]However in very short-run, the mobility may not be present; and the method of Linear Programming which we will use to solve the problems is not far from the comparative advantage approach.

Yet comparative costs are not inimical to opportunity costs. Embodied in these costs ratios is an implicit "could have been" system of computation. Suppose two countries I and II where in country I, commodity A requires 10 units of labour and commodity B 5 units; while in country II, commodity A requires 20 units of labour and commodity B 5 units of labour. Country II is confronted with two opportunities as far as commodity A is concerned: buying A at home at a cost of 4B or buying in country I at a cost of 2B. Evidently it is more advantageous to import A and export B to pay for it. This is as far as the theory goes; it compares opportunities, but it does not determine these costs themselves. However, handled this way, the notion of comparative costs is close to the opportunity costs.

(3) Money Costs

The defects of the real costs as a unit of account can be remedied by the use of money as a common denominator. Money costs are then a way to evaluate real costs and thus solve the accounting problem. Yet up to now, for example, in Cournot, money costs have been used in the form of historical costs, past, present, or future. In this respect, they are inadequate, and sometimes misleading if used as a basis to make managerial decisions. It is the subject of the balance of this section to prove that point. However, without the historical connotation they are not antithetical to opportunity costs. There is such a thing as money opportunity

costs; but here money serves only as a unit of account.

Showing the inadequacy of money costs is a lengthy business because various situations are possible.

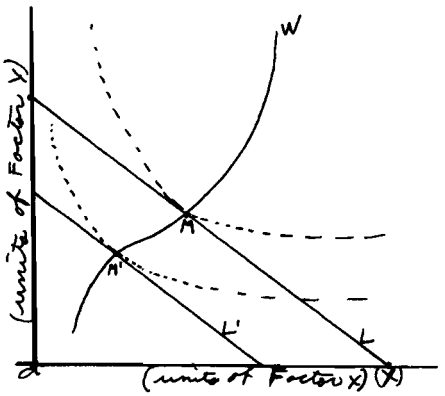
(a) Single Product: Long-Run

To begin with a simple situation, let us take the firm which produces only one commodity A. This firm, like any firm, must solve two types of problems:

(i) How to produce the given product.

(ii) What quantity of this product should be produced.

(i) The firm having decided to produce commodity A (although as will be seen below p.121 this is begging the question) has to decide first <u>how</u> to do it; that is what factors of production to use and in what quantity. The money cost theory of the firm would proceed as follows:



With a given sum of money L, it can hire OX of factor X and OY of factor Y; where $OX = \frac{L}{p_x}$ (p_x being the price of factor X), and $OY = \frac{L}{p_y}$. This line YX is the isocost L. Its a straight line depicting the various combinations of X and Y that costs a constant amount L. If various values of L are used, a whole map of isocost curves can be constructed. Similarly, we can plot a map of isoquants, each isoquant showing the various combinations of X and Y that will produce a constant amount

of product A. (For the construction of isoquants see below p.214). The cheapest way to produce a given quantity of A is combination M, and M' for another quantity etc. whose least cost way of being produced lies on what is called the expansion path (the series of points of tangency between the isocosts and the isoquants). Along the expansion path,

$$\frac{MP_x}{MP_y} = \frac{p_x}{p_y} .$$

If the supply for the factors of production is not perfectly elastic, the isocosts are concave (from below), and consequently the least cost combination involve:

$$\frac{MP_x}{MP_y} = \frac{MRC_x}{MRC_y}$$

(where MRC_x is the change in the cost of hiring factor by one more unit). By pairing isocosts and isoquants (see below p.285) we can establish the Long Run cost curve of the firm which shows the cheapest way to produce various outputs. With varying returns to scale or varying prices of the factors of production, the cost curve is likely to be U shaped. At the least cost points M, M' etc. of figure page 116, we have, (since $MI_x = 1/MP_x$), $p_x/p_y = MI_y/MI_x$ or $p_xMI_x = p_yMI_y$; this last expression is equal to MC for the output in question because from the least cost point the cost of increasing output by one unit is either p_xMI_x if factor x is used or p_yMI_y if factor y is used. Strictly speaking this is true only if the increase in output is infinitely small. Consequently

the least cost combination implies:

$$p_x = MC.MP_x \quad \text{and} \quad p_y = MC.MP_y.$$

This equation determines how factors should be combined; it does not say what quantity of factors should be hired because this obviously depends upon the output that is wanted to be produced. This brings us to the second problem which consists in how many units of product A should be produced.

(ii) To maximize profits the firm should increase output until MC = MR.

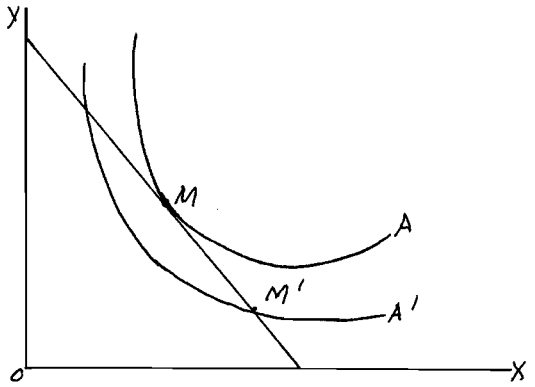
When this equation will be satisfied this will solve the second problem. That the firm must increase output until MC = MR is a self evident proposition implied in the postulate of profit maximization (97, p. 169). As a corollary, this proposition (at the point of equilibrium) implies in terms of factor hiring that $p_j = MRP_j$. This last equation defines equilibrium in terms of hiring of factors. Now, this equation is always verified except if MC $\neq$ MR and vice versa.[7]

Finally it can be shown (as Walras thought (220, p. 552) that the least cost condition (i) and profit maximization are two different conditions because one can be satisfied without the other being so.[7]

---

[7] See appendix F.

Implied in this analysis is some form of opportunity cost. If condition (i) is not satisfied that means that we are operating at M' (with the same costs we produce a lower output A'). The opportunity cost of operating there (outside the expansion path) is A - A' in terms of product.



On the other hand if the firm does not maximize profits the internal opportunity cost[8] of each unit of factor is greater than its outside opportunity cost.[9] Each factor will be underutilized and the firm will be losing the opportunity of gaining the triangle BCD of product A (in diagram No. 2 appendix F). Rule (i) has an important application. In the case of pure competition the MC curve (as long as the average costs are covered) becomes the supply curve of the firm. In the case of monopoly the MC curve is the guiding line (with which MR should be equated) to decide about output.

This approach while fundamentally right can lead to wrong decisions if these curves are built exclusively on

---

[8]By internal opportunity cost is meant the loss in revenue to the firm if this unit of factor is not used by the firm in that particular employment. This is sometimes equal to the marginal revenue product of this resource to the firm. [illegible]

[9]By outside opportunity cost is meant the loss in value in alternative uses outside the firm, for example this is equal to the market price of the unit of the factor.

outlay or money, data as is done by the accountants. This will be in the case of the firm who owns factors of production which have alternative uses outside the firm. This is encountered when people farm their own land, provide their own labour force or where being in other types of businesses they provide their own money capital, or entrepreneurial labor. Some costs are not outlays and as such do not appear in the books of the firm. Suppose a firm just breaking even, where its LRAC accounts for outlays only; here the owners miss a chance of getting some remuneration for their services, in providing money capital or labor or both, if they would simply rent their resources to outsiders. These revenues are the outside opportunity costs. Consequently if the firm was formerly breaking even, it will now make losses, if these outside opportunity costs are added, or in other words, by getting out of business and working outside these owners would now have positive revenues. Consequently monetary models must be corrected with opportunity costs to enable the manager to make the right decision. This is common knowledge today among economists although few textbooks take care to spell out clearly these opportunity costs and to include them in their cost curves.[10]

---

[10]Actually of all those within my knowledge, only Stigler (204) takes care to do it. Yet his approach is single product. While internal opportunity costs are the most important, especially in the planning stage where there is no such thing as a single product firm.

Finally, even if these problems of how to produce are solved, the initial question is still unanswered: what to produce? Until now we worked under the simplifying hypothesis that commodity A is to be produced. But in the long-run, since we are in the planning stage, the crucial decision is first of all what to produce. If the firm decides to produce A, with the same productive resources it will not be able to produce commodities B, C, D (which could potentially be produced). Consequently alternative products are a source of opportunity costs. From the point of view of the products, the opportunity cost of producing A is the value of B that cannot be produced with a given amount of resources. If some A and some B are produced, the cost of A produced is equal to the value of B that could have been produced if no A had been produced, less the value of B actually produced. Given some A produced, the marginal opportunity cost of producing one more unit of A is the quantity of B that will have to be cancelled.

We have now arrived at the multiproduct analysis. This should show how insufficient is the single product approach. In the planning stage, there is no such thing as a single product firm, except if its owned factors or the factors it can hire are completely specialized individually and as a combination. Let us note that in such a case these factors have no alternative uses except in firms competing in this same product; consequently the rewards of these factors is

exclusively in the form of pure rents. Granted these special conditions, in the long run the supply of the firm is governed solely by the sum of money costs plus the outside opportunity costs of its owned factors. If the restriction of complete specialization of factors is removed (individually or as a group), alternative products must be contemplated within the firm. Obviously an entrepreneur cannot decide that his firm will be a single product firm (as might be under constant costs and pure competition) (135, p. 275) except if he has compared the possible revenues of A with the revenues from B, supposing the firm could either make A or B.

This is why we must wait until the next section to picture the mechanism that will enable an entrepreneur to take this decision. Logically the entrepreneur must decide first, what to produce, next how much to produce, how to produce it, etc. And the what to produce can only be decided in a multiproduct firm model.

(b) Single Product: Short-Run

When we move from the long run to the short-run the money costs are even more misleading. It is customary in the literature to distinguish in this period two types of costs:

(i) fixed-costs which are outlays which do not vary during the period,

(ii) variable costs which are outlays that vary with output.

This distinction made, the textbooks go on stipulating that the MC curve, as long as the variable costs are covered, becomes the supply curve of the firm (182, p. 475).

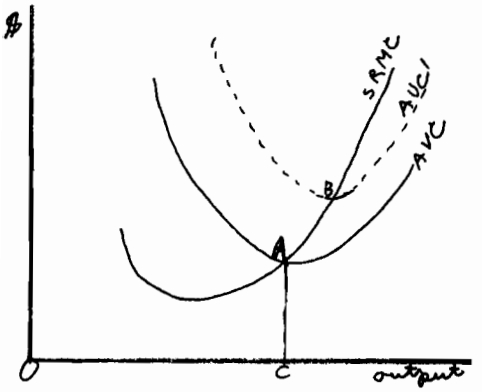
It has long been obvious to many people that the length of the time period considered by a certain decision determines what costs will be relevant. Since Jevons, in the short-run, we use to say that bygones are bygones. This is why the textbooks show that the fixed costs have no effect on price or output in the short-run.

The blind following of this simple money costs approach might involve the firm in not maximizing profits. The reason hinges on the way fixed costs are usually defined. If the fixed costs are loosely defined as contractual expenses that cannot be varied in the time period contemplated (181, pp. 67-68), it may happen that although the expenses are fixed, the units of factors appropriated through these expenses may have alternative uses outside the firm even in short run. For instance, a retail store may have a ten year lease, yet it might be possible to sublet in whole or in part for short period. Any revenue that can be earned outside the firm in the short-run by the fixed factors, represent an opportunity cost of using them inside the firm. In this case these revenues must be added to the usual variable costs when determining the shut-down point.

Similarly, the owned fixed factors may not have alternative uses outside the firm at the moment, but they may have

alternative uses in the future inside or outside the firm. The discounted value of these alternatives uses through time must then be added to the present variable costs.

Finally goodwill considerations, or the costs of shutting off and restarting the plant, or getting back a clientèle, must be added. The same with other more subtle elements such as the loss of control of the firm, the reduced status of the managers, if the plant is closed down, not to mention the financial problems of liquidity involved.



The graphical solution is given at the left. In the purely monetary approach, the supply curve is MC as long as it is superior to CA. When opportunity costs, mentioned above are introduced, the firm will supply as long as MC is greater than AVC', where AVC' is obtained by loading the monetary variable costs with the opportunity costs.

Alternately the monetary approach can be made right by redefining the fixed costs as the expenses paid for factors which have no alternative uses during the period considered.

If we want to study the case where, in the short-run, some factors that are in variable supply are also under the jurisdiction by the firm, for example, labor services, monetary accounting is also misleading. If these factors have

alternative uses in the short-period considered, the revenues they could earn outside should also be imputed as a cost to the firm and added to the monetary variable costs. The effect on the short-run supply curve of the firm will be similar as in the previous case.

Introducing monopolistic elements in the model will not change the basic principle of operation. Yet some alterations should be made: for instance, it is well known that the supply curve of a monopolistic firm cannot be drawn (178, p. 85; 230, p. 122). The modification is then to equate MR to MC, instead of P to MC, while taking care to use MC only in its permissible range. In addition, the price and output are determined as usual.

As in the long-run case, this analysis is in marginal terms. That is the short-run equilibrium of the firm is attained by equating SRMC to MR.

In a way the SRAC curve represents a sort of short-run least cost way to produce various outputs. This requires that the variable factor is employed until its $MRC_X = MRP_X$. For example in the competitive case we always have: $MC = p_X.MI_X$ (since factor x is the only one that may change the costs). Consequently when the firm is in short-run equilibrium: $MR = MC = p_X.MI_X$. In other words, $p_X = MR.MP_X = MRP_X$ so that in equilibrium, a quantity X_1 of X will be used. Being in marginal terms, this analysis

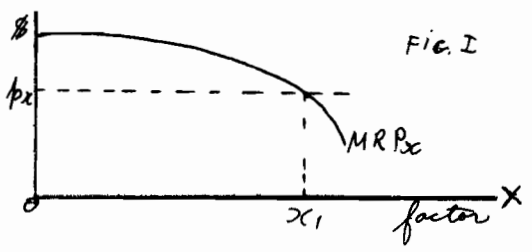
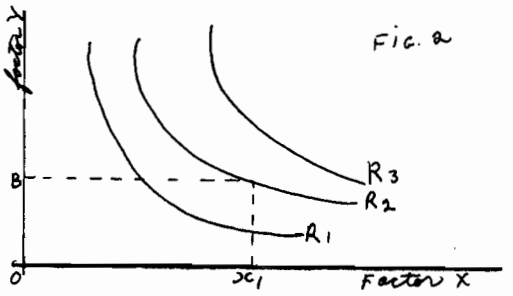


Fig. I

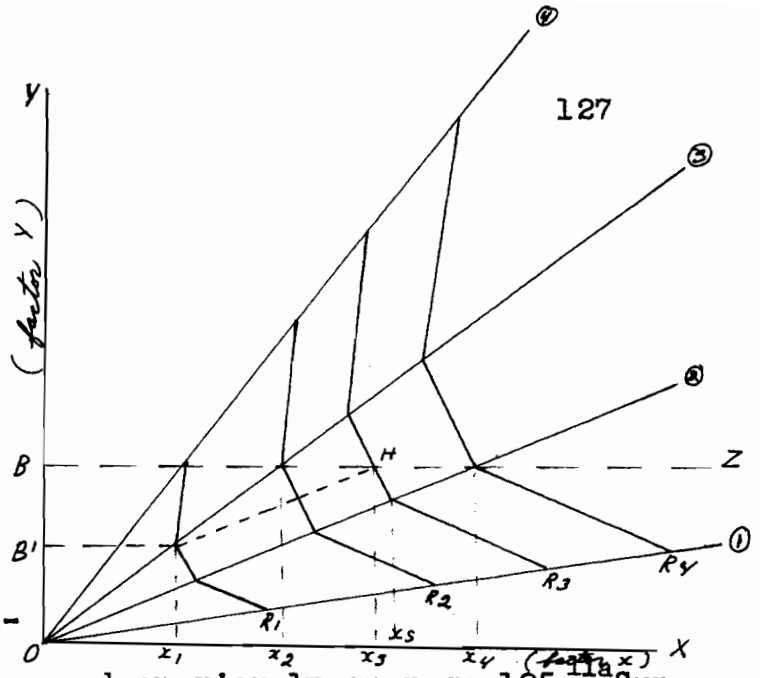
needs an infinite number of processes, e.g. infinitely variable proportions. Such that if Y is fixed at OB, a quantity R_2 of A will be produced because maximizing profits requires that an amount of X_1 of X to be used.



(see computation in the previous diagram). In the short-run, even if the firm is single product, there is a further source of opportunity costs which requires a special method of computation. It is the case, also encountered in the long-run, where the number of processes to carry on production is limited. The first difficulty is in the construction of the map of isoquants; or more exactly there are no isoquants but simply lines of output.[11] In such a context only the points on the product lines (Nos. 1, 2, 3, 4, in the diagram below) are points representing outputs that can be produced by pure substitution of factors, all other points are attained by substitution between processes. This can be explained as follows. Suppose a production 'surface' where rays Nos. 1, 2, 3, 4, represent the only possible technical combinations of factors. In short-run, the firm may be endowed with only OB of factor Y (the fixed factor). Therefore on this BZ

---

[11] See below p. 214 for a detailed construction of this map of isoquants.



cross-section we may visualize a continuous TP_X curve only if we utilize portions of the production surface which are not real. From this, a MP_X curve can be deduced; then a MRP_X curve which might look like an ordinary MPR_X curve like the one used previously on page 125.[11a] Supplying the p_X might determine the employment of x_3 of X (or the production of R_3 of product A) to maximize the profits of the firm in short-run. This introduces us to an apparently insoluble situation if the internal opportunity costs are computed exclusively by the straight money cost approach. R_3 has to be produced, but there is no way through conventional analysis, to ascertain the best way to produce it. The only opportunity cost of a factor computed through money analysis is the MRP of one factor X. And in this case it is inoperative because of the limited availability of processes. Another possible non-optimum solution attainable by straight money cost analysis is to use x_5 of X; but here some Y is unemployed. Therefore straight money cost analysis breaks down. What is needed is a method of computing the Internal opportunity cost of both factors and a scheme to combine them in a least cost manner. This is done through Linear Programming which utilizes substitution between processes in lieu of varying exclusively the proportions of factors used. This method will show

---

11a With the difference that it will be a step function instead of a smooth function and some of its points ,e.g. x_3 and R_3 are theoretical.

how to produce R_3 with no more than $0x_3$ of X and OB of Y. Here is how:

The procedure is simply to draw from H a parallel either to the line of process No. 1 or 2 until it meets line of process No. 2 (or 1). This determines the process-mix that will be used.

Starting with process No. 2, OB_1 of Y and $0x_1$ of X are used in process No. 2, thus producing R_1 of A. Next, BB' of Y and x_1x_3 of X are combined through process No. 1, thus producing R_3-R_1 of product A. The sum of these two outputs is R_3 of A, and the quantity of the factors used is OB of Y and $0x_3$ of X.

In this particular case (single product firm under limited processes), the opportunity cost of using a particular unit of a factor in a particular process is the foregone product that could be produced by using this particular unit of factor in another process.

In terms of period analysis, we can distinguish still a further case: the very short period, where all the factors are in fixed supply. This is a situation often encountered in practice; the quantity of most resources cannot be varied overnight. Two situations can be distinquished:

(i) the firm can use an infinite number of processes,

(ii) only a limited number of processes are available.

(i) If the supply of all the factors are fixed, marginal analysis has no room to operate. Maximization under these

conditions asks for the greatest production as possible (highest isoquant) except if MR becomes negative. Referring to diagram No. 2 p. 126, if the firm is stuck with OB of Y and OX_1 of X it should produce R_2 of A with combination M. Money costs do not matter here, neither do outside opportunity costs which do not exist, because if the period of time is too short to hire new units of factors, it is also too short to hire out the factors presently under the control of the firm.

(ii) Under the limited processes, the situation may present the same problem as met previously in short-run. Going back to the figure on page 127 we may suppose the firm is endowed with OB of Y and Ox_3 of X. The solution is also found by the method of linear programming provided we can use many processes simultaneously.

MULTIPRODUCT FIRM

A multiproduct firm is one where two or more products are simultaneously produced. There are two sets of reasons why a firm may produce many products: (1) technical production conditions, and (2) revenue or costs interrelationships, for example, complementarity.

On the technical side two situations are encountered:

(i) fixed proportions, and

(ii) variable proportions.

Furthermore, as before, two main time periods may be distinguished.

We must then study the multiproduct firm under two main types of situations, regarding the relations (in production or in sales) between the products: interdependence and no interdependence. Not only could a firm produce two products, but their demand curves might be interrelated in the sense that the more it sells of A, rightward or leftward shifts in the demand curve of product B occur. Similarly, the more of A is produced, the less or the more costly it becomes to produce B. Finally more complicated cases imply simultaneous interrelations on both the sales and costs of the products involved.

Let us begin our analysis with the case of no interdependence whatsoever between the products. A question should be answered beforehand: why should a firm produce many products if there is no interdependence? The answer may be stated in terms of the utilization of some hidden factors, for example entrepreneurship, which very often has some unused capacity; also, monopolistic competition may be a breeding ground for multiproducts.

Multiproduct Firm

No Interdependence

Long Run:

Fixed Proportions.

Let us suppose that even in the long run, the technical conditions of production are such that at least two products A and B must be produced in an invariable proportion.

There is no way here to determine MC of A and MC of B. Only a single cost function can be isolated, making sure that it includes all the money costs of the hired factors and all the outside opportunity costs of the owned factors.

The equilibrium of the firm is where:

$$MC_{a+b} = r.MR_a + s.MR_b$$

where r and s are respectively the number of units of A and B that are produced with each batch of production (222, p. 292 ff.) and MR_a and MR_b the various marginal revenues per unit at which these units of products can be disposed of in the market.

Example: Suppose each batch produces 2 units of A and 4 units of B, out of one unit of K the basic raw material, then MC_{a+b} = cost of buying and processing one unit of K. The entrepreneur should increase production until

$$MC_{a+b} = 2MR_a + 4MR_b.$$

Yet a further condition must be added here: none of the MRs must be negative. If such a landmark is overpassed, it will pay to destroy the excess of product (A or B) when the MR of the product in question is negative. This is so except where it costs more to destroy it or leave it out than the negative result added. This provision shows how simple marginal money analysis may lead to errors. Incidently, if the same problem is handled through L.P. this provision is automatically taken care of.

No Interdependence

Long Run

Variable Proportions:

Two situations can be studied here:

(i) The production of either of the products can be undertaken under constant or variable costs and furthermore one product alone can be produced if desired. This is not, properly speaking, a case of joint production. Yet it is in such a model that the prospective single product firm in the planning stage can determine which product or products it will handle.

(ii) The production of either product cannot be undertaken separately; this is the real case of joint production. The costs may or may not be constant; however, for the moment, the costs of the products are not related.

Let us study situation (i); some of the analysis used below applies also to situation (ii).

To illustrate this case let us take the firm which is contemplating (in the planning stage) producing either product A or product B or both. With regard to the cost situation, let us suppose it is endowed with a Ricardian cost function i.e. constant costs of production which are also easily distinguishable.

The costs of either product A or B are computed by adding all the money costs of the hired factors plus the outside opportunity costs of the factors owned by the firm. But the contemplation of the production of many products is

another source of costs. Obviously both types of factors, owned or hired, now have alternative uses within the firm. When a certain amount of funds is devoted to factors that work on A, they do not produce B and vice versa. There is then a new source of opportunity costs.

In terms of equilibrium of the firm, each production should be pushed until $MC_a = MR_a$ and $MC_b = MR_b$. However, if the costs are constants and include the money costs and outside opportunity costs, we cannot use this simple rule to decide which product(s) will be produced, because the "internal" opportunity costs are not taken care of. Let us see why:

Suppose, with the data provided above, that $MC_a < P_a$ and $MC_b < P_b$; according to the simple rule both productions should be increased. However we are saying they perhaps are not. The reasoning is as follows:

$P_a - MC_a = R_a$ (the profit on a unit of A) and

$P_b - MC_b = R_b$ (the profit on a unit of B).

In such a case, provided the funds are limited, when the production of A is increased, the costs involved are MC_a plus the opportunity of making a profit on B (since if we devote funds (and factors) to the production of A they are not producing B). Consequently the complete costs of producing A is: $MC_a + R_b$. Similarly for B whose complete cost is now $MC_b + R_a$.

Supposing that MC_a is \$4 and P_a is \$5, and MC_b is \$5 and P_b \$8. In such a case it can be shown that A will

never be produced, if you remember that this is a purely competitive case, constant costs and a given capital fund. The reason is that the complete cost of A is $7 which is greater than the MR_a, while the complete costs of B are only $6 which is smaller than MR_b.[12]

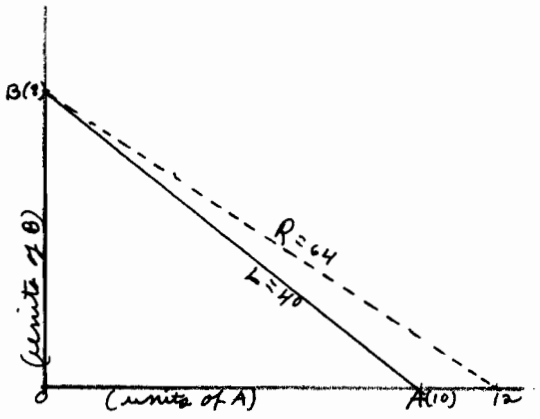
This simple example illustrates what we mean by the complete opportunity costs approach: it includes foregone profits. An alternative and more exact method would have been to compute the alternative revenue that could have been earned in B by devoting the same amount of funds in B as in A.

The foregoing example can also be represented graphically with the use of indifference curves, using a comparative cost approach. (This is one case where this classical method of analysis can yield the right solution). First, let us build a map of cost contours; a cost contour (181, p. 373) represents the various amounts of A and B that can be produced with a given amount of funds (costs) L, making sure, to correct the classical comparative costs in such a way that they include the money costs of the factors hired plus the outside opportunity costs of the owned factors.

---

[12]This example is not rigorously exact because the amount of funds devoted to one unit of product A is different from the amount involved in the production of B. We will shortly take care of that.

With a certain (limited) amount of funds,[13] if all the resources are devoted to A, OA will be produced, and if all are devoted to B, OB is produced. Having postulated in our former example that the costs are constant, $OA = L/MC^a$, and $OB = L/MC^b$. Furthermore, BA will show the different combinations of A and B that can be produced with a given L. It will be a straight line because its slope at any point is MC^a/MC^b.[14] Let us note in passing that this slope is a ratio of direct costs and not a ratio of "complete" opportunity costs as defined in our example of the previous page. Similarly we can construct iso-revenue



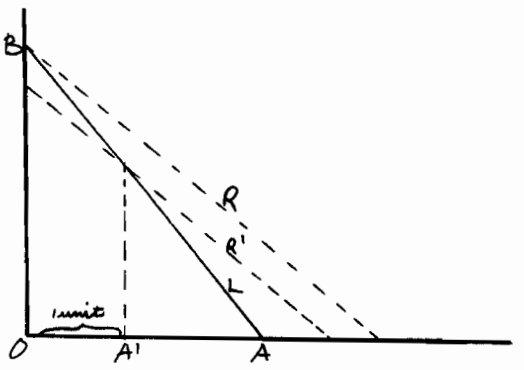
---

[13]Here is the justification for our approach in terms of time period. Previously in the reasoning we were employing expressions such: for a given amount of costs, only L/MC_X of X can be produced. This is not a short-run limitation, which in this context is characterized by a completely inelastic supply of one factor. Moreover, money is not used here as a factor of production but as a means to obtain them. The limitation of the amount of money capital available to the firm is realistic both in the short-run and in the long-run. As far as we are concerned here, this is an external limitation imposed on the firm. And it is not reasonable to postulate the firm as having access to unlimited amounts of funds, even in the long-run. Money capital can be integrated into the theory of the firm but only in much wider models of the firm, as those tentatively drawn by Boulding (25), and Gabor and Pearce (90).

[14]See, the slope at any point is $OB/OA = L/MC^b/L/MC^a = MC^a/MC^b$. This will be constant because both MC^a and MC^b are postulated as constant.

curves, each curve depicting the various combinations of A and B that will yield a certain constant revenue R. The construction of the R curves is simple: to obtain a given R we need R/P^a of product A, if only A is used to obtain this revenue; or if only B is used, we will need R/P^b to raise it. By exploring various Rs we can obtain a whole map of them where the farther to the right a particular R curve is, the greater the revenue it represents. If the firm has L = 40 amount of funds to invest, it will maximize profits by attaining the highest revenue possible. This graphically is done at point B where the L curve touches the R = 64 revenue curve. Consequently, a combination of OB of B and zero of A is the one that will maximize the profits of the firm. In the case of constant costs and prices, these curves will usually touch only at their extremities, thus entailing the complete specialization of the firm in one product; only in the case where the slopes of R and L are equal can any combination of A and B be satisfactory.

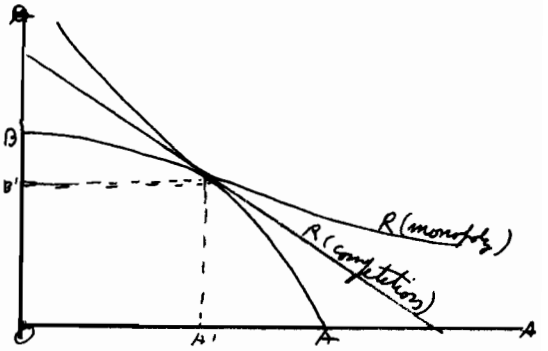
We can now introduce the "internal" opportunity cost of producing one unit of commodity A, that is, the "internal" cost of changing the optimum combination by producing one more A and less B (because we work with limited funds). Under constant costs, each time one more



unit of A is produced the production of B is diminished by MC^a/MC^b; the "internal" opportunity cost is then $MC^a/MC^b.p^b$. The loss in value is this amount minus pa. Graphically (above) this is equal to $R - R'$. The direct costs (L) do not enter into the picture because they are constant; in money terms in the above example this amounts to $2.40.

Consequently we must make one remark at this point: It might be inexact (135, p. 286) to call the slope of the cost contour curve a ratio of marginal opportunity costs, because in this slope nothing is said about the "internal" opportunity costs. Actually the cost-contour curve represents the comparative direct costs (money plus outside opportunity costs), but not the "internal". In situations of no interdependence between the products, the comparative costs can yield a correct solution, but since foregone profits are not included, it will fail in the case of interdependence (see below pp. 148 ff.

Having established our method of analysis we can proceed rather swiftly with the other situations, always under the assumption of a limited budget (capital) of the firm. For instance the case of increasing MCs:



In this case the cost contour curve will be concave to the origin. Its slope at any point is MC^a/MC^b (both determined by partial differentiation). Optimum proportions between the products is determined by the point of

tangency with the highest iso-revenue curve (a straight line in competition and a parabola in the case of monopoly). Any combination of A and B other than D will entail an internal opportunity cost R - R' (as determined previously). As before (p. 137 above) the internal opportunity cost of one more unit of product A is $MC^a/MC^b.P^a$, in the case of competition, and $MC^a/MC^b.MR^b$ in the case of monopoly. The ratio of opportunity costs can then be written as $\frac{(MC^a)^2.P^b}{(MC^b)^2.P^a}$, in competition.[15a] What precedes is a two products model. The general conditions of equilibrium, in terms of outputs and products produced and quantity of factors hired, of a general multiproduct long run model are as follows:[15]

1) Any output of any product should be produced under least cost conditions. Symbolically, if i is any product a, b, c, ... n, and j is any factor x, y, ... m, the least cost condition can be written as

$$MRC_j = MC^i.MP^i_j.$$

2) Profit maximization requires that each product output is expanded until $MR^i = MC^i$. This implying that the optimum rule for hiring factors has been satisfied; that is,

$$MRC_j = MRP^i_j.$$

---

[15] See appendix G for a derivation.

[15a] The reader might realize by now that our approach is not fruitless; it yields results different from those who assume that the slope of the transformation curve yields the ratio of opportunity costs.

In other words, these conditions mean that in equilibrium:

1) Profit maximization stipulates that there should be no profit earned by transferring productive power from one selling market to another, and that each factor is employed until its cost is equal to its value to the firm.

2) The least cost rule (implied in (1) above) stipulates that there should be no profit earned by utilizing different combinations of factors in the production of each product.

Marginal opportunity costs are obtained as follows:

1) the opportunity cost of employing one unit of factor j in employment a is its greates MRP^{i}_{j} in alternative products employments; or in alternative processes, or alternatively,

2) the marginal opportunity cost of producing one more unit of a product i is the value of the foregone products b, c, ... n that could be produced with the same amount of resources necessary to produce one unit of i. This is equal to:

(i) the outside opportunity cost of the resources used, (money costs of the hired factors plus the outside opportunity costs of the owned factors), plus (ii): the profit $MR^{i} - MC^{i}$ (called here internal opportunity cost) that could be earned on the alternative product. This leads to the rule

that in equilibrium each dollar spent in the form of money costs on hired factors or each dollar value of outside opportunity cost on owned factors, should get the same return in any product, or process.

This provision is not necessary in the case of constant costs, perfect competition and unlimited capital funds.

Multiproduct

Long Run

Variable Proportions

(ii) Joint production

Let us now take the case of the firm which technically has to produce jointly two or more products, yet there is no direct way to distinguish the costs of production of each. This is the famous Beef and Hide case; however since we are in the Long Run there might be ways to vary the proportions of the products produced, for example, through cross-breeding, varying the feeding formula. In most cases this cannot go on indefinitely (varying the proportions) because one restriction imposed by this case is that there is no way to produce exclusively one product, and this is the difference between this case and the previous one.

The first step to solve the problem of the multiproduct firm is to try to isolate the particular costs of each product, (remembering that they are not given directly). One way to do it is explore all the total costs figures corresponding to all the possible outputs and all the variable

proportions that could be used. In the case of two products, we obtain a figure in three dimensions, where the cost data form a cost surface (204, p. 131). Let us note that this cost surface takes care only of the direct costs: money costs and outside opportunity costs of the owned factors. There is no way to read the "internal"opportunity costs on this surface; to do that we need another piece of information: the revenue function. Finally, because of the restriction of joint production (at least a bit of both must be produced) this surface will not cover, when projected on the BOA plane, all the first quadrant. Mathematically speaking this aggregate cost function can take various forms; homogeneous of first degree, homogeneous of other degrees, and finally non-homogeneous. Yet to be economically relevant it must conform to the various kinds of MC situations usually explored by economic theory: constant MC; eventually increasing MC; diminishing MC. To isolate the particular MC of a particular product, at a certain level of production, we simply have to take, at that point, the partial derivative of the cost function $L = f(A,B)$ with respect to A or B.

Examples:

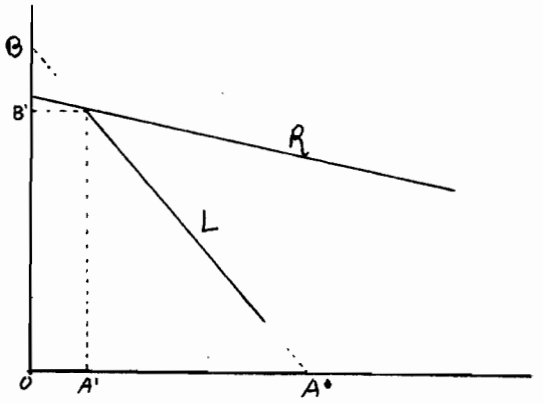
(1) Constant MC would mean that at any point of the surface (or any combination or level of output) that:

$$\frac{\partial L}{\partial A} = k, \text{ and } \frac{\partial L}{\partial B} = j.$$

With this information, the firm can now determine which

proportion of A and B it will produce as follows:

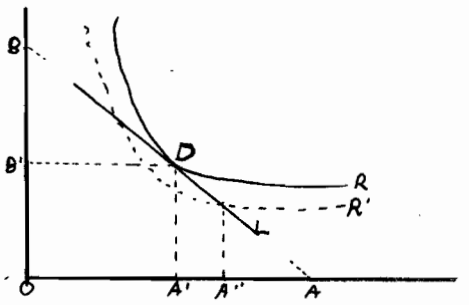
With a given money capital L a cost contour curve can be built; it will be similar to the one of p. 135 except that it will not extend until the axes. The cost contour will be a straight line because in this hypothesis the MCs are constant. Introducing the various iso-revenue curves will induce the firm to produce OB' of B and OA' of A. Here as before the comparative cost approach does not yield the internal opportunity cost, but they can be determined as previously (see above p. 136).



To elaborate furthermore our analytical apparatus, we must note that the comparative cost technique can deal with monopolistic cases and changing marginal costs.

(1) Constant costs (MC) and monopoly in selling:

The only modification to introduce is in the iso-revenue curve which is now convex toward the origin. With given L, the R revenue curve is the highest attainable. Consequently OB' of B will be produced and OA' of A. From this combination, the internal opportunity cost



of producing one more A (and less B) is equal to R - R'. The case of increasing marginal costs could be handled similarly, while utilizing our method established previously p. 137.

Multiproduct

No Interdependence

Short Run

(i) Fixed Proportions:

The short-run limitations here are concerned with funds (similar to L.R.) and also with inelasticity of the supply of at least one factor of production.

These new conditions will not basically change the principles which were devised in the L.R. case to deal with multiproduct. However, there should be some modifications of the costs data. Generally speaking, being interested exclusively in MC, the cost function used will be a variable cost one: it will include only the money variable costs plus the outside opportunity costs of the variable factors that are owned by the firm. The distinction between fixed and variable owned factors is that in the S.R. the period is too short for some factors to be hired to outsiders; these factors are the fixed owned factors. With this in mind, the case of fixed proportions can be disposed easily. The firm will be in equilibrium in S.R. when: $SRMC_{a+b} = r.MR^a + s.MR^b$, with the usual added condition that no MR is negative. From here on, the analysis proceeds as in the L.R.

(ii) Variable Proportions

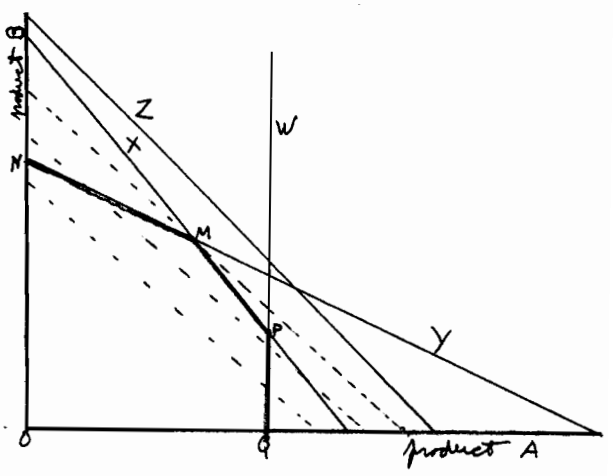
Here the time period has the same significance as in the case of Fixed Proportions (above). The same modifications of the costs data are necessary. The various diagrams remain essentially the same, except that the various BA curves pp. 136, 137, 142, will account for different quantities of product. For most outputs, the range of adjustment of factors is limited in such a way that it is now more costly to produce most outputs. This is the familiar proposition that the SRMC is different from the LRMC.

Here as before, money costs can be very misleading. The example on page 133 can be reworded in short-run marginal cost terms, and the same comparative costs apparatus can solve the problem of which product(s) should be produced. Furthermore, the assumption of limited funds (to build the L curves) is still more plausible in short-run than in the Long Run. Regarding the funds involved, the distinction is as follows: supposing the firm does not run into debts, in LR the L curve can encompass all the capital funds of the firm, while in SR only the working capital is available to build the L curve.

"Internal" opportunity costs, meaning the loss in revenue from switching from one product to another, are still present.

Regarding our general formulation (pp. 138-139) where many products (more than two) and factors are present, the

whole set of equations can be used after making the following adjustment: supposing factor x is the fixed factor in SR, we can still use the system of equations after having taken x off from all equations involving the x. Factor x now becomes a hidden factor, like the entrepreneurial factor in our LR models. The advantage of using general production functions as we did is that when one factor becomes fixed, we can use the same technique simply by ignoring factor x. Yet this advantage becomes a deficiency by removing from the picture the "hidden" factors which have a definite influence on the production function. This is a case where linear programming can usefully be employed. It is possible in such a framework to include in to the analysis both the fixed and variable factors (within limits) and the many products that may be produced. An extreme case is provided by the following illustrations. The firm can produce many products e.g. A, B. All its factors are in fixed supply and are distributed over four departments, x, y, w, z. The method of production requires that each product goes through all four departments. Each process involves constant returns. Each department has a maximum capacity with respect to the products. In the diagram (65, pp. 797-802) these capacities are pictured by the boundary lines, X, Y, W, Z, where ONMPQ is the area



of possible productions. The maximum production is naturally at the boundary NMPQ which becomes a product transformation curve. Its slope (when it has one) does not yield the ratio of opportunity costs of the two products but as before (see before p. 137) they can be discovered as follows. The opportunity cost of producing one more unit of product A is equal to the slope of the NMPQ curve multiplied by P^b. (When this slope exists). Superimposing a set of iso-revenues curves will give M as the optimum combination of A and B that should be produced. Here facilities X and Y will be used to capacity while facilities W and Z will be partly used. The solution being at a corner of the products transformation curve, this prevents the marginal approach to be able to evaluate the opportunity cost of one product into the other. However through L.P. the right solution is obtained.

More complicated problems involving three products and three processes could be handled by transformation planes and iso-revenues surfaces, but for m products and n processes, we must part with geometry and use algebra. At the moment the Simplex Method is the best method to deal with these problems. This algebra as contrasted with the former one, used in the previous section, leads surely and swiftly to the optimum solution, and especially permits us to recognize it when we meet it (28, p. 82ff). Finally activity analysis (132, chap. III) can deal with joint inputs, joint outputs, and joint demand.

Multiproduct

Interdependence

Long-Run

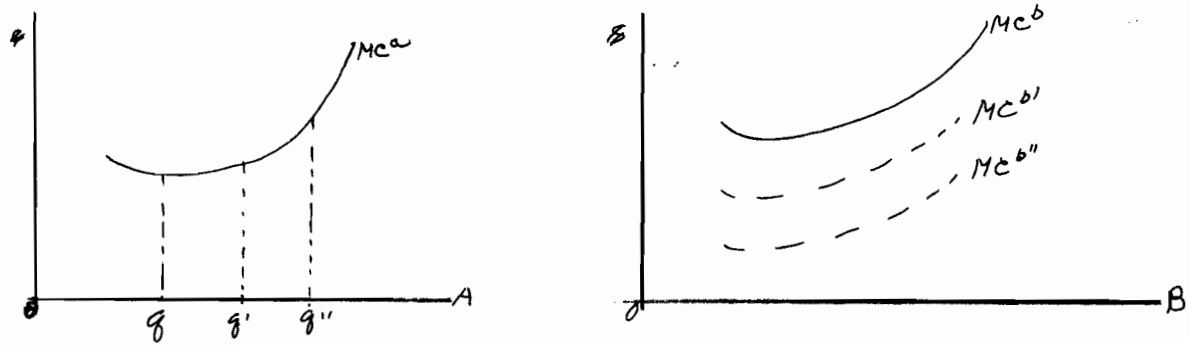
Cost Interdependence

The firm studied previously had none of its costs or revenues of any of its products interrelated in any way. The reasons why firms would still produce many products in these circumstances was that the firm had super abundant funds which could be spread over many products or that some entrepreneurial excess capacity was present. However now is the time to study the firm where it has other reasons to produce simultaneously many products.

Let us begin by cost interdependence. In such a case the situation is thus: when two products A and B are independently produced their MC curves are MC^a and MC^b. However when produced simultaneously, two cases are possible:

(1) Complementarity in Production

Each time one more unit of A is produced, it becomes easier to produce B; this is translated by downward shifts in the MC^b curve. What happens can be pictured as follows:



where MC^b corresponds to output Oq of A

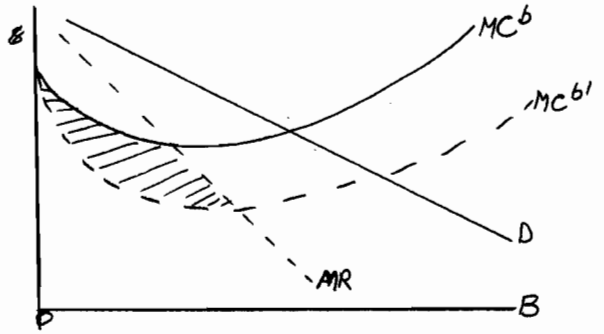
$MC^{b'}$ corresponds to output Oq' of A

$MC^{b''}$ corresponds to output Oq'' of A etc....

Similar relations may hold for changes in output in B which would then induce shifts in the MC^a curve. Symbolically we have taken:

$$MC^a = f(Q^b) \quad \text{and} \quad MC^b = f(Q^a).$$

From the point of view of A, the "complete" cost of changing its output by one unit is equal to the direct MC^a minus the change in profits occasioned by the downward shift in MC^b. Symbolically: the complete marginal opportunity costs are: $MC^a - \Delta P_?^b$, where $\Delta P_?^b$ is equal to the hatched portion of the diagram at the right. When one more A is produced MC^b shifts to $MC^{b'}$, output of B is increased (not necessarily by one unit of B), and total profits out of the sale of B are increased by the change in total revenues minus the change in total costs. To be in equilibrium the firm should then increase output until its <u>complete marginal opportunity costs</u> are equal to marginal revenue:

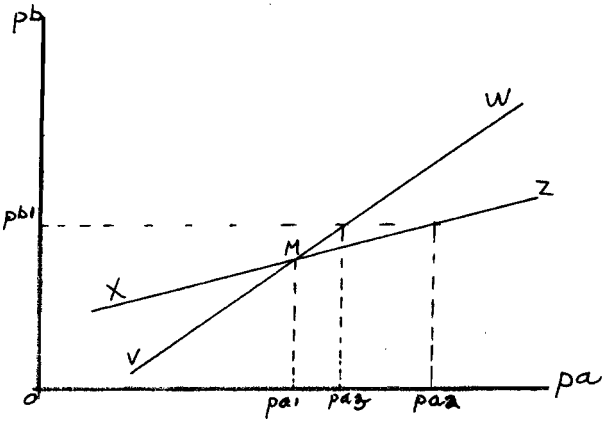


$$MC^a - \Delta P_?^b = MR^a;$$

note here that the internal opportunity costs $\Delta P_?^b$ are negative.

This is not the end of the story, because since the

output of B is changed to reap additional profits, complementarity in production between B and A will in turn shift MC^a, thus starting a new chain of reactions. Whether these things go on indefinitely or not is not essential to obtain a point of equilibrium because it is likely that MR^a and MR^b will diminish with each increase in output. The condition of equilibrium is that there is a point where aggregate $MC^{a+b} = MR^{a+b}$ (these expressions including all the cross-effects). This point can be determined with the familiar tool of price reaction curves (48, pp. 280-281). The general shape of the curve showing the reactions in prices of B to changes in prices of A are depicted by line XZ. Proof: If P^a is increased, less A is produced, then MC^b shifts upward, reducing output of B and increasing the price of B.



We can then match all the prices of A with all the prices of B that will maximize profits as far as A is concerned.

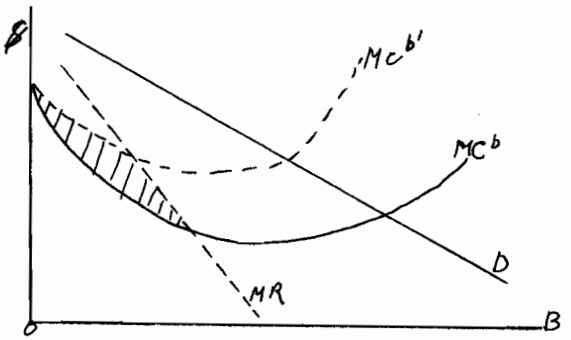
Similarly, A being complementary to B, any rise in price of B will reduce quantity of B produced and thus shift upward the MC^a which will have the effect of increasing the price P^a. Line VW is then upward sloping when depicting the reactions in price of A to changes in the prices of B.

The system is in equilibrium at M because if the price

of A is changed from P^{a1} to P^{a2}, the reaction in B will be to name P^{b1} (all this read on XZ). However, if P^{b1} is named, there will be a reaction in P^{a}, (now we read along VW), P^{a3} is then named. This in turn will induce another readjustment in the price of B, until we move to M, where reactions stop because both prices are then compatible.

(2) Rivalry in Production

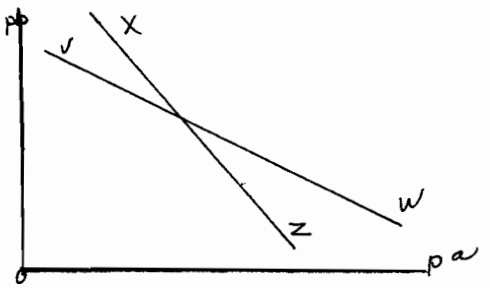
The case of rival production involves just the opposite of the previous case. As more of a product A is produced it becomes more expensive to produce B. As more A is produced, MC^b shifts upward with each increase in the production of A. If one more A is produced, MC^b shifts to $MC^{b'}$, inducing a readjustment in output of B and in the profit situation of B, equal to the hatched area in diagram at right. Consequently the opportunity cost of producing one more unit of A is $MC^a + \Delta P_n^b$ (where ΔP_n^b is equal to the loss in profits in B). Provisionally, then, the firm should increase the production of A until $MC^a + \Delta P_n^b = MR^a$. Yet, as in the case of complementarity, this is not the end of the story, because readjustments in output of B induce in their turn downward shifts in the MC^a, thus increasing the revenues in A by ΔC^a, where ΔC^a is the change in costs in A due to a change in output



in B. This will induce a further change in output in A which will induce a further change in output of B. Shortly speaking we can say that the complete opportunity costs of a new unit of A is:

$MC^a + \Delta R_h^b - \Delta C^a$; Output in A should then be increased until this expression is equal to MR^a.

The same process is true from the point of view of B. As in the case of complementarity, the whole system will come in equilibrium when $MC^{a+b} = MR^{a+b}$. As usual, price reaction curves can be constructed and their intersection will bring the set of prices in A and B that will be compatible. The shape of these reaction curves will generally be the one illustrated in the diagram at the right. They will slope downward because, for instance, if the price of A is increased less of A will be sold, and produced,



thus inducing a downward shift in MC^b, which will induce a decrease in the price of B. This is similarly so for the other point of view.

Before passing to interdependence in sales, let us pause to note that the introduction of the particles ΔR_h and ΔC clearly shows that direct monetary costs are insufficient to solve the problem here. "Internal" opportunity costs are essential where there is interdependence.

Furthermore the inquiry was pursued without introducing

the effect of limited funds. When limited funds are introduced, the firm will reach equilibrium when the revenue per marginal dollar invested is the same in any direction (on product A or B), and, for instance, the marginal opportunity cost of producing one more A (under the hypothesis of limited funds) is equal to the revenue in B that a sum k of costs invested in B would produce; where $k = MC^a \pm \Delta p_n^b \pm \Delta c^a$, and the revenue that B would produce is computed with the old MC^b. (The signs depends on whether the relations are of complementarity or rivalry).

Multiproduct
Interdependence
Long Run
Sales Interdependence

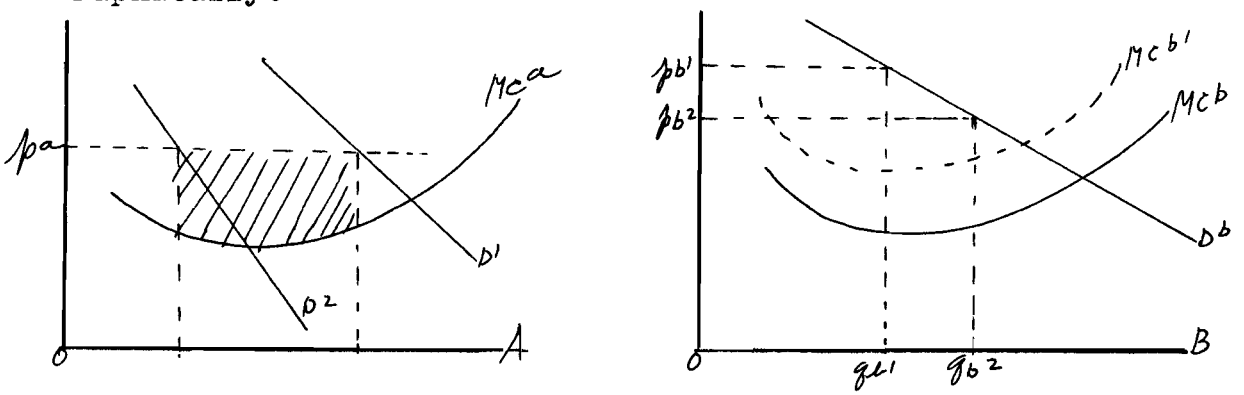
Products may be related not only in production but also in selling. For instance, the more we sell of A, the easier or harder it becomes to sell B. That is each time we sell one more unit of A, there is a shift in the demand curve of B; and vice versa. Let us study the sales relations while stipulating independence in production. Products can be related in their sales either through complementarity or through substitution.

(1) Substitution

When two products A and B are substitutes (in selling), each time the price of B is decreased the demand curve for A

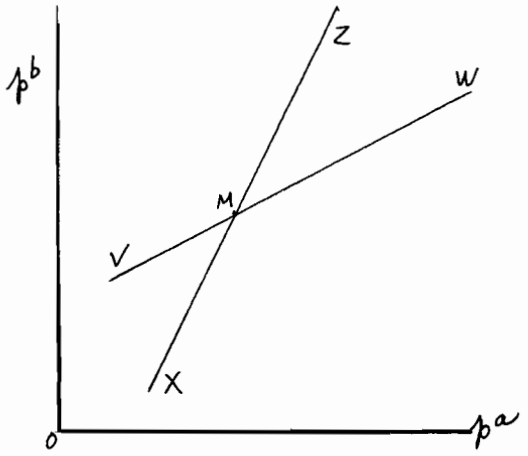
shifts to the left. Because, in this case more of B is sold, and consequently, B and A being substitutes, the customers now need less of A. Their demand for A decreases. All this, if the price of A is kept constant (48, p.p. 286-294).

Graphically:



When the firm expands sales and production of B by reducing the price B from p_{b1} to p_{b2}, the demand curve of A shifts from D_1 to D_2; reducing the profits earned in A (while keeping the price of A constant) by the hatched area ΔP_{π}^{a}. Consequently the <u>opportunity cost</u> of producing one more unit of B is not only the direct cost MC^b, but MC^b plus the foregone profits on A. This opportunity cost $MC^{b'} = MC^b + \Delta P_{\pi}^{a}$. Keeping the price of A constant would induce the firm to expand the production of B until $MC^b + \Delta P_{\pi}^{a} = MR^b$; or it would induce it to construct a new MC curve for B which would be a marginal opportunity cost curve $MC^{b'}$. Of course, the same mechanism can be worked out from the point of view of A, where reasoning on the assumption that the price of B is constant, would induce the firm to expand the production of A (and change its price) until $MC^a + \Delta P_{\pi}^{b} = MR^a$.

However, when considering the system as a whole, we are then in a world of mutual reactions, each adjustment in one product inducing another readjustment in the other. Yet from each product point of view, for example, A, there are as many optimum prices of A as there are possible prices of B. This is similarly so for B. This enables us to construct price reaction curves. For instance, XZ shows the



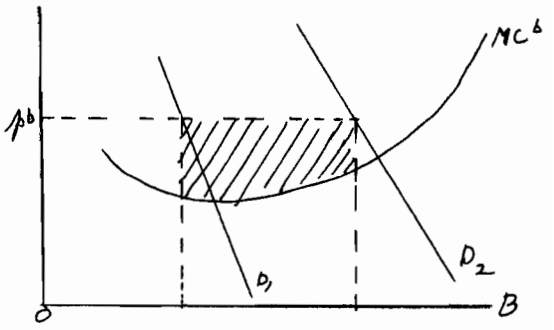
adjustments in P^a as P^b (considered constant from A) is allowed to vary. XZ is sloping upward, because as P^b increases, the demand curve for A shifts to the right, presumably involving a readjustment in the price of A that will increase this price. This occurs especially if the induced changes in profits in B (through readjustments in A) are quite constant in value. Similarly, WV which shows reactions in the price of B as the price of A (considered constant from B) is allowed to vary. If the price reactions lines have the shape depicted above the system will come in equilibrium at M, where the prices are mutually consistent. To save space we will not deal with the mechanism of equilibrium at point M; we may refer the reader to our previous section of costs.

As mentioned by Coase (48, p. 288) these reactions

lines will probably not have these simple shapes, since the magnitudes in foregone profits ΔP_π are a function of the MC and of the relative shifts in demand. Yet the main item to remember from this analysis, is that to take the right decision the entrepreneur needs an estimate of the "internal" opportunity costs ΔR_π. Foregone profits are what matters.

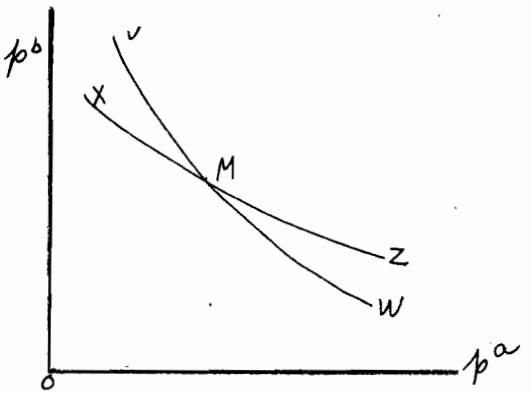
(2) Complementarity

In this case the reasoning is similar to our previous case of substitution except that the terms are reversed. The products are complements when if more of A is used, the demand curve for B shifts to the right, and vice versa. Now, more A will be used only if the price of A decreases. Consequently p^a decreasing entails a rightward shift in B. If now we are considering the cost of producing and selling one more unit of A, this is equal to MC^a minus the change in profits in B due to the shift in demand curve B as a reaction to selling more A, the price of B being kept constant. For a given price of B, then, the price of A should be such that:



$MC^a - \Delta P_\pi^b = MR^a$ (whereas before ΔP_π^b is an internal opportunity cost measure; negative in this case). This enables us to draw a new $MC^{a'}$, a marginal opportunity cost curve, (not drawn here, but lower than the direct cost one). Yet, changing the price of B will determine a reaction in A, and

thus a new price of A (figured out with a new marginal opportunity cost curve). Exploring a whole series of prices of B (being the constant) yields a corresponding series of prices of A. This is our first reaction curve XZ. It will slope downward, because the higher the price of B (here the constant) the lower the demand for A, and thus the lower the price of A, especially if the changes in profits in B (due to the readjustment in B) are enormous. XZ then shows the readjustments in the price A as the price of B (constant) varies. A similar set of relations can be worked out from the point of view of B, the price of A being considered the constant. Output of B is expanded (for each price of A) until $MC^b - \Delta P_\Omega^a = MR^b$. The collection of various prices of A, yields (through these reactions) a series of equilibrium outputs of A and consequently of prices of B. This is enough to construct the price reaction curve VW, which will slope downward for similar reasons as in the case of XZ. The system will come at rest at point M. As in the previous case, <u>marginal opportunity costs</u> are those that count.

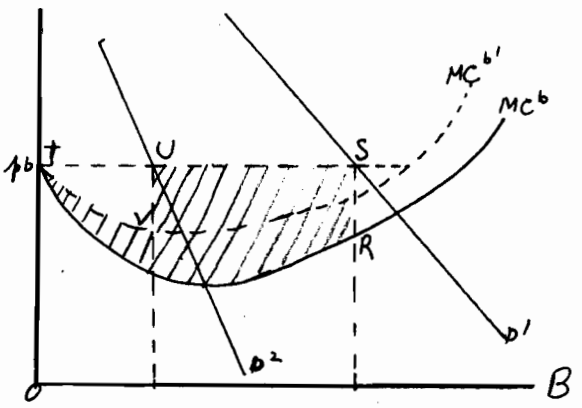


(3) <u>Simultaneous interdependence in sales and production</u>

Only a few cases will be explored here to give a general idea of the principle at stake.

(3i) Rivalry in Production and Substitution in Sales

To analyse this case let us suppose for the moment that the price of B is constant. The price of B being constant permits us to obtain a stable demand curve in A. From this the marginal cost of producing one more unit of A is: $MC^{a'} = MC^{a}$ (direct) plus the loss in revenue in B, (ΔPR^{b}) minus the beneficial effect of reduced output of B on the cost of producing A, (ΔC^{a}). This occurs where ΔPR^{b} is the loss in B due to two causes: (1) a leftward shift of demand curve of B, and (2) an upward shift in the cost curve of B, because more A is now produced.



This is equal to the hatched area or: QRS - TUV. This occurs also where ΔC^{a} is the change in total cost of producing A due to a reduction in output in B. This change in cost is a reduction in costs: $TC_1 - TC_2$. Consequently the production of A, given the price of P^{b}, should be extended until: $MC^{a} + \Delta PR^{b} - \Delta C^{a} = MR^{a}$.

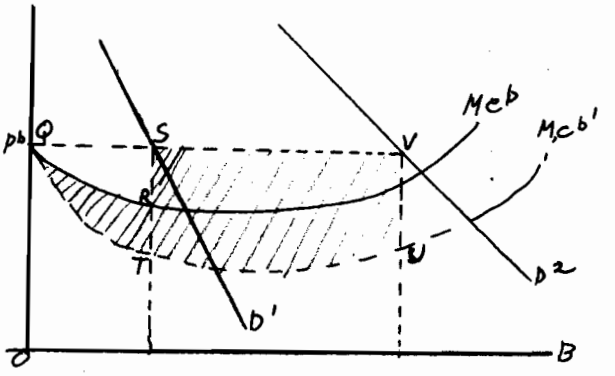
One P^{b} determines, through readjustment, an optimum P^{a}. If a whole series of P^{b} is explored this will yield a series of optimum P^{a}; and this gives us our first price reaction line XZ.

We could work similarly from the point of view of B, keeping the price of A constant now, and this would yield,

by exploring a whole series of P^a, a series of optimum P^b; plotting this information would yield the price reaction line VW. And as before the system will come in equilibrium at M. Generally, the greater the P^b (on XZ) the greater the demand for A, and the greater the price of A, if we suppose that $\triangle PR^b$ are very important and $\triangle C^a$ not important. This accounts for the upward sloping curve XZ which is similarly so for VW. It must be noted here that because the interactions are rather complicated and their importance not determined, the Marginal opportunity cost curves may have many shapes.

(3ii) Complementarity, Both in Sales and in Production

Suppose the price of B is fixed, this gives us a stable demand curve for A. From this we can calculate the MC of producing one more unit of A as follows:



$MC^{a'} = MC^a$ (direct) $- \triangle PR^b - \triangle C^a$. Where $\triangle PR^b$ is the gain in profits in B because: (1) more sales in B because the rightward shift in the demand for B, and (2) the downward shift in the cost curve of B. This is so because of complementarity. And where $\triangle C^a$ is the reduction in the total costs of producing A because now more B are produced: $TC_1 - TC_2$. In the diagram above $\triangle PR^b$ is the hatched surface of QRS - QUV. As in previous cases, price reactions curves

can be drawn, and compatible prices and outputs determined.

All these examples illustrate thoroughly how the notion of opportunity cost, especially the internal opportunity costs: ΔPR and ΔC, are important in order to make decisions. Straight money costs, even when added with outside opportunity costs for the owned factors, are clearly insufficient. The above analysis was pursued under unlimited funds, yet when limited funds are introduced, the rule devised previously in the case of no interdependence (pp. 138-140) applies also in this case.

(3iii) Price Discrimination

It remains before closing this section, to study a simple case of jointness both in production and in sale: this is the case commonly known as price discrimination. It is well known (178, chap. 15) that if the markets for a homogeneous commodity can be separated and if the particular demand curves have different elasticities (at the same price), it might pay the firm to practice price discrimination. This model is very well suited for the multiproduct firm whose products are not too different in production but appeal to different classes of customers because of small differences. This is the classical example of the wine merchant who put a big barrel out of the view of his customers has installed three taps labeled X, XX, and XXX, the Xs supposedly denoting the quality. He might sell these "brands" of wine at different prices, thus practicing price discrimination. More

plausible examples are found in the fields of automobile, home appliances, clothing, where one basic model can be adorned with various gadgets which in the eyes of the customers change the model from a De Luxe model to a Super DeLuxe.

One way to bring these various products to a model of homogeneous product price discrimination is to subtract from the demand curve of each product the special costs imputable to the differentiation. We are left with demand curves for products A, B, C ... N which although different in sales, possess a common production facility and a common MC (the MC of the basic product).

In this case the complete opportunity cost of producing and selling one more unit of A is equal to the basic marginal cost plus the highest profit that could be earned by this unit of product in any other markets B, C, ...; $MC^a = MC + \Delta PR^b$ or c, or d ... . The firm will be in equilibrium when $MR^a = MR^b = ... = MC$; this is brought about by adjusting the respective outputs of A, B, ... N.

Multiproduct
Interdependence
Short-Run

Short-run considerations do not alter the general method of approach. It might create accounting problems regarding the isolation of the respective MC, but besides that there is nothing special. After having converted the data

from LRMC to SRMC, the analysis proceeds as above. It would then be tedious to repeat the analysis.

Limitations:

However successful this type of analysis may be, even when based on marginal opportunity costs, it can become inoperative in the following cases:

(1) Where instead of two products we have to deal with n products, all interrelated in a different way, some in rivalry, others in complementarity in production some as substitutes, others are complements in sales, while others are completely independent.

In this case the geometry cannot be used at all, while the preceding algebra becomes almost impossible to build up, especially the marginal cost equations for each product, because of the difficulty of including properly all the interrelationships among products.

(2) When the production function involves only a limited number of processes; the isoquants and the transformation curves will entail corners which prevent the use of marginal analysis.

(3) Finally, when the time period is so short that all the factors are in fixed supply. Marginal analysis cannot determine which products to make and which factors or facilities to use, and to what extent. This last sentence implies that there may even be cases where it might pay not to use all the facilities at the disposition of the firm.

These three cases can be dealt in one way or another by the methods of Linear Programming or activity analysis.[16] Linear Programming is still in the process of development and has not reached the stage where it could handle easily all the problems of the firm presently handled by economic theory. As was said previously, increasing returns are not amenable to it, neither are the problems of exchange (132, p. 7); the monopolistic competition case can be dealt with only in a restricted way and the effect of changes in demand are not clearly brought about.

Yet, if we use both tools of analysis, marginal analysis and linear programming, depending upon the situation faced, we realize that the opportunity cost is the only sure concept to use in the conduct of a firm. This is obvious in short period where by hypothesis some factor(s) are in short-supply. But this is also true in long periods, since there also are some hidden fixed factors such as the entrepreneur whose opportunity costs must be computed and of course alternative products are present. Other types of costs might be misleading. "What the firm is not doing but could do is frequently the critical cost consideration" (62, p. 260). In other words, "bygones are bygones" is the golden rule on which to run a business. By contributing this notion of opportunity cost the economists contributed the essential

---

[16]See previous illustration above p. 127.

philosophy of business administration. This fact is not publicized enough today. Businessmen[17] are sometimes ready to admit that economists could be used as sales and costs forecasters, and market researchers, but these are mainly statistical jobs. The economists are also credited with a 'comprehension' of a market mechanism, and even with picturing the network of relationships in which the firm is willy-nilly engaged with the other economic units or with other sectors of the economy, for example, Labor, Government, etc.

Without being unimportant, many of these contributions of the economist are probably very weak in the business world: his demand theory is practically worthless (146); his market theories are made either of untenable behaviour hypothesis (Cournot, Edgeworth) or of unrealistic approximations (marginal cost pricing) or of rules of thumb (full cost pricing).

On the other hand, his "opportunity cost" theory is becoming more and more scientifically sound, because it is now based either on calculus or on linear programming or activity analysis, and game theory.[18] The day is coming when it will be the job of the economist to tell the businessman

---

[17]The following remarks were written with an eye on: Proceedings Operations Research Seminar, (152).

[18]About game theory, my incompetence in this matter forbids me to work out its relevance in terms of opportunity cost analysis.

how to operate his business.[19] The opportunity cost principle is the basic principle to operate any business; as such, the contributions of the economist is outstanding. As the ghost of Keynes is haunting most government planning bureaus, so the ghosts of Jevons and Wieser should be found in business offices. I say "should" because unfortunately not all businesses use the golden "rule" especially in short-run situations where rule of thumb methods are used.

At the moment we have no scientific measuring rod to determine whether or not the firms are efficient. At the moment the public and the directors of the going concerns judge the performance of the various businesses through the

---

[19]The usefulness of economics has been changing at a tremendous pace. Not long ago, Pigou (165, p. 137) held that "it is not the business of economists to teach woollen manufacturers how to make and sell wool, or brewers how to make and sell beer, or any businessmen how to do their job." This statement is still respected today; see Dr. E. F. Beach (152, p. 171). Yet today many economists are telling oil producers how to mix the ingredients to make gasoline, also how to program their capital expansion and how to keep inventories. Consequently the economists are doing exactly what Pigou forbade them to do. They are doing that because they are better equipped and they have better methods of investigations, for example, linear programming, game theory, information theory, decision theory, inventory theory.

The economist takes today decisions on scientific reasoning; the same decisions which formerly were made by the businessmen on a few hunches, and regarded at that time as something like black magic.

Recently Business Week (May 7, 1960) reported under the heading "Can Machines Replace the Boss?" that Prof. H. A. Simon advanced the idea that many management functions would be programmed and performed by computers. This means that important decisions pertaining to output, quality, market strategy, will be made by the economist and his team of experts, i.e. engineers, mathematicians, lawyers, marketing experts.

financial reports of the accountants. This is very unfortunate because the accountants use costs concepts that are not relevant to evaluate the efficiency of the firms, especially in the field of multiproducts.

Many accountants in this situation try, through what can be called rules of thumb, to saddle each unit of output with a portion of the costs such that in the end all the costs are apportioned. The technique of opportionment is pompously named: cost accounting and forms an important part of the curriculum of accountants. Unfortunately, as far as the management of the firm is concerned, much of this is useless. Because the costs that are relevant to the firm, and especially the multiproduct firm, are the opportunity cost. The relevant costs in economics or in the management of a firm are neither the past, actual, nor even expected money flows, but the foregone alternatives. The accountants deal mainly in post-mortem exercises: the firm having sold so many units of products A, B, C, ... N; the exercise consists in allocating, through various indexes, the overhead costs to these products. It is agreed that it is a legitimate occupation to allocate direct variable costs; however as compared to the problem at stake this amounts to a triviality. The real problem is with the overhead common costs. Does it make sense to allocate them? Would it not be better to work from the opposite, instead of saddling each unit of output with costs; would it not be preferable and more logical to figure

out the opportunity cost of each unit of factor, and then allocate it properly?

Before answering that, we should explore the nature of the costs used by the accountant:

Firstly, the accountant uses only historical flows of money as the basis to determine costs. Things that were never the subject of a contract involving flows of money do not enter into his notion of costs. In other words, alternative courses of action are never the subject of his calculations. He is preoccupied with what happened or is happening or will happen but not with what could have happened or could happen.

Secondly, costs occur for him either through: (i) use or (ii) the flow of time. For instance raw material becomes a cost if used in producing a commodity. If not used, its value piles up in inventories, and these historical money flows are not allocated. Similarly for a piece of equipment, if not in use, (and abstracting of time) it is not a cost. It becomes a cost only as far as used (the historical money flows become costs according to the rate of use or at the rate of time flows). Thirdly, the costs for the accountant are objective. The rules for determining them follow from the principle outlined above. They also become standardized.[20]

---

[20]However the concept of costs itself is something quite hazy for the accountants, for example the following confession: "It is not amiss to the student of cost accounting, at the outset, that there is no such thing as an exact cost or true cost." Van Sickle, (215, p. 5).

Fourthly, for the accountant, costs are caused by output; consequently he has the tendency to calculate the cost per unit of output.[21] Suppose[22] that during an accounting period (quite short) fifty raincoats had been produced at historical money flows of $20.00 for raw material and $80.00 of overhead common costs. Total costs, then, are $100.00. Suppose these raincoats, each using the same amount of raw material were labeled and sold as follows:

20	labeled	A	were	sold	at	$4.00	a piece
10	"	B	"	"	"	3.00	"
10	"	C	"	"	"	2.00	"
10	"	D	"	"	"	1.00	"

The accountant will say that we made more profits on the raincoats sold at $4.00; more exactly that we made a profit of $2.00 each, while on those labeled B we made only $1.00, and that we broke even of the C ones, and lost $1.00 each on the D coats. This kind of reasoning is completely irrelevant from the point of evaluating the performance of the managers. First of all because these different "brands" of raincoats may simply be a strategy (price discrimination) to raise $140.00 in revenues; secondly, cancelling the production of the D raincoats will simply increase the overhead that is

---

[21] "Cost accounting work, when there are joint products, is essentially one of a mathematical computation of unit costs." Neuner (148, p. 453).

[22] This example is adapted from Thirlby (212, pp. 32-49). A similar example is worked out in Neuner (148, pp. 452-456).

allocated to the other types of raincoats.

Allocating, burdening or charging particular units of products with certain historical costs is then solving a pseudo-problem.

Comparatively, the costs of the economist are:
(1) Subjective: costs are not the objective historical money flows but the displaced opportunities, and this value has to be subjective because as far as the decision-making is concerned the opportunities to become costs must be seen or perceived by the decision-maker. Consequently, no outsider can determine the cost in this situation because his economic horizon and knowledge of opportunities might be different. He might "see" more or less than the present manager. Consequently, the accountant uses only the rock bottom value or outside opportunity cost of money itself.
(2) Costs are ephemeral: there is no such thing as historical opportunity costs. A cost occurs only when a decision is taken; once a machine has been bought bygones are bygones, the only sensible thing to do is to make the most of it. Each time a new decision is taken, even if the accounting period is not finished, a "new" cost is set up, and a new period begins. In fact this analysis leads us to a series of periods during which we have no money costs to guide us.

The Work of the Accountant

Having dealt with the nature of cost for the accountant, the following shows how they apply their notion of cost:

Dealing exclusively with money flows, they divided them into

(1) Overhead or common costs to all products, for example, salary of the president.

(2) Traceable direct costs; costs that are indisputably imputable to a particular product.

Some of these costs are for items that will last more than the present accounting period; they are then expenditures for fixed assets, thus conceptually fixed costs. Some other items will not last the complete accounting period, they are classified as liquid assets, and their corresponding expenses are variable costs.

Summarizing we have:

(1) Common costs (overhead) which are (a) fixed or (b) variable,

(2) direct costs (traceable) which are (a) fixed or (b) variable.

As said previously it is agreed that (2b) should be allocated to the product; this is done simply by subtracting these costs from the sales price. Regarding (2a), they should be taken care of only in the planning stage. In the ordinary conduct of the business they are bygones and any post-mortem profit and loss statement allocating the depreciation charge to a particular product is information of no managerial value.

With regard to (1a) there is a fortiori no sense in allocating them; none of the indexes, for example, floor space,

direct labor, and, worse, sales figures, make sense.[23] "These bases or formulae are essentially arbitrary in the sense that equally defensible formulae can give quite different cost estimates."[24] Particular costs are consequently interdependent and thus lead nowhere near a determinate solution.

About (1b) two points might be given:

In the planning stage they must be taken into account. Particular marginal costs are isolated by taking derivatives from the total cost surface (a surface build à la Stigler) (204, p. 131). No fancy indexes should be used to proceed to the allocation.

In the short-period, if changes in overhead can be correlated with changes in output of product(s) then these changes should be allocated to the products responsible for these changes in cost. This is nothing new, because in this case (1b) becomes (2b).

The above applies only in the cases of variable proportions in the product mix. When fixed proportions prevail, it is absolutely useless to allocate costs among products.

---

[23]"The generally accepted costing procedure...where joint products are produced...is to allocate the costs of production on the basis of sales prices..." Keller (116, p. 207).

[24]Joel Dean, (62, p. 315). Also "The greatest influence upon the determination of what comprises cost, however, is the individual in charge of the accounting procedure in each business concern..." (Van Sickle 215, p. 5).

All this does not mean that the businessmen are duped by their accountants. The reason is that first of all the decisions are rarely taken from past accounting data. Some decisions involving great sums of money were taken on the basis of computations covering no more than the back of a pack of cigarettes. Secondly, the accountants themselves are conscious of the fact, yet they are not free to vary their methods. Up to now their main efforts were devoted to coping with legal and tax requirements. Only really big ventures can afford managerial accountants. If these people want to be effective they will have to follow the economist's footprints in this matter.

If the accountants are unable to determine the efficiency of the firm, can we rely on the economist to do so? In other words can the economist compute adequately the opportunity cost of one more unit of a particular product or, alternatively, the opportunity cost of using one more unit of resource in a particular employment? Historically, the economists were initially quite successful in establishing the proposition that allocation of resources is optimum when the coordination is done by the "impersonal forces of the perfect market". Within the firm, it is thought that the resources are optimally allocated when the manager follows the rule MC = MR. However, within the firm what assures us that the data used by the managers are the "best" ones?

It was quite easy to show that the market price mechanism would bring optimum allocation of resources because the market mechanism could "evaluate" the "opportunity cost" of a unit of factor or product. But since there is no price mechanism within the firm, but instead authoritative coordination , (47, pp. 332-335)$^{w\ell}_{x}$ can only hope that the entrepreneur is efficient. It will be shown later[25] that the conventional production function does not lend itself easily to the drawing of a map of the optimum isoquants. The effect of using standard production functions lead to the situation where:

(1) "the economist's prescription...are non-operational"

(2) "his descriptive assumption of profit maximization is not subject to direct empirical test" (119, p. 70).

For instance in the case of p. 127 (above) although the production function is seen as a family of processes it is unable to show that two processes should be used simultaneously.[26] What is needed is a method of imputation within the firm such

---

[25]See below p. 184 ff.

[26]Dorfman, (64, p. 15): "The production function thus is a tool for exhibiting and comparing different but related processes. What it fails to present adequately is the consequence of using several processes in parallel, and such combinations of processes are characteristic of modern industry."

that this computational procedure will "imitate the action of a competitive market with regard to resources not actually traded." (119, p. 99). In other words we need a method to figure the value of marginal product for each unit of factor as more or less of this factor and other factors are used: and this process of imputation is made through L.P. in its dual form. These imputed values become accounting prices or internal opportunity costs, (15, p. 851) and with these prices we are able to use the rules of managing the firm since we are in possession of the elusive prices (costs) that the cost-accountant was trying to determine. In other cases, for instance where there is an infinity of processes marginal analysis is able to do the same job.

L.P. and marginal analysis are simply mathematical methods and as such are complementary to each other.[27]

Yet we have seen that there are cases where L.P. is superior to the marginal analysis whether in its pure money form or when it includes the outside opportunity costs. For instance when one or all the factors are in limited supply, the marginal analysis stipulates that the fixed factors have no cost in the period considered. While with L.P. we are

---

[27]Allen, (11, p. 608): "The theory of the firm can proceed either in marginal terms following the 'classical' or traditional pattern or in the large by means of linear programming and activity analysis."

able to compute the "internal opportunity costs" of such fixed factors. For instance in the case of limited processes, there might be an opportunity cost (in the form of foregone production of commodity A) in using the scarce resources in some process instead of in some other. Using our former example and diagram of page 127 it makes no difference, as said previously, whether the firm chooses to move initially on process 2 or process 1. Suppose it has initially devoted some resources on process No. 2 such that OB' of Y and Ox_1 of X have been spent; being at point M, the entrepreneur may ask whether he should go on with process No. 2 or take another process; BB' of Y remaining to be used if necessary. If BB'of Y is used in process 2 (helped by the necessary X) R_2-R_1 of product A can be produced. On the other hand if BB' of Y is used in process No. 1 R_3-R_1 can be produced. Consequently if process 2 is used all the way R_3-R_2 of commodity A will not be produced and this represents the internal opportunity cost of utilizing BB' of Y in process No. 2. It is these hidden internal opportunity costs that are crucial for management decisions, not the outpocket costs, especially in very short periods.

This is the point that has been emphasized repeatedly in the present section on Costs. We are accustomed since Coase (47, p. 334) to saying that the "distinguishing mark of a firm is the suppression of the price mechanism," at its place a planning system is functioning. As seen previously

this should not be interpreted as a complete doing away with prices. To allocate properly its resources, especially in short-run, the manager needs to know the "internal prices" of his resources, these prices being the "internal" opportunity costs of these resources.

It is unfortunate that at present the textbooks in price theory rarely spell out adequately the implications and fundamental importance of the opportunity costs. As when reading Marshall, when digging for them we always find them, but they are not in the fore-front. This is sometimes misleading. For instance, Boulding (24, p. 480) when studying the supply of the competitive industry pictures firms with various productivities (cost curves). This enables him to show that the lowest money cost firm is the first firm to "come" in the industry and the highest money cost firm, the last to come in. However to have different costs the firms have entrepreneurs or other factors which must differ in ability. The "best" entrepreneur runs the firm which has lowest money cost. In Boulding's example a small reduction in price might oblige the highest money cost firm to be the <u>first</u> to leave the field. Yet if opportunity costs of these exceptional factors are introduced this solution might be wrong. In fact the first firm to leave or the last to come in might be the lowest cost firm because presumably it is the one with the most valuable entrepreneur and thus with the highest opportunity cost. Managers with high opportunity costs are much more "mobile" than others (178, p. 106).

HISTORICAL SECTION

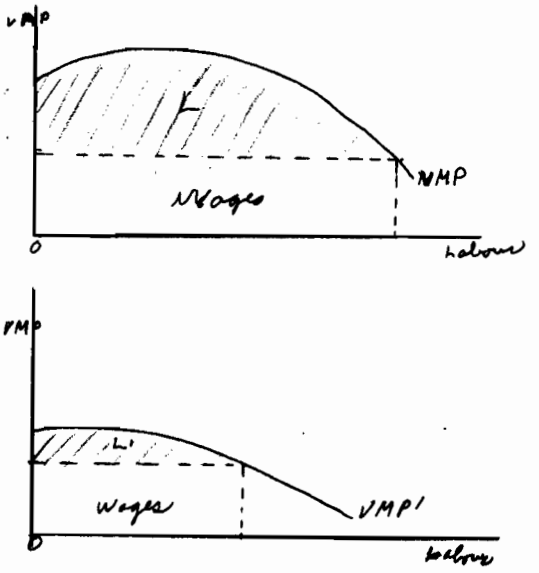
Marshall distinguished between money and real costs and correctly stipulated that "when considering costs from the point of view of the capitalist employer, we of course measure them in money" (p. 291 of Principles). Opportunity costs show up in the form of imputed costs for the managerial services of owners, and interest not paid on subscribed capital. He says (136, pp. 395-6) that these items should be included into the "ledger" of the going concern.

In another connection, (p. 361) while deciding what to grow on a piece of land, Marshall begins by stating correctly that "if one crop shows signs of being more remunerative than before relatively to others, the cultivators will devote more of their land and resources to it." This is how Marshall comes to study a multiproduct firm similar to the one we outlined on page 133 above. Yet in a footnote (p. 362n) he shows that he might have disagreed with our opportunity cost approach. He reproaches Jevons for having debited to a rival product the rent per acre which a first product was yielding. However his own solution utilizing the rent per acre posterior to the switching of one acre to a new employment, might not be correct also.

Our solution in terms of profit foregone as an opportunity cost might be much simpler:

When the original amount of land is devoted to product

A, an amount L of rent is earned (see the hatched area in the diagram at the right). When one acre of land is transferred to the production of product B, two things happen:



(1) the production of A is reduced and its price rises,

(2) a different VMP curve (lower) is ruling.

In the new situation, a new total rent L' is earned on the land devoted to A. (see hatched area). The opportunity cost of one acre transferred to B and which should be charged to the production of B is L - L'.

Consequently Jevons would be right only in the case of pure competition matched with constant marginal product for any output or any amount of land.[28]

The short-run supply of the firm is also well treated (pp. 298, 299, 300, 311-12). Nowhere is it identified with the portion of the SRMC curve over its AVC. Spoiling the market is for Marshall a powerful deterrent to using the MC curve as a supply curve over all its length. He is then aware of all these hidden opportunity costs, without incorporating

---

[28]The same topic has been discussed in the literature from another angle by Brown (32, pp. 498-500).

all of them into the SRMC curve.

What is said of Marshall applies to many other writers in this period. The real costs are no more in use and the analysis proceeds in money terms. Yet the main contribution of the period is the concept of costs as a sum of foregone utilities: the opportunity cost. The expression 'opportunity cost' was invented only towards the end of the period.[29] but the concept dates further back. Naturally it is an Austrian, Wieser, in 1876, who must be credited with its formulation.[30] Wieser and his followers extended to the cost side the subjective theory of value. If the value of a unit of a product is measured by its marginal utility its costs can similarly be measured by the foregone utilities of the other commodities that could have been produced with the same resources.[31]

The first English presentation of this theory occurs in Wicksteed (202, p. 43). Opportunity cost as a mode of "costing" is present in most of the authors when they come to value the productive services according to the value of their marginal product in their best alternative employment. Yet

---

[29]Most people, for example Viner (216, p. 468) and Schumpeter (185, p. 917) attribute it to Green.

[30]Schumpeter (185, p. 1044, n. 45) remarks that everything that could be said about opportunity costs had been said by the Austrians, notably by Bohm-Bawerk.

[31]This is sometimes called the Wieser Law. "Such a conception of costs first systematically developed by Wieser and made familiar in English speaking areas by Green, Wicksteed, Davenport, Knight, Anderson..." Robbins (172, p. 2).

in many writers, for example, Bohm-Bawerk, Marshall, Wicksteed, Edgeworth,[32] it is not explicitly mentioned.

The exact merits of Walras are harder to establish. When he solved his general equilibrium system, he was implicitly using a generalized system of opportunity costs. Maximum aggregate output, with fixed supply of resources (a basic assumption in Walras), requires that only the best alternatives are used.[33] Solving such a system of equations (including the demand equations) amounts to using implicitly an opportunity cost approach.[34]

Stigler (202, p. 230-231) mentioned that this can also be interpreted as containing also a real cost theory since, if one of the products is "leisure", the problem of allocation becomes one of choosing between commodities and leisure. Yet if both are valued in utilities, the formulation in alternative

---

[32]"The doctrine of alternative or opportunity cost is not explicitly mentioned by Marshall. The idea of competition of various uses for given resources is, of course, fundamental to his treatise and it may well be described as the underlying theme of the entire discussion in Book V" (202, p. 66). Also: "alternative cost theory...is strongly implicit in Edgeworth's writings" (202, pp. 111-112).

[33](202, p. 264). The same may be said of Wicksell whose "essential point (in the allocation of resources) is that the return from a given resource will be equalized in all its occupations...which at the same time maximizes total product."

[34]In such a system "the idea of mutual determination arises, goods constitute reciprocal demands for one another. Or one good limits the demand for another, since in order to obtain one good, we must give up another--the 'opportunity cost idea'. Haney (98, p. 782).

uses is here the counterpart of a real cost approach where the utilities of the commodities are in competition with the disutilities of working (not enjoying leisure). In such a perspective one cannot agree with Stigler (p. 231) that there is an opposition between real costs and opportunity costs.

Walras' analysis is not only "compatible" with opportunity costs in general but also with the various ways of formulating them: calculus, and linear programming. In his equations he used straight marginal analysis, but his fixed coefficients, his production function (homogeneous of degree one) provide an ideal setting for linear programming (106, p. 208).

Pareto came to contribute when he worked on the mutation of the classical comparative costs into the Austrian opportunity costs. Pareto, feeling the inadequacy of the comparative costs in real terms, added opportunity costs under the name of "ophélimité indirecte" (217, p. 509). This "ophélimité" being identical to the benefits that could be reaped on the land (which has no real cost) used to produce goods for exports.

In the U.S., Davenport (98, p. 519; 185, p. 917.n.20) reduced costs to the opportunities (utilities) foregone by entrepreneurs and is credited with the most exhaustive treatment of the subject.

A choice field of application for this theory was the problem of multiple products. The first important study of

multiple products is found in Walras and Pareto (157, p. 609). When they worked a general equilibrium system they had to consider the possible production of n goods, where the conditions of sale and production are those of independence. (See Pareto p. 608). They isolate the MC of each product and the firm reaches equilibrium when $MC^a = MR^a$, $MC^b = MR^b$, etc..

Similarly, Marshall's firm is not a single product firm. As elsewhere Marshall sacrifices rigour in favor of realism. This he does in Book V, Chap. VII, where he correctly establishes the condition of equilibrium of the firm in the case where it is possible to separate the marginal cost of each product. However (p. 328) he does not resist the temptation of attempting to allocate overhead, for example, "general expenses of the business", on the basis of the relative importance of the respective prime costs. He realizes that this is difficult to do, especially in the case of loss-leaders and generally when products are interrelated both in selling and in production. Joint production is well illustrated, (chapter VI), by the example of beef and hide, wheat and straw, and is pictured--in the framework of both fixed and variable proportions, (for example, p. 323, when he talks of "judicious crossing in sheep",) where it is possible to discover the respective costs.

Joint demand is less well represented when one realizes that Marshall means indirect or derived demand for some factors

of production who usually work together, for example, carpenters and plasterers. On that topic the best piece of analysis of the period is found in Edgeworth (73, pp. 84-91) where interdependence in production is characteristically written as:

"$f(x + \Delta x, y + \Delta y) - f(x + \Delta x, y) - f(x, y + \Delta y) + f(x,y) < 0$", (p. 86).

This is the case of rivalry in production while presumably if the expression > 0, would mean complementarity in production. On page 85 he builds a cost surface similar to the one used above on p.214, where the partial derivatives measure the marginal cost of any product. This type of analysis was coming in the forefront because of its connection with applied economics, mainly in the field of price discrimination by the Railways. In this matter Edgeworth mentions J. S. Mill, Biskerdike, Maurice Clark, Taussig and Seligman who studied this aspect of the theory.

Since then, the concept of opportunity cost has had an active career. Its incorporation into new tools of economic analysis of the firm, such as linear programming, surely opened new horizons to the possible use of the principle. Similarly the incorporation of the opportunity costs into the indifference curve analysis of international trade by Haberler was another step; this is not to mention other items, such as the user-cost of Keynes.

CHAPTER V

THE FIRM 1870-1920

THE LAWS OF RETURNS

SECTION I. PRODUCTION FUNCTION

Previously, in Production Theory, it was pointed out that the authors did not see the production problem in isolation but as a part of a greater problem of distribution. To be able to solve this general problem, one must put it into its proper perspective: the production function. With the production function, the problem of distribution (how much each unit of factor is worth), and the problem of production (how much of each factor to use), are solved simultaneously.

Generally speaking, this can be done by what is called the use of the marginal productivity theory. Yet, before developing such a theory, the authors needed, at least, to have an "inkling" of a production function. Here is why: the production function describes "the relationship between inputs of productive services per unit of time and output of products per unit of time."[1] One production function itself can be represented by one product contour map, or a map of isoquants.[2] Up to quite recently, the traditional theory of

---

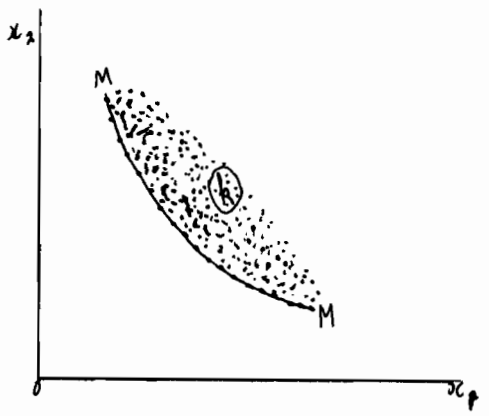
[1]Stigler (203, p. 107). See also the exposition of Schumpeter, (185, pp. 1027-1032); Carlson (34, pp. 14-15); Dorfman, Samuelson and Solow (67, p. 201).

[2]See appendix C,

the firm assumed, more or less, that the production function was continuous, so that the production surface was continuous, and therefore that the isoquants were continuous curves convex to the origin. One advantage of continuity is to permit the use of calculus, the main tool to expressing the various theorems of the traditional theory.

However continuity implies the following properties of the production function: (1) any combination of factors corresponds to a definite quantity of product, (2) there is an infinite number of these combinations, (3) where the production surface is reduced to a map of isoquants, it is assumed that the isoquants are the optimum ones. What precedes characterizes the conventional production function in such a way as to render it non-operational in many circumstances.

1) For instance, the production function has been defined above as the set of relations between inputs and outputs.



In a sense this is quite vague. To produce a given output there might be many ways to use the inputs. There might be n combinations of inputs that will produce k units of output. Diagrammatically the set k of of points in the plane X_1OX_2 might the locus of the combinations of X_1 and X_2 that will

produce at least k units of output, and curiously in some cases, not more.[3] In terms of conventional analysis many isoquants all representing the same output, can be drawn in the area k. However in conventional analysis only the points on MM are used, the others are discarded as non optimum. This discarding is regarded merely as a technical procedure.[4] Yet in terms of data usually available to the economist, they never present themselves under the form of an optimum production function as pictured by the textbooks, but as a limited series of production possibilities through a variety of processes (119, p. 70; 67, p. 131). Consequently what assures us that the "right" discarding has been done? These remarks suggest that actually one important problem has been whisked away by the conventional production function, because as such it presupposes that the problem of physical maximization (the retention of the "best" combination to produce k of output and the discarding of the others) of output out of given inputs has been solved (119, p. 70). In other words, an exercise in physical imputation is necessary beforehand. This process of imputation can be brought about only by a

---

[3]When a minimum amount of vitamins is required, this can be satisfied with a given combination of goods; yet if more food is absorbed, the excess of vitamins is lost because the organism has only a certain capacity of absorption.

[4]"The purely technical maximization problem may be said to be solved by the very definition of our production function" Carlson (34, p. 15).

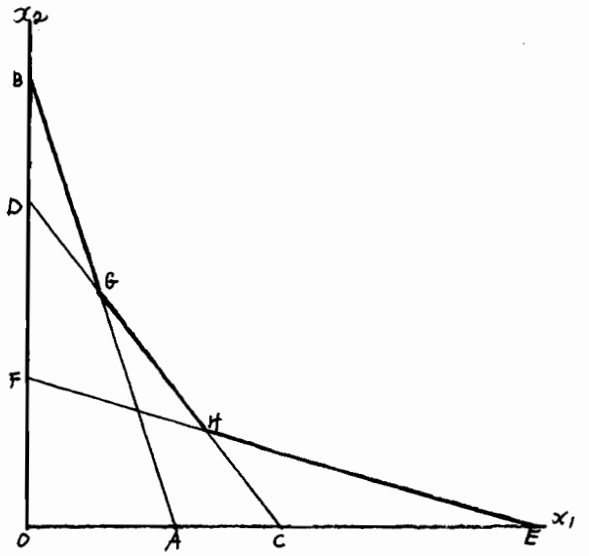
method of calculation that will evaluate "internally" the opportunity costs of the various quantities of factors.

The following classical example (the Diet problem) helps illustrate the point (67, p. 9 ff.). The problem is to construct the optimum isoquant that will depict the various combinations of factor x_1 and x_2 that will produce at least k units of output. Suppose the following problem exists: We can feed a hog and make it ready for meat on two types of food x_1 and x_2. However to transform the hog into ham! certain nutritional requirements must be satisfied, for example, calories, vitamins, minerals. Let's label C_1, C_2, C_3 the minimum physical requirements (in the respective quantities of units of the nutritional requirements). Now, foods x_1 and x_2 have different capacities to satisfy these requirements.

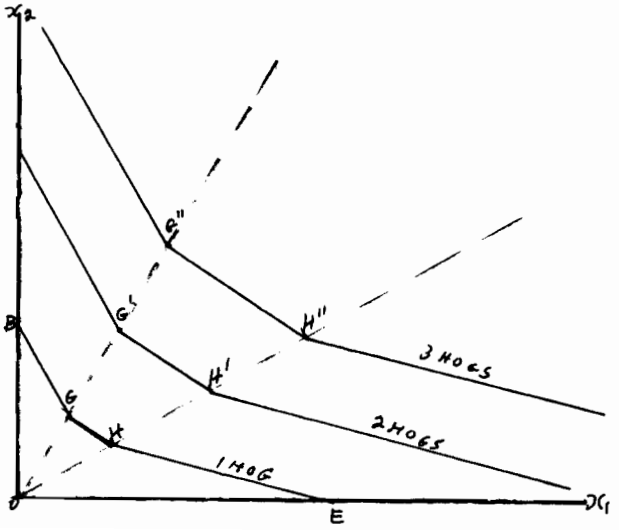
Suppose $C_1/d_{11} = OA$, the quantity of x_1 (in physical units) necessary to satisfy requirement C_1. Where d_{11} is the capacity per physical unit of x_1 to satisfy the requirement C_1.

Similarly $C_1/d_{21} = OB$, the quantity of x_2 necessary to satisfy requirement C_1. Graphically AB represents the various combinations of x_1 and x_2 that will satisfy the minimum requirements of C_1.

Similarly we could compute C_2/d_{22} = OC; and C_2/d_{22} = OD where DC are the combinations of x_1 and x_2 which will satisfy requirement C_2. And C_3/d_{13} = OE and C_3/d_{23} = OF where FE represents the combinations of x_1 and x_2 which will satisfy requirement C_3. From this it comes out that only the combinations of x_1 and x_2 on or in the half-space at the right of BGHE will simultaneously satisfy the nutritional requirement to feed one hog ready for market. BGHE is then the isoquant representing a product of one hog. Note that we can now without any danger eliminate all the combinations at the right of BGHE as non-optimum.



The method of handling the problem above, is called linear-programming. It can also solve similar problems involving m inputs and n restraints. Being in possession of this basic isoquant, we can after that build out the complete map of isoquants as demanded by the general production function. If constant returns to scale are

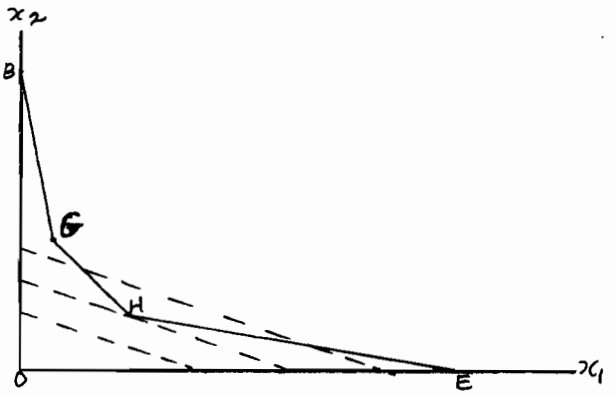


postulated we will have: OH = HH' = H'H'' = etc... where the first isoquant accounts for 1 hog, and the second for 2 hogs. Varying returns to scale could also be pictured either by using non-linear programming. The only restriction is that we cannot use the first approach in cases of increasing returns. The reason is that mathematically linear programming depends on the property of convexity, and convexity and increasing returns are inimical.[5]

We are not at the end of our journey. Marginal analysis determines which combination of factors should be used by introducing isocosts in the picture. However, even if there is an infinite number of ways to produce a given output, marginal analysis might not work properly if the number of restraints is finite, because in this case the isoquant will then have kinks. In this case, there will be no point of tangency between the isocost and the isoquant but points of contacts; although obviously the point of contact depicts the cheapest combination of factors to produce a given quantity of output. At each of these corners (points of contact) $P_{x1}/P_{x2} \neq MP_{x1}/MP_{x2}$, consequently no solution is possible in marginal terms, because there is no ratio of MP_{x1}/MP_{x2} at

---

[5]Linear theory since it is based upon convexity cannot be extended, without losing most of its virtues, into the realm of increasing returns." Hicks, (101, pp. 688-689).



H that can satisfy the definition of a slope or marginal rate of substitution of x_1 for x_2. Another awkward case for marginal analysis is where the isocost is tangent to the isoquant over a long stretch for example, HG. In this case there is a range of indeterminacy, any combination of x_1 and x_2 in that range will minimize costs, or maximize output with given costs.

2) A second source of trouble for conventional marginal analysis, if found when given, even in the long run, the number of methods or processes that can be used to produce a given output, is limited. Suppose, with two factors, x_1 and x_2, we can produce output k of a certain commodity only by combining them in four possible proportions: and that there are no economies of scale.

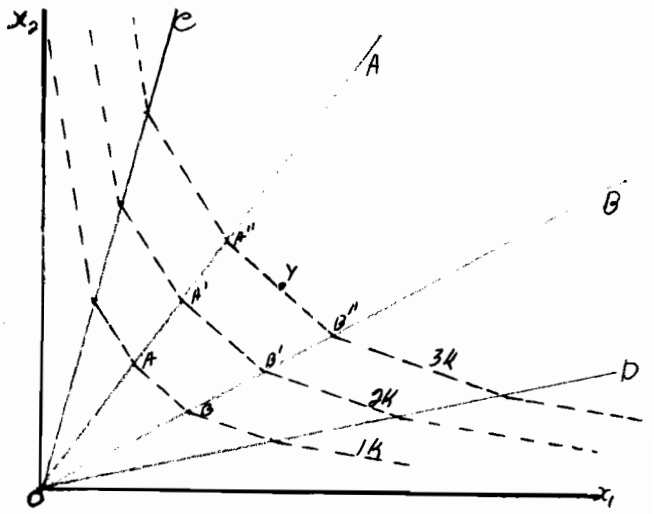
Process A, requires $3x_2$, $1x_1$, produce 1k

Process B, requires $1x_2$, $2x_1$, produce 1k

Process C, requires $7x_2$, $1/4x_1$, produce 1k

Process D, requires $\frac{1}{2}x_2$, $5x_1$, produce 1k

We can construct the following lines of output (65, pp. 804 ff.). OA represents the various combinations of x_2 and x_1 that corresponds to the first possible process; OB corresponds



to the other possible combinations of x_1 and x_2, etc. On these lines let us mark various outputs, 1k, 2k, 3k, ...that correspond to the various combinations of x_1 and x_2. The distance between each point on the various lines is constant on a given line (because of constant returns to scale) and depends upon technical considerations. Let us note that the dotted lines joining points of equal output are <u>not</u> isoquants. Because by hypothesis, for instance, point Y represents a combination of x_1 and x_2 which is not permissible; it does not correspond to a valid proportion. Here as before, the introduction of isocosts will not yield a solution in marginal terms because of the corner solution, if the slope of the isocost is different from the slope of the isoquant. If the slopes are identical, the solution is either at point A or B but not on the intermediary points, because of the hypothesis about the permissible proportions. However, through linear programming, it might be possible to combine x_1 and x_2 simultaneously in two processes. If this is so, all the combinations of x_1 and x_2 on the line AB are now possible.[6]

---

[6]See before p. 127.

Substitution in terms of processes, is something which the conventional production analysis did not explore explicitly.

3) The third point is that the conventional production analysis does not take explicitly into account the factors of production that are in fixed supply. This is so especially in long run situations, where for instance the entrepreneurial factor never appears as one of the axes. Sune Carlson (34, p. 14) defines his production function "in relation to a given plant; that is certain fixed services." Linear programming essentially takes into consideration any factor that might be in short supply, and especially these ones.[7] In a sense, the linear programming approach has the tendency to reduce all problems to a short-run approach (106, p. 109). However, being able to compute the internal opportunity costs of the "hidden" fixed factors through linear programming, even in the planning stage, (conventional long run) is surely a definite advantage.

Now whatever the basic property of the production function (continuous or discontinuous), mathematically the function can either be homogeneous, thus satisfying the Euler Theorem, or non-homogeneous; it can also be linear or non-

---

[7] "On the other hand, in the programming formulation most explicitly when the restraints are linear, the quantities of the fixed productive factors are central to the problem, because they are essential data in determining what the firm can and cannot do while in the conventional formulation these same fixed factors are regarded as being somewhat aside the problem, just because their quantities are fixed and predetermined." (67, p. 202).

linear. The economic implications of these varieties of the production function can be fully appreciated only by studying the laws of returns as will be done below.[8]

In those days, there was a marked preference, for example, by Walras, Wicksell, Wicksteed, for using a homogeneous function of degree one (185, p. 1040 n. 31). Finally, it is easy to agree that the function is stable only if the "stage of the arts" does not change.

Logically, the production function is prior to marginal productivity theory because before being able to isolate the contributions of a factor at the margin, we must first build the function in which it is used. In mathematics the partial derivative of a function with respect to a variable is usually considered after the function of which this variable is a part of has been established. In other words, it is only when economists have visualized the production process as a process where output results from functionally related inputs, that they were led to conceptualize a principle of substitution, or a marginal productivity theory.

Leon Walras[9] seems to be the first to "give a symbolic

---

[8]To make life easier for us, we will use exclusively continuous functions.

[9]Stigler, (203, p. 110). Yet most people agree that in fact the "inklings" of the production function are a bit older. "Von Thunen had already deduced empirical production functions for his famous estate" (ibid., p. 110). And Schumpeter (185, p. 1032) adds that "the production function itself appeared in 'classic' times under the name of the State of the Arts."

statement of the production function...which is usually written, $P = \phi(A,B,C,...)$, where P is Product and A,B,C,... are productive services, all per unit of time." Yet Walras' production function is one which Schumpeter (185, p. 1034) calls degenerate because it used fixed coefficients of production where no substitution between differentfactors or production is possible. He later (185, p. 1034) in 1896 at the suggestion of Barone introduced variable coefficients in his 'equations de fabrications' and then only "where he is concerned with economic progress".[10] In the meantime Marshall "did not avail himself of the production function explicitly" (185, p. 1032), while, on the other hand, Wicksteed wrote one; similarly Wicksell used one which was homogeneous of the first degree. "By 1900 then the production function had established itself, as a result of the efforts of many minds, ... (Berry, Edgeworth, Marshall, Barone, Walras, and Wicksteed)." (185, p. 1032).

SECTION II: SUBSTITUTION PRINCIPLE

Once the authors were in possession of such an apparatus, a little experimentation introduced them to the notion of substituability between factors. Here "the contribution

---

[10]Prof. W. Jaffé in (220, p. 527, note 3). These tergiversations are quite an enigma because very diverse opinions have been expressed on this subject. For instance Prof. Jaffé (ibid., p. 527) says that Walras proceeded initially in this way to make the exposition more simple, while Hutchison (110, p. 205) says that it might as well be through ignorance that Walras did that.

consists in the realization that the proportions in which productive agents may be combined to secure the same product are variable.... The significance of the principle of variation of proportions is apparent. It leads directly to marginal productivity theory of distribution. Until the principle of proportionality was fully developed furthermore, no satisfactory solution to the problem of resource allocation was possible. Finally, as long as discussion ran in terms of fixed proportions between productive agents (or as long as the question was ignored), the individual firm could not be used for purposes of analysis" (202, pp. 149-150).

The classical theory had incorporated a certain amount of substitution between the factors; for instance, Ricardo recognized some substitution between machinery and labor (see his famous chapter). Moreover, on a given quantity of land, it was known that by varying the proportion of capital-labor to land, diminishing returns would ensue. But the classical theory never envisaged many factors competing to be used in the same operation.

Menger seems to be the first to mention the principle of substitution, he did it in 1871.[11] Following him, Wicksteed,

---

[11] See for instance: Biet (21, p. 49); Schumpeter, (185, p. 917.n.19); Stigler (202, p. 150): "This formulation of the principle of variation of proportions as a general rule governing all resources is one of Menger's greatest achievements, one which he is not required to share with either Jevons or Walras."

Marshall and Pareto succeeded in freeing the theory by using variable coefficients.[12a] Wicksteed was the first to formulate the clearest exposition of the principle of substitution.[12b] The position of some others, for example, Clark, is obscure (21, p. 49).

With the production function plus the substitution principle we are getting near marginal productivity theory. Only one necessary step now remains to be taken: to substitute factors at the margin and isolate the corresponding change in output imputable to the variation in the quantity of the factors. And this step, the economists of the 1870s were in position to make. Indeed they had, as said before, all the constitutive elements for a marginal productivity theory: (1) With the principle of substitution they were in position to establish what would be called later a law of variable proportions; if substitution is possible that means that the proportions can be varied. (2) They also had a law of diminishing returns; this was supplying a mode of behaviour of the variables involved. Moreover, as will be seen later, a theory of marginal productivity leads to absurd

---

[12a] One cannot agree with Stigler (202, p. 346) that "Marshall never considers explicitly the scope of possible substitution between productive services." Although one must concede that the substitution in Marshall is not always at the margin (136, pp. 295-299) yet in some other places (136, pp. 295, 426) marginal substitution is present. [12b] In English "Wicksteed was to elaborate the central doctrine that all factors of production can be substituted for one another ..." (202, p. 333). "No part of the theory of production is more admirably developed than the principle of substitution.... He assumes the possibility of substituting any productive factor for any other productive factor within fairly wide limits." (202, p. 48).

results if the law of diminishing returns does not hold.

Yet, it took time to coordinate these elements. The theory of marginal productivity appeared only in 1891 (21, p. 32). Only in 1894 did Wicksteed make the statement that "in investigating the laws of distribution it has been usual to take each of the great factors of production...to inquire into...the special nature of the service that it renders and...to deduce a special law regulating (its) share of the product" (185, p. 1033).

The theory of production and distribution had, then, to travel a long way from the classical theory where Smith and its successors never faced the problem of determining how a product could be imputed to resources which cooperated in its production. Moreover, the marginal productivity theory is twofold: it has first to be expressed in physical terms and later in value terms. The value is obtained either by multiplying the marginal contribution in physical terms by the price per unit of the product or by the marginal revenue per unit of product.

The first pillar of marginal productivity theory is then physical. In other words, we must be able to appreciate the marginal physical contribution of each unit of factor. On this topic some work had originally been done in the preceding period in the form of the study of the laws of returns. However, these laws were not in a very good shape. Consequently, the first elucidation that had to be done was in the form of tidying these laws.

SECTION III: LAWS OF RETURNS

The classics were aware, of course, that increasing all factors of production will cause total product to increase. This is true at least for a certain range of values of the respective factors. They were also aware of some substitution, for example, the possibility of using more or less of one factor, and that this had effects on the quantity of product produced.

SECTION IV: LAW OF RETURNS: FIRST BRANCH

With this equipment the following experiment is possible. Suppose a production function involving two factors x and y. Keeping y constant (Land in the classical example), let us vary factor x continuously from zero to almost infinity. What has to be established is the particular way (if any) total product varies as a consequence. One way to characterize these variations is to relate (in terms of size) increments in x with the corresponding increments in TP (total product). The ratio of these two increments expresses the marginal physical productivity (MPP_x) of a factor x of production. However, as will be seen later, economically speaking, marginal productivity theory makes sense only if MPP_x eventually diminishes. In other words if economic theory is to account for the productive process it must use a production function which should at least have one property: it must be bounded. This is why, in all the derivations of the laws of returns, it is crucial to establish a law of

variable proportions that will be compatible with this state of affairs.

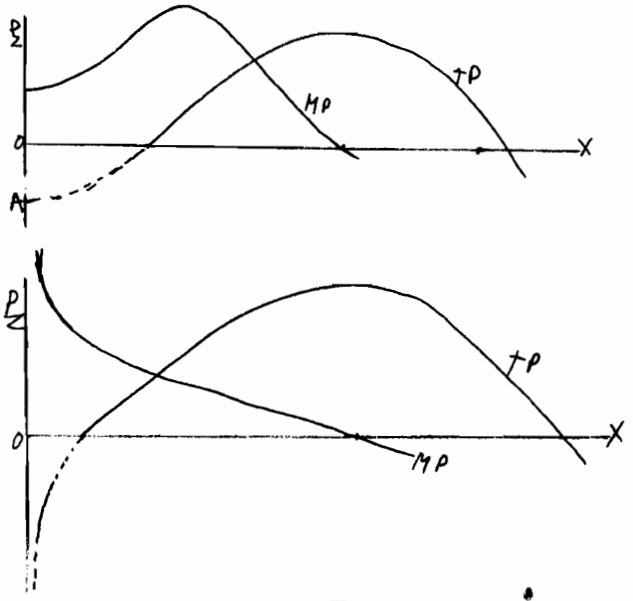
The boundedness of the production function is an external constraint imposed by the necessity of economic relevance. It is not suggested that any production function that may be thought of will automatically satisfy this requirement, but that if some do, it is only through accident not because of a mathematical necessity. As will be seen in the historical section below the authors of 1870s were misled by their insistance in devising an a priori proof of the necessity of such a bounded function.

With this in mind let us explore the types of functions that will comply with the constraint, taking good care to formulate it in the right manner, because the formulation itself is very important. The classicists, and expecially Ricardo, West, Cournot and Senior had formulated the law in an incremental form: they held that output will increase as more of the variable factor x is used, and that eventually these increments in product will diminish. Some, for example, Turgot, even held that there might be an initial phase of increasingly larger increments before the size of these increments start diminishing until they become nil or negative.

This general shape of the TP curve is accepted. It must be supplemented with other key values like what is the value of the TP when $x = 0$, etc.. Here observation shows that when x is zero TP can either be positive, zero and even

negative. Further, it might need many units of x for any production to occur at all. Consequently, the TP curve is bell shaped and starts anywhere, and then rises for some time. TP can increase with increasing increments (Turgot case) or at a constant or at diminishing increments (classical case). The classicists thought the last stage was absolutely essential to having a bounded function. However K. Menger (143) showed that this is not necessary when the increments are neither equal nor infinitely small.

If calculus is used (instead of reasoning the way the classicists and the authors of 1870s did with arithmetical finite increments) the MPP curve corresponding to a bell shaped TP curve may start any where on the Z axis (the product axis) since MPP is the slope of TP. Example: suppose the TP curve meets the X axis at A; it is obvious that at A, the curve TP has a slope which is greater than 0. And consequently, that the MPP curve should not originate at A but have a positive value when X is equal to 0. We can even consider a case where TP begins as an asymptote to Z and consequently where the MPP curve will start at Z equals to infinity. The custom however is to start the MPP curve for values of x where the



TP curve meets any of the axes. If this procedure is followed, the MP will rise for some time, and then eventually decreases. The MPP curve does not say anything new that is not already in the TP curve; it just says it differently.

There is also another way to express the data of the TP curve, and this is the form of the average product per unit of variable factor used. Historically speaking, the laws of returns were often expressed and are still today, for example in Lerner, (124, p. 155), in the average form instead as the incremental or calculus form. In this case $AP_X = \frac{TP}{X}$ and diminishing returns there means diminishing average product per unit of factor used. We had to wait for Edgeworth to disentangle the two formulations.

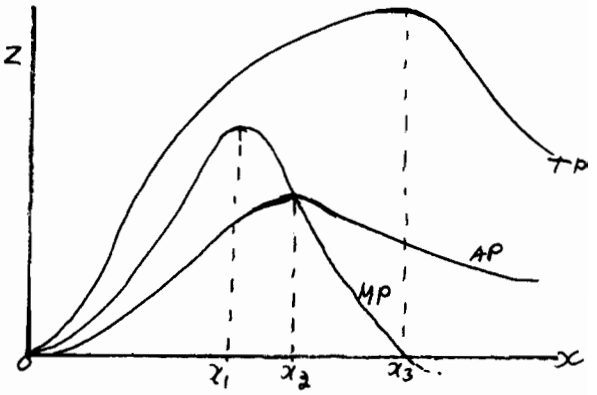
It is obvious that given a certain MP curve, the TP and AP curves follow as consequences. Actually the relations between the MP, and the TP and the AP curves can be rigorously illustrated with the aid of a homogeneous production of degree one like: $Z = \frac{x^2 \cdot y^2}{(x + y)^3}$. This function can legitimately be used because it is bounded on any of its cross-sections parallel to either of the axes.[13]

Since all the curves discussed above use the same set

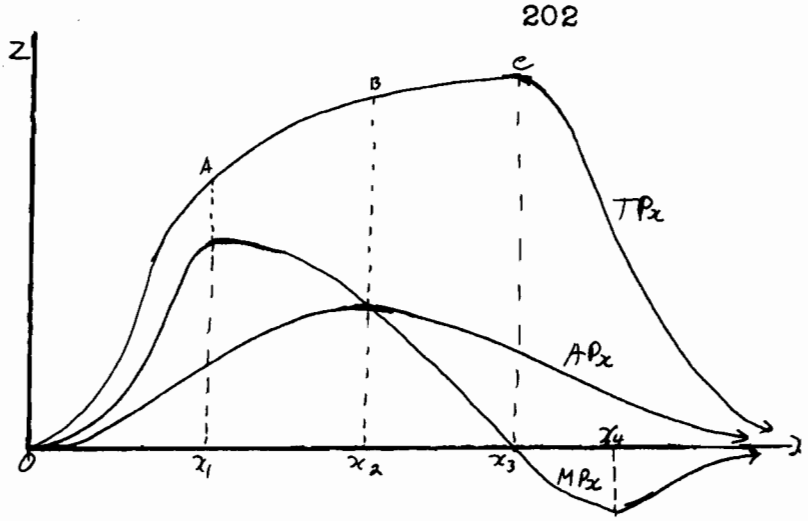
---

[13]The shapes of the TP, AP and MP curves are interrelated through their mathematical relations. For instance, it can be established through mathematical relations that the AP curve reaches a maximum when MP = AP; that AP increases only as long as MP > AP, and vice versa; that MP = 0 when TP is maximum, etc.. See (203, pp. 42-51) and (124, pp. 80-83).

of axes, they can simultaneously be plotted on the same diagram. In this diagram an homogeneous production function of degree one has been pictured, and consequently the TP curve starts at the origin. There are of course many other functions which do meet the requirement of boundness and yet are not homogeneous of degree one. These functions are legitimate but may present a variety of shapes. For instance in some cases the TP curve may originate on the X axis, denoting that a very small amount of variable factor is not enough to produce anything, and may even produce waste (represented by negative product, for example, the negative phase of the TP curve). In other cases the TP curve may originate on the Z axis, denoting that the variable factor is not absolutely necessary to produce some units of product.



Historically, the laws of return have been formulated indistinctly in two formulations: 1) in terms of AP and 2) in terms of MP. For instance in the classicists and others, the law of diminishing returns sometimes meant decreasing MP and at some other times diminishing AP. It can be shown that these two formulations are not equivalent if the increments in X are taken as very small. This can be seen by inspecting our next diagram which pictures the production function used previously. Here in the range of



$x_1 - x_2$ we have diminishing MP with rising AP. The two formulations are then not interchangeable as was assumed by Bohm-Bawerk and others, see below 255.

The complete set of values AP_x, MP_x and TP_x is as follows:

In the ranges

$x_0 - x_1$	: MP ↑	and	AP ↑	
$x_1 - x_2$	: MP ↓		AP ↑	(where ↑ means
$x_2 - x_3$	: MP ↓		AP ↓	increasing,
$x_3 - x_4$	: MP ↓		AP ↓	and ↓ means
$x_4 - x$	: MP ↑		AP ↓	decreasing.

These relations in no way contradict the former relations established by the average-marginal arithmetical relations, as can be verified in the diagram above where when: MP > AP then AP ↑ ; and where MP = AP, this implies that AP is constant, and where MP < AP, implies that AP ↓ . Also the diagram shows that MP ↓ is compatible with TP ↑ and that <u>MP ↑ is compatible with TP ↓</u> ; and AP ↓ is compatible with TP ↑ . Besides formulating the laws in forms amenable to computation by calculus (as above), there is another category of formulations in incremental terms. The difference between this category and the previous one is that the incremental type does not have to satisfy the requirements of calculus, for example, infinitely small variations, continuity, but accommodate itself

with discrete changes. This is the form which was used from the classical period up to the 1930s (143). We have, then, three ways to formulate the law of diminishing returns:

(1) in terms of diminishing marginal product, for example, where the variations in x are very small,

(2) in terms of diminishing average product, (variations in x being very small),

(3) in terms of diminishing product increments, (the discrete case).

K. Menger accounts for three formulations here: (that is there are three main types in which the people of 1776-1930 could formulate the law of diminishing returns in incremental terms).

(a) If x is the fixed factor, and y the variable factor, and h the increase in y, then:

$$Z(x,y_2 + h) - Z(x,y_2) < Z(x,y_1 + h) - Z(x,y_1)$$

(b) $$\frac{Z(x,y + h_2) - Z(x,y)}{Z(x,y + h_1) - Z(x,y)} < \frac{h_2}{h_1}$$

where h_1 and h_2 are the finite amounts by which factor y is varied.[14]

(c) $$\frac{Z(x,y''') - Z(x,y'')}{y''' - y''} < \frac{Z(x,y''') - Z(x,y')}{y''' - y'} < \frac{Z(x,y'') - Z(x,y')}{y'' - y'}$$

where y being the variable factor, and $\bar{y} < y' < y'' < y'''$

---

[14]This formulation is almost identical to the one of Edgeworth (73, p. 66, n.2).

for sufficiently large ys.

Let us now establish a few propositions regarding these various formulations:

If the production function is bounded and monotonically increasing (on any of its cross-sections) it is mathematically necessary that the first derivative of the function, for example, the MP, has eventually a decreasing phase. Similarly, such a function implies that its average value (Z/X) eventually decreases.

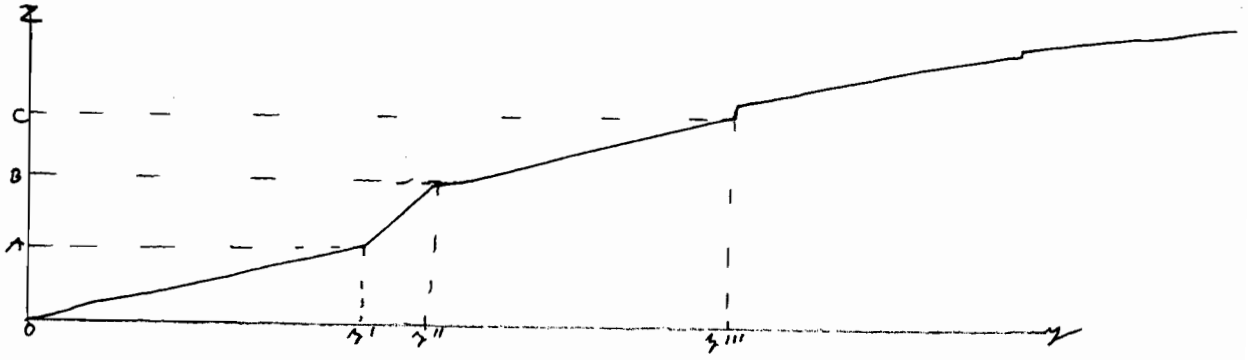
(2) The matter is much different if we formulate the law of diminishing returns in terms of increments (finite, unequal, as it was often done by Edgeworth, Bohm-Bawerk, etc.). In this case "the boundness of the production function implies neither the proposition of diminishing product increments nor the proposition of diminishing average product, but that it actually implies parts of both propositions." (143, p. 445). The reasoning is based on a function devised by F. Alt (143, p. 446):

Suppose the function Z(x,y) tends toward a limit of 1. Also that the series of values of the factors used are such: x is the fixed factor at 1, and

x:	y:	Z(x,y) =
$x' = 1$	$y' = n + \frac{1}{2^{n+2}}$	$Z(1,y') = \frac{2^n - 1}{2^n} + \frac{1}{2^{n+2}}$
$x'' = 1$	$y'' = n + 1$	$Z(1,y'') = \frac{2^{n+1} - 1}{2^{n+1}}$
$x''' = 1$	$y''' = n + 1 + \frac{1}{2^{n+3}}$	$Z(1,y''') = \frac{2^{n+1} - 1}{n + 1} + \frac{1}{2^{n+3}}$
etc..	....	....

Here it is easily seen that $y' < y'' < y'''$

Graphically the function will look like this:



Now if the law of diminishing returns is formulated in incremental terms instead of in calculus form, it should yield:

$$\frac{B - A}{y'' - y'} > \frac{C - B}{y''' - y'}$$, this corresponds to form 3(c).

However in the Alt function, the corresponding computations yield: $\frac{Z(1,y''') - Z(1,y'')}{y''' - y''} > \frac{Z(1,y'') - Z(1,y')}{y'' - y'}$

which is a negation of the law of diminishing returns. These computations show that although the function is bounded, this does not necessarily imply the proposition of diminishing product increment when computed as the authors of 1870s did.

Similarly this function also fails (in the incremental form) to satisfy the diminishing average product formulation. From the previous computations we have:

$x = 1$, $y = n$, then $Z(1,n) = \frac{2^n - 1}{2n}$. Consequently for any n, (n sufficiently large) we have:

$$\frac{2^n-1}{n2^n} < \frac{\frac{2^n-1}{2^n} + \frac{1}{2^{n+2}}}{n + \frac{1}{2^{n+2}}} \quad \text{or} \quad \frac{2^n-1}{n2^n} < \frac{\frac{2^n-1}{2^n} + \frac{1}{2^{n+2}}}{n - \frac{1}{2^{n+2}}}$$

which also fails to justify the law of diminishing average product.[15]

It remains now to study the basic hypothesis: the boundedness of the production function which up to now was introduced into the problem only as an external constraint. Is it possible to establish a priori the necessity of such a hypothesis? There were various attempts at doing so, especially in the 1870s. The classicists had used something like a reductio ad absurdum a priori proof. Not only was it accepted by most people, but many of them, for example, Wicksell, Von Mises, Bohm-Bawerk, Clark and Cassell tried to improve this "theoretical proof" of the law of diminishing returns. This type of proof is still in circulation today, for example, Stakelberg (201, p. 31 ff.) and is reproduced in Stigler (203, pp. 118-120).

It is usually presented as follows: the idea is to set up a case where the function is not bounded and later to show that this implies an absurdity. Supposed an unbounded total product curve based on the following table:

1	man	working	10	acres	of	land	produces	10	bushels:	MP	is	then 10
2	men	"	"	"	"	"	"	22	"	"	"	" 12
3	"	"	"	"	"	"	"	36	"	"	"	" 14
4	"	"	"	"	"	"	"	52	"	"	"	" 16
5	"	"	"	"	"	"	"	70	"	"	"	" 18
6	"	"	"	"	"	"	"	90	"	"	"	" 20
.												
.												
10	"	"	"	"	"	"	"	190	"	"	"	" 28

---

[15] What precedes is a free adaptation of Menger's reasoning.

This table presupposes that:

if 1 man working 10 acres produced 10 bushels, then any man working 10 acres can produce 10 bushels,

if 2 men working 10 acres produced 22 bushels; then each man working 5 acres can then produce 11 bushels,

if 3 men working 10 acres produced 36 bushels; then each man working 3 and 1/3 acres will produce 12 bushels;

when 5 men are used the critical figures are 2 acres and 15 bushels per man;

when 10 men are used the critical figures are 1 acre and 19 bushels per man;

such that when 10,000 men will be used on these 10 acres, each man utilizing 1/1000 of an acre could produce something like 10,000 bushels. Put to its limits this reasoning implies that a man restricted to using a flower pot could produce the world's output. The absurdity, according to the tenants of this proof, consists in having an increasing productivity per man as the quantity of the fixed factor per man diminishes. Yet this reasoning is not adequate because the amount of fixed factor per man is arrived at merely by an arithmetical deduction. It does not signify at all that the productive process consists in "allocating" to each man a piece of land, so that with increasing number of men this space is shrinking. It may be that the productive process is changing with each addition of new workers, for example, a better division of labor is enforced involving the specialization of each worker

Consequently, it is perfectly possible that an increasing total product occurs with each addition of new workers. When there is division of labor, many things can happen in terms of productivity. Imagine a farm where each laborer is an agronomist. It may happen that, with a new addition to the staff, a scientific committee is formed. Through team work, it may happen that the team is more productive than any of its parts separately, so that not only will TP increase, but also the Mp of the last addition in the labor force. Intuitively we may think that this cannot go on indefinitely, but who knows? A single project like the atomic bomb or the design of a rocket to go to the moon might involve continuous increasing TP and MP for each new scientist hired.

The classical 'flower pot' reasoning would hold if the productive process would not change with each addition of worker, that is each time a new worker is hired land is redistributed so that each worker works alone a smaller plot. But in this case the reasoning is tautological.

If one has to deal with this <u>a priori proof in the incremental form</u>, as used by the proponents themselves, the following methods devised by K. Menger can be used:

(1) The classical 'flower pot' approach:

When formulated in incremental form, since there is no necessary connection between boundness and diminishing increments, boundness cannot be inferred from diminishing increments and vice versa. It is possible that the production

be bounded on every piece of land, and yet not to have diminishing product increments.

(2) The 'flower pot' argument was reformulated in a more sophisticated form by Wicksell and reads as follows: "The validity of the law of returns on land does not require experimental proof, but presents itself as a logical postulate or corollary. For if it could be shown that twice the labor and twice the capital applied to a piece of land yielded twice the product and even better results would be achieved per unit of capital and labor if the labor and capital at hand were concentrated on half the formerly used area, the remainder to be used for grazing, forestry or some productive purpose which requires no labor" (143, p. 462). Symbolically Wicksell says this: $Z(x, \lambda y) < \lambda Z(x,y)$ which is simply saying in words that the production has the mathematical properties of sub-homogeneity, and super-additivity, and that land is an independent input. In other words, Wicksell says that the law of diminishing returns is compatible with the mathematical properties of a function, just mentioned. However he has not proved that any economic production function must possess these properties. Consequently he has not proved that inevitably all economic production must comply with the law of diminishing returns.

Moreover, it has been proved by Menger that even if a function

(1) is sub-homogeneous: $Z(x,y) \geq Z(\lambda x, \lambda y)$, where $\lambda > 1$,

(2) is super-additive: $Z(x_1,y_1) + Z(x_2,y_2) \leq Z(x_1 + x_2, y_1 + y_2)$

"the product of an outlay cannot be increased by division of the outlay into two separate parts,"

(3) is bounded; and (4) is monotonic with respect to x and y and even (5) super-homogeneous (that is $\lambda \geq 1$ in the equation of sub-homogeneity) there can still be cases where the property of diminishing product increments will not be satisfied. For instance, Menger proposes the function

$$Z(x,y) = \frac{k(k + 1)(x + y)}{(k + 1)(k + 2)}$$

where $kx < y < (k + 1)x$) and k takes any integer value) which satisfies all the five properties used above and yet the proposition of diminishing product increment is not satisfied.[16]

This should suffice to show that it is at the moment impossible to give an <u>a priori</u> proof of the law of diminishing returns. Yet Menger assures us that a weaker version of the law of returns involving only non-increasing (instead of diminishing) product increment can be proved with the aid of the

---

[16]Menger (143, p. 478) remarks that this type of reasoning is valid only if it is understood that we do not consider ranges where the production function is decreasing. This restriction permits the use of the property of super-additivity. This is obtained by stipulating that the entrepreneur will never use factors in the region where their marginal productivity is negative. Yet since it will be shown below that the entrepreneur may operate outside the ridge lines, this may restrict a bit the scope of application of this reasoning.

properties so far considered as long as they include the one of super-additivity. This might have been suspected by the reader since the Alt function used above p. 204 tended toward constant product increment. Consequently that function would have satisfied the following weaker formulation of the law of returns:

$$Z(x, y_2 + h) - Z(x,y_2) \leq Z(x,y_1 + h) - Z(x,y_1)$$

while the average product formulation would now read as follows: "for every piece of land, the average product of outlays is not increasing if the outlay exceeds a certain level" (143, Section D). We are left with only one basis to justify the existence of the postulate of diminishing returns: the empirical one.

What precedes accounts for the behavior of the productivity of the variable factor. It is time now to study the productivity of the fixed factor, and in general to determine how all the various productivities of any factor vary if at all.

To do this let us use as a framework the homogeneous production function of degree one used previously:

$$Z = \frac{x^2y^2}{(x + y)^3} .$$

This function can be used legitimally because it is compatible with either the strong or the weak formulation of the law of returns. Graphically it will look like figure p. 202 including the fact that all the curves tend asymtotically

toward the x axis. We can now apportion various areas under the TP curve to various stages:

Stage I:

Stage I is made of the ranges:

$0\text{-}x_1$ where $MP_x\uparrow > AP_x\uparrow > 0$; and $TP_x\uparrow > 0$

$x_1\text{-}x_2$ $MP_x\downarrow > AP_x\uparrow > 0$; $TP_x\uparrow > 0$

Stage II:

$x_2\text{-}x_3$ $MP_x\downarrow < AP_x\downarrow > 0$ $TP_x\uparrow > 0$

Stage III:

$x_3\text{-}x_4$ $0 > MP_x\downarrow < AP_x\downarrow > 0$ $TP_x\downarrow > 0$

$x_4\text{-}x_\infty$ $0 > MP_x\uparrow < AP_x\downarrow > 0$ $TP_x\downarrow > 0$

Note that the various stages are established in regard to factor X only. However, it is now easy to determine the status of the productivity of the other fixed factory y in relation to x as x varies. Since the above function is homogeneous of degree one the properties of the Euler Theorem hold as follows: If $P = f(x,y)$ then

$$P = \frac{\partial Z}{\partial X} \cdot X + \frac{\partial Z}{\partial Y} \cdot Y$$

When x_2 is used, we know that $AP_x = MP_x$. With the Euler formula we can obtain the value of MP_y at that point. If $AP_x = MP_x$; then $\frac{P}{X} = MP_x$ or $P = MP_x \cdot X$, in such a case $MP_y \cdot Y$ must then equal 0. Y being greater than 0, than MP_y must be equal to 0 at that level of employment of x. More than that, we can prove with the same formula that in the range

$0\text{-}x_2$, that $MP_y < 0$. The demonstration is based on the fact that in that range $MP_x > AP_x$. Before x_2, then, $\frac{P}{X} < MP_x$; or $P < MP_x.X$. This obliges us to write $P + K = MP_x.X$, where K is a positive quantity that changes the inequality into an equality. We obtain: $P = MP_x.X - K$; we know also that $P = MP_x.X + MP_y.Y$; than $MP_y.Y = -K$. Since $-K < 0$ than $MP_y.Y < 0$; since Y is greater than 0, than in the range $0\text{-}x_2$, $MP_y < 0$. By the same method we can prove that when $AP_x > MP_x$, then $MP_y > 0$. Another critical point is x_3. At that point $MP_x = 0$, consequently the Euler formula at that point can be written as $P = MP_y.Y$, or $\frac{P}{Y} = MP_y$ or at that point $x_3, AP_y = MP_y$.

These critical points x_2 and x_3 determine three economic regions which have the following properties:

Stage I:

$0\text{-}x_2$: in that stage $AP_x < MP_x$, and $MP_y < 0$

at $x_2, AP_x = MP_x$; and $MP_y = 0$.

Stage II:

$x_2\text{-}x_3$: in that stage $AP_x > MP_x$, and $MP_y > 0$

at x_3, $MP_x = 0$ and $AP_y = MP_y$.

Stage III:

$x_3\text{-}x_\infty$: in that stage $AP_x > MP_x$, $MP_y > 0$ and $MP_x < 0$

Many people do not see the point of talking of marginal productivity of factor Y while by hypothesis the amount of this factor remains constant at k, i.e. If we do not vary factor Y how can we know how its marginal productivity changes?

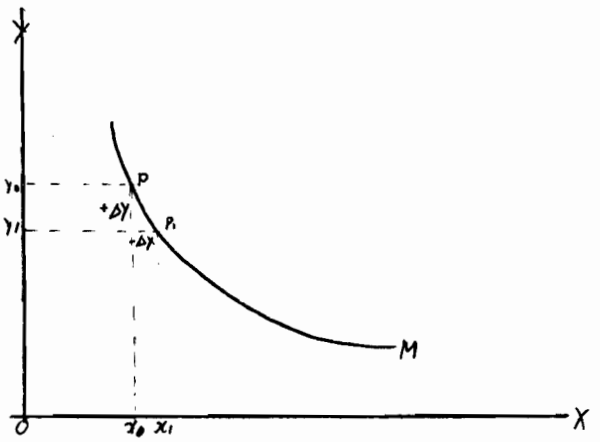
To obtain a clear picture of what is involved let us revisit the original function. This production function when both X and Y are variable determines a surface in space, called a production surface which will look like a mountain. In such a geometrical figure, keeping one factor constant is equivalent to take a cross-section (parallel to an axis). We might say that the figure can be completely reconstructed by an infinite number of cross-sections parallel to the X axis, there will be one cross-section for each possible value of Y. Any of these cross-sections is called a cross-section in X. Figure p.202 pictures one of these cross-sections in X; this one when y is equal to k. The same mountain can also be built with an infinite number of cross-sections parallel to the Y axis, and there will be as many cross-sections in Y as there are possible values for X. Such a cross-section is called a cross-section in Y.

The cross-sections in X are normal planes to the cross-sections in Y, they intersect at an angle of 90 degrees. Here two sets of cross-sections are involved. Consequently the measures of productivity of factor X <u>i.e.</u> MP_X and AP_X, are measured exclusively on cross-sections in X, and conversely the measures of productivity of Y, <u>i.e.</u> MP_Y and AP_Y can <u>only</u> be measured on a cross-section in Y.

Now each point on the production surface is common to two sections (one in X and one in Y) passing by this point. This has the effect of permitting to read simultaneously four

values, AP_X and MP_X as measured on a cross-section in X, and two values, AP_Y and MP_Y as measured on the other cross-sections in question passing by that point. The values of MP_X and MP_Y are easily obtained by taking the partial derivatives with respect to X and with respect to Y. The values of AP_X and AP_Y are simply P/x_o and P/y_o, where x_o and y_o are the coordinates of the point P_o on the production surface.

We can now reduce the three dimensions diagram to a two dimensions diagram by constructing a map of isoquants. In mathematical terms this means projecting on the XY plane the level curves of the production surface. Each isoquant (or level curve) is the locus of a definite level of product. Product is of course constant along an isoquant. If the factors of production are imperfect substitutes, the isoquant will be convex to the origin. The exact curvature is governed by the production function. Along the isoquant, we can measure the marginal rate of substitution of one factor for another. This is equivalent to the slope of the isoquant. Here is why: the marginal rate of substitution of X for Y is the ratio of the change in amount needed in Y to compensate for the loss of one unit of X, in such a way that output is not changed.



$M.R.S._{x\,y} = \frac{+\Delta Y}{-\Delta X}$ = slope of isoquant at P if ΔX tends

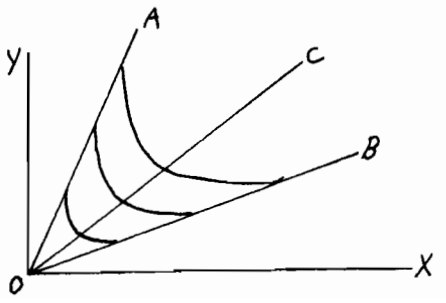
toward zero. Roughly speaking moving from P' to P involves $-\Delta X.MP_X + \Delta Y.MP_y = 0$; when X is reduced, the product is reduced; and the amount of reduction is ΔX times its productivity, and similarly in the case of Y, when the factor Y is increased, the product increases by ΔY.times the productivity of Y. Obviously this is strictly true only if ΔX and ΔY are infinitely small, otherwise this would involve four cross-sections: two in X, one running on the path of Y_oP and the other on the path of Y_1P_1; and two in Y, one running on the path of X_oP and the other on the path of X_1P_1. and this would mean two different MP_x (one on each cross-section) and two different MP_y. However as ΔX tends toward zero the two cross-sections in X are for practical purposes reduced to one, (similarly for Y), such that our computations do not suffer too much. We can then write:

$$\Delta X.\ MP_X = \Delta Y.MP_y \text{ or } \frac{MP_x}{MP_y} = \frac{\Delta Y}{\Delta X}.$$

Consequently at any point P on the isoquant, the slope =

$$mrs_{xy} = \frac{\Delta Y}{\Delta X} = \frac{MP_x}{MP_y}$$

If we draw a sufficient number of these level curves (isoquants) and join them by a line cutting them at a point where their tangents have the same slope, we will obtain



what is called an isocline, OC in figure on page 216. These isolines will be straight lines at least in the case where the production function is homogeneous of degree one (10, p. 320). Two isoclines are of particular interest: the first one OA meets the isoquants when the slopes of their tangents are ∞, and the second, OB, meets them when their slopes are zero. These two isoclines will be in our case straight lines and will be hereafter called ridge lines, because supposedly a firm will never utilize factors outside these lines (we shall see later that this is not always true). Now if on OA the slope of the isoquants is infinite, that means that $MP_X/MP_y = \infty$, this is true only if $MP_y = 0$. On OA then, on the various cross-sections in Y passing by, MPy = 0 on any of these cross-sections. On OA then $MP_y = 0$. Similarly on OB, the slope of the isoquants is 0, then $MP_X/MP_y = 0$, and this is true only if $MP_X = 0$. Then on the cross-sections in X when they meet OB, at that point on each of these cross-sections, $MP_X = 0$. Consequently OB is the locus of $MP_X = 0$.

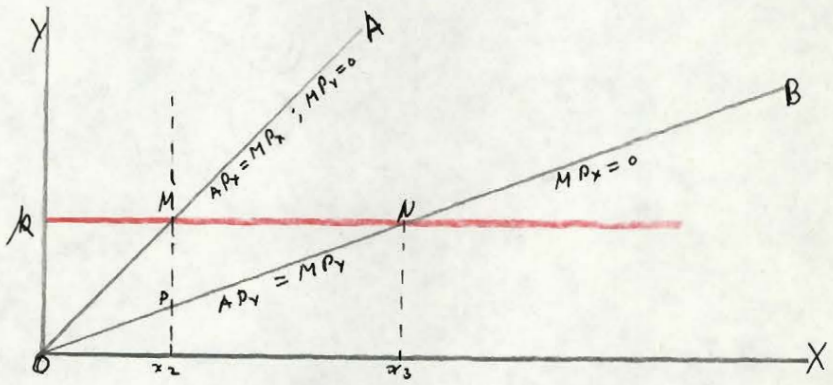
Let us establish now that on OA, $MP_X = AP_X$, that is on (any) cross-section in X meeting OA at M (see diagram below), at that point if we read on this particular cross-section the value of MP_X equals the value of AP_X. This is so because at the same point M reading on the other cross-section Y that passes by M (on OA) we have already established

that precisely at that point on the cross-section Y, $MP_y = 0$. Utilizing the Euler formula, if $MP_y = 0$ at M, then at M,

$$P = \frac{\partial Z}{\partial X}.X \quad \text{or} \quad \frac{Z}{X} = \frac{\partial Z}{\partial X} \quad \text{or} \quad AP_X = MP_X.$$

Consequently on any cross-section in X, $AP_X = MP_X$ when this cross-section meets OA, etc. On OA then, $AP_X = MP_X$ on the various cross-sections in X and $MP_y = 0$, on the various cross-sections in Y.

With the same kind of reasoning we will obtain that on OB, $AP_y = MP_y$ on the cross-sections in Y and $MP_x = 0$ on the cross-sections in X.



Let us now take a particular cross-section in X (where $y = k$). It will have a shape similar to the one of figure p. 202 and the following relations between the two figures can be established:

(1) The space k to M of the production surface (figure p.218) corresponds to the distance O to B on the TP curve of figure p. 202. Consequently in the figure above from k to M,

$MP_X \nearrow 0$, $AP_X \uparrow > 0$, and this corresponds to stage one of factor X.

(2) Point B of figure p.202 corresponds to point M of figure p. 218 because both points asks for $MP_X = AP_X$, and $MP_y = 0$.

(3) Space B to C in figure p.202 corresponds to space M to N in figure p.218 (as far as X is concerned). Consequently in that range corresponds to stage two of factor X, where $MP_X \downarrow > 0$, and $AP_X \downarrow > 0$.

(4) Point C of figure p.202 corresponds to point N of figure p. 218 because both points asks for $MP_X = 0$, and $AP_y = MP_y$, when each measure is read on its respective cross-section.

(5) From point C onward in figure p. 202 corresponds to point N onward (moving to the right) in figure p. 218 (always when we move on a particular cross-section in x, that is along kN in figure p. 218. In this range $MP_X < 0$, (as seen in figure p. 202) and this corresponds to stage three of factor X.

Since M and N can be anywhere on OA and OB (figure p. 218), region YOA corresponds to stage one of factor X, and region AOB to stage two of factor X, and finally region BOX corresponds to stage three of factor X. We can now make a similar reasoning for cross-sections in Y. If they are studied in figure p. 218, we could study a typical cross-section in Y

as one which would pass through points P and M. We could similarly use figure p. 202 to represent this cross-section, if we label its horizontal axis Y instead of X. The TP curve of figure p. 202 now describes the behaviour of AP_y and MP_y. The production function used being homogeneous of degree one, the relations 1 to 5 established previously for X will hold for Y but in the direction of the cross-sections in Y.

(1) From x to P (figure p. 218) corresponds also to the distance O to B on the TP curve of figure p. 202. Consequently (according to the information collected in figure p. 202) the distance x_2P has $MP_y > 0$ and $AP_y \uparrow > 0$, and this corresponds to stage one of factor Y.

(2) Point B of figure p. 202 now corresponds to point P of figure p. 218 consequently at P,

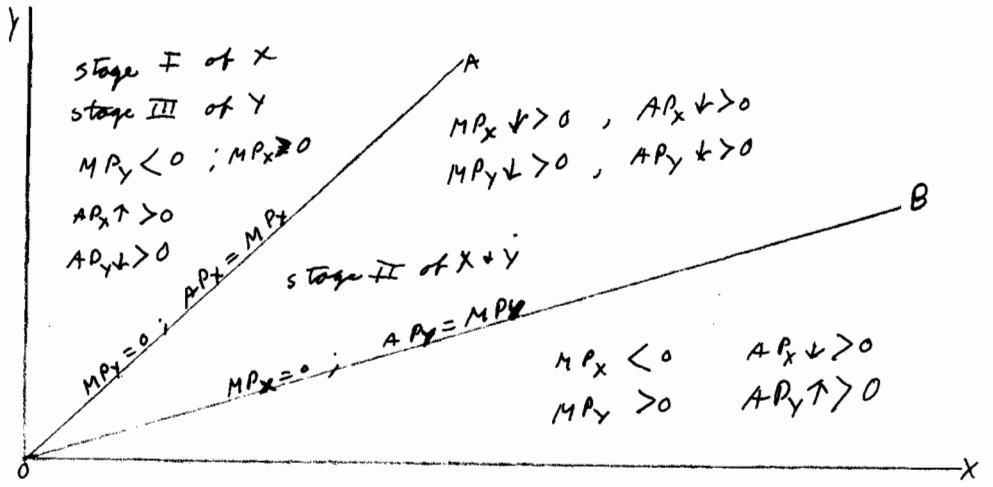
$$AP_y = MP_y, \text{ and } MP_x = 0$$

(3) Space B to C of figure p. 202 now corresponds to space P to M of figure p. 218 (as far as Y is concerned). Consequently in that range corresponds to stage two of factor Y, where $MP_y \downarrow > 0$ and $AP_y \downarrow > 0$.

(4) Point C of figure p. 202 now corresponds to point M of figure p. 218; because both asks for $MP_y = 0$, and $AP_x = MP_x$, when each measure of productivity is read on its respective cross-section.

(5) From point C onward in figure p. 202 corresponds to point M upward in figure p. 218 (when moving on the cross-section in Y,x_2M. In this range $MP_y < 0$ (as seen in figure p. 202) and this corresponds to stage three of factor Y.

Consequently, region YOA corresponds to stage III of Y,
region AOB corresponds to stage II of Y,
region BOX corresponds to stage I of Y.



This diagram above introduces us to the proposition often met in the literature that if the average productivity of the variable factor increases, the marginal productivity of the other (fixed) factor must be negative.[17] In other words if

---

[17] This is implied in Knight (117, p. 102) when he states: "Unless the return per unit of any agency does decrease it is not productive at all; its use adds nothing to the output of the combination. If we imagine increasing returns the agency is negatively productive." This law is used to determine a priori the area of factor utilization.

$AP_X \uparrow > 0$, then $MP_y < 0$. Now since AP_X changes in value, the corresponding MP_y changes also.

Actually we can draw (p. 224) an illustration of what happens to MP_y when the utilization of factor X is changed. Suppose we take as a reference one particular cross-section in X which may look like figure p. 202. This cross-section in X has as many meeting points with different cross-sections in Y as there are different values of X. Furthermore, as shown previously, at each meeting point we can read on a particular cross-section in Y a particular set of values for AP_y and MP_y. Now the MP_y for instance will surely vary as we move from one cross-section in Y to another cross-section in Y as X increases. The exact behaviour of MP_y during this experiment can be determined as follows:

By the Euler formula we have established that as long as AP_X increases (that is $AP_X < MP_X$), MP_y is negative. In figure p. 218 it means that in the range kM, $MP_y < 0$. We also have determined that at the right of M, MP_y is > 0. We have then two landmarks to help us determine how the value of MP_y varies. If MP_y passes from negative to positive, it must have at least one rising phase. Yet it might not be continually rising, because even if MP_y is negative for some time this does not rule out the possibility of a declining phase followed by a rising phase. Suppose <u>k</u> a big value, consequently moving on the X cross-section kM means meeting initially a cross-section in Y in a region where its MP_y is

much at the right of x_4 (see figure p. 202), and where MP_y is near zero since the MP_y line is asymptotic to the absissca in this region. However the next cross-section met by kM will be at an earlier phase of its MP_y curve, and consequently at a value of MP_y is greater (in absolute terms) than the one before. (We are getting nearer to x_4 by moving from right to left in diagram figure p. 202). This will go on until when we are at M (figure p. 218) we are meeting a cross-section in Y where its MP_y value is at x_3 (figure p. 202). For short, moving in figure p. 218 from k to M means moving from right to left in figure p. 202 and thus exploring the values of MP_y at the right of x_3. For instance there might be a point on KM which corresponds to meeting a cross-section in Y when its MP_y is at x_4, (lowest point). Presumably the plotting of the various values of MP_y encountered by kM would be something like this: Point R corresponds to point x_4 in figure p. 202 (when it represents a cross-section in Y). and M corresponds to point x_3 of figure p. 202. However, the shape of MP_y will depend on the value of k. If k is small the kM cross-section in X will meet the cross-sections in Y rather early after their MP_y has attained 0, and consequently the first diagram might

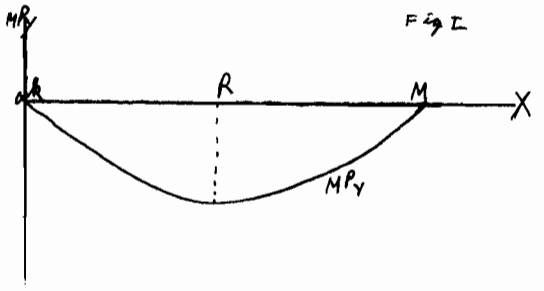


Fig I

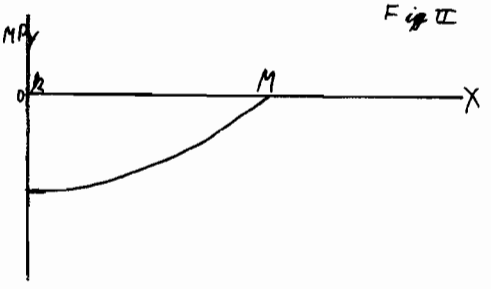
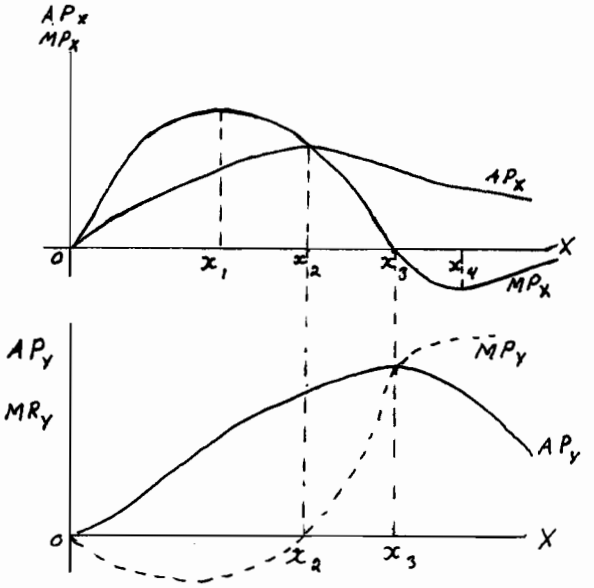


Fig II

better look like the second. We can now give a complete picture of the relations between the various productivities as follows. The first diagram depicts how productivity in X varies along one cross-section in X. In the second diagram we correlate the values in Y as read on the various cross-sections in Y at the point when they meet the cross-section X depicted above.



A few words of explanation might be necessary about the values of Y:

(1) The AP_y curve:

This curve is always positive because TP is always positive. It rises until x_3 because on kM (figure p. 218) TP reaches a maximum at x_3. All this because AP_y is TP_x (along kMN) divided by a constant amount Y_k. Therefore after x_3 AP_y diminishes and finally becomes asymptotic to X.

(2) MP_y curve:

As seen before the initial phase from 0 to x_2 is uncertain in shape (although it is certainly negative). It depends upon the value of k. At x_2 we are on ridge line 0A, MP_y is then zero. It rises hereafter because we know that at x_3 it is equal to AP_y (we are on ridge line 0B). Its exact

shape is also uncertain after that, for example, it will rise for a certain time, and depends upon the value of k. However it will end as an asymptote to the abscissa. Finally the second diagram shows an apparent anomaly. It seems to violate the rule connecting the average and marginal values. The present relation is only an apparent contradiction because the values of Y depicted pertains to different cross-sections while the rule applies to only one cross-section. Before leaving this part a few remarks should be made: Knight's statement is far from being always true. It is surely true as shown above in the case of a homogeneous function of degree one, but it is <u>not</u> true in the case of a function having the shape of a saddle. In this case using half of the saddle, and utilizing the cross-section in X which passed by the saddle point, MP_X and AP_X will be increasing continually as X increases, while simultaneously MP_y is continually zero. We should also note (in the case of a function which is homogeneous of degree one) that the productivity of the fixed factor changes with the rate of utilization of factor X. On the other hand if both factors are kept in a fixed proportion, for example, moving along an isocline, the productivity (MP_X and MP_y) does not change. This shows that the size of the firm (in this case) does not change the productivity of the factors.

What precedes it quite well known, and many textbooks talk about the marginal and average productivities of the

two factors.[18] Yet, they do not make it clear (with the exception of Stigler) that the precise relations between the productivities described previously hold only in the special case of production function which is homogeneous of degree one and conforms to the following formula:[19]

$$Z = \frac{x^{\frac{\alpha}{\alpha-1}} \cdot y^{\frac{\beta}{\beta-1}}}{\left[(\beta-1)x + (\alpha-1)y\right]^{\frac{\alpha(\beta-1)}{(\alpha-1)(\beta-1)}}}$$

Otherwise it is very unlikely that they will hold. For instance if the function is homogeneous but of degree n, although the production surface will have generally the same shape, the isocline on which $MP_y = 0$ will not coincide with the isocline where $AP_x = MP_x$ (204, p. 302). A bit of computation with the Euler formula in this case will show that when $MP_y = 0$, then $AP_x = MP_x/n$; and consequently $AP_x < MP_x$, (all this if $n > 1$). A fortiori this lack of synchronization might be greater in the case of non-homogeneous functions. Moreover, the exact drawing of a homogeneous production function should be recognized. Many people use a reasoning involving a homogeneous production function of degree one (or what they call constant returns to scale), yet when they draw it they use curved ridge lines (187, p. 119; 122, p. 134).

---

[18]See Weintraub (222), Scitovsky (187), Dehem (63), Leftwich (122), Stigler (204).

[19]Furnished to us by Dr. Carman Miller, Department of Mathematics, University of Saskatchewan. Its detailed derivation is found in appendix E.

This is mathematically unsound since the ridge line is a locus where there the slopes of the tangents to the isoquants are all equal; and in our case it is well established that all the isoclines are straight lines (10, p. 320).

The catalogue of fantastic behaviours attributed to homogeneous functions of degree one (constant returns to scale) is very wide. For instance, one doubtful practice[20] is to postulate such a function and draw the corresponding figure in such a way that total product becomes nil on two radii from the origin.[21]

Another fallacy, while talking about a homogeneous function of first degree consists in saying that "there will always be diminishing marginal physical productivity along a horizontal or vertical line like MM in figure 87a" (208, pp. 225-227). And these authors spend a few pages proving that point. Obviously if one refers to our function drawn p. 202, their statement is not always true. Some other people think that the homogeneous functions of degree one are all of the Douglas' form $P = \sqrt{A.B}$. That is no maximum is reached on any cross-section. It should now be clear that this form is only one form among many others that are

---

[20] See Cassel, (53, p. 110). The homogeneous function of degree one is postulated pp. 105-106. See also Dehem (63, p. 61).

[21] My mathematician was not able to find any homogeneous function of degree one that would satisfy Dehem and Cassel. Yet it is only a presumption that they are wrong.

possible.

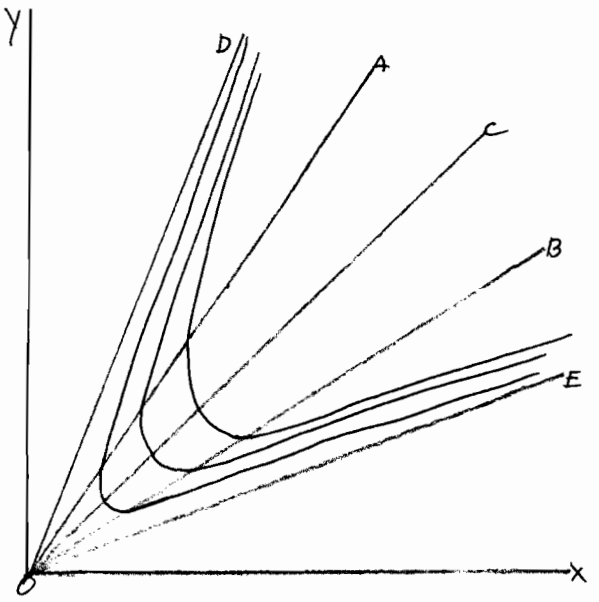
It is time now to move to another type of function (homogeneous) which unfortunately will not satisfy the precise relationships established above, but will satisfy another constraint: total product can become nil in the first quadrant, (this while satisfying the first constraint in the form of diminishing MP). It has been contemplated for some time[22] that on the grounds of economic relevance that total product might be nil initially on any cross-section, and might reach zero later for finite quantities of any of the factors. In the words of Cassel: "The curves (TP) do not begin at the origin but at some point on the X axis to the right of it, because before the ratio of the excessive factor to the deficient one becomes infinite (as it is at the origin) the product is reduced to zero" (35, pp. 108-109). We can even go further and introduce negative products as Knight does in the drawing of the TP curve (117, p. 100). By negative product we mean waste which is costly to remove. For instance, one man trying to run a blast furnace alone, if the fixed factor is indivisible and unadaptable, may produce only waste. All this to say that there are economic reasons why we should build a model (a production function) which would have as one of its properties

---

[22]Knight (117, p. 100), also Cassel,(35, p. 109).

that the TP becomes nil in the first quadrant. The following function:[23]

$Z = Z(x,y) = \sigma(x - By)(y - \alpha x)$ where σ is a positive constant and Z a homogeneous function of degree n = 2, does that. This function pictured in a three dimensional diagram will look like the section of a cone. The three dimensional diagram (as previously) can be reduced to a diagram in two dimensions by projecting on the YOX plane the level curves of the production surface. This homogeneous function of degree two has the following properties:



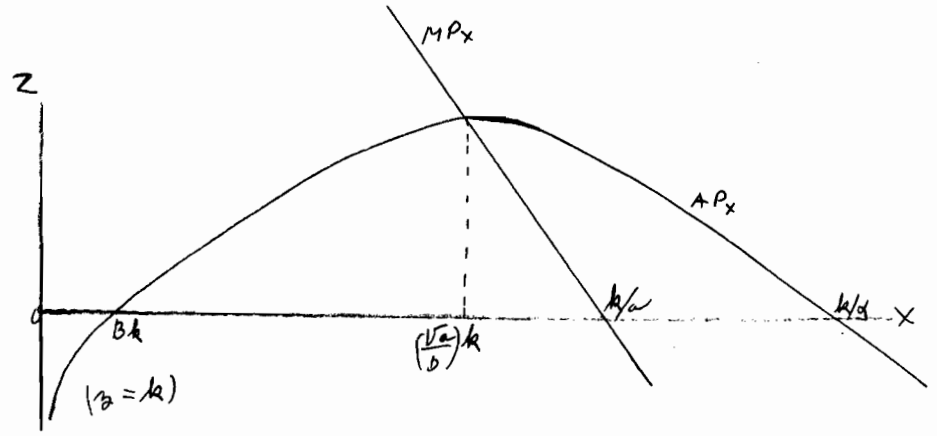
(1) No level curve (isoquant) accounting for a positive product will be found in the region YOD or EOX. All isoquants are asymptotic to these curves, OD and OE, which are radii from the origin.

(2) The isoclines are straight lines from the origin. For instance, radius OA is the locus where the slopes of the tangents to the various isoquants have a slope of ∞.

---

[23]This function has been also furnished to me by Dr. Miller and appears in appendix E.

Similarly OB is the locus where the slopes of the tangents to the isoquants have a slope of 0. A cross-section in X (when Y = k) looks like the figure below. The points of correspondence between figure below and figure p. 229 are such:



Point B_k of 230 corresponds to a point on OD in fig. 229 or 232

.. $(\sqrt{a}/b)k$ " " " OC " " "

.. k/a " " " OB " " "

.. k/d " " " OE " " "

In both diagrams figure p. 229 and 232, the location of line OC must be determined, because in such a function it can be shown that the locus of points of maximum AP_X (OC), does not coincide with the ridge line OA, ($MP_Y = 0$). This is seen either in the appendix on this function or by using the short reasoning in terms of the Euler Theorem: On OA the Euler formula gives $nZ = MP_X.X$, consequently on OA, $AP_X < MP_X$. If we take as a reference a cross-section like kN (in figure p. 232) point M is on OA, but now this is not a point of maximum AP_X. Since at M, $AP_X < MP_X$, AP_X is still

increasing. Yet it must reach a maximum before the second ridge line OB, because on OB, $MP_X = 0$, consequently AP_X has reached its maximum before the point N (in figure p. 232). Suppose point R is this point where AP_X is maximum, then OC must pass by that point, for example, OC is between OA and OB. A similar reasoning in Y would show that the locus of $AP_y = MP_y$ (maximum AP_y) is also between OA and OB; moreover in the function used above, the locus of maximum AP_y is also the locus of maximum AP_X.[24]

On any cross section in X (or cross-section in Y):

$MP_x \downarrow > 0$ until it meets OB, and is negative herafter

$MP_y \downarrow > 0$ until it meets OA, and is negative herafter

and:

$AP_x \uparrow < 0$ in the range between OY & OD

$AP_x \uparrow > 0$ in the range between OD & OC

$AP_x \downarrow > 0$ in the range between OC & OE

$AP_x \downarrow < 0$ in the range between OE & OX

and

$AP_y \uparrow < 0$ in the range between OX & OE

$AP_y \uparrow > 0$ in the range between OE & OC

$AP_y \downarrow > 0$ in the range between OC & OD

$AP_y \downarrow < 0$ in the range between OD & OY

$TP < 0$ in the range between OY & OD and OE & OX

---

[24]See appendix E.

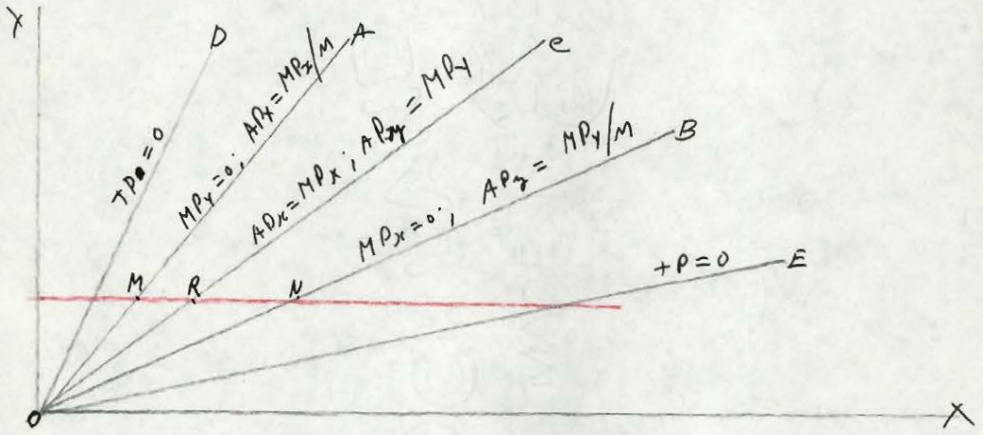
TP $>$ 0 in the range between OD & OE

We can summarize this information by region as follows:

Region	MP_x	MP_y	AP_x	AP_y	TP
YOD	↓ +	↓ −	↑ −	↓ −	−
DOA	↓ +	↓ −	↑ +	↓ +	+
AOC	↓ +	↓ +	↑ +	↓ +	+
COB	↓ +	↓ +	↓ +	↑ +	+
BOE	↓ −	↓ +	↓ +	↑ +	+
EOX	↓ −	↓ +	↓ −	↑ −	−

(where ↑↓ represents increasing or decreasing; and + − represents greater or smaller than zero)

A small amount of these informations can be shown on the following diagram:



As before we can distinguish stages as follows:

Stage I of X corresponds to space YOC
I of Y corresponds to space XOC

Stage II of X corresponds to space COB
II of Y corresponds to space COA

Stage III of X corresponds to space XOB
III of Y corresponds to space YOA

It is obvious that the stages of the two factors do not coincide; and this will cause embarrassing problems in terms of areas of factors utilization (seen later), while OA and OB (by definition) are the ridge lines.

We are now ready to discuss the theory of the area of factor utilization (in the long run first). In so doing, we will speak in general terms, that is, this theory will be studied in terms of many types of functions. We will look at four cases: both in Long Run and in Short Run.

(A) LONG RUN

Case I: Factor X and Factor Y cost nothing:

First, in such a case, regions of negative products can be ruled out a priori, and for any types of functions.

Secondly, production should be inside ridge lines (if the revenue factor is ruled out). Let us see how: From the point of view of factor X, (moving on a cross-section in X), production will not be extended further than point N, (ridge line OB) because TP is maximum at that point; any further use of X (even if costs nothing) will simply reduce total output. Similarly moving on a cross-section in Y, production will not be extended further than point M (ridge line OA) because as far as Y is concerned TP is maximum on OA. A priori, then production should be made inside the ridge lines.

Thirdly, we can delimit still further the area of factor utilization by utilizing the Knight's rule which reads as follows:[25] "unless the return per unit (AP) of any agency

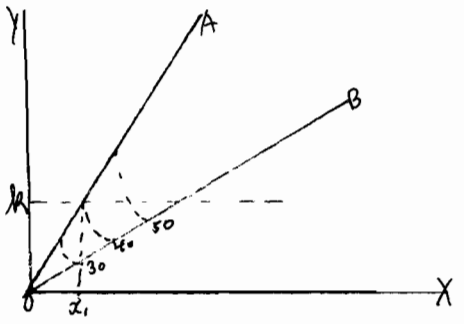
---

[25]Knight (117, p. 102). Let us notice there that Knight's rule has been devised along a single cross-section; it will be interesting to see what it gives in a larger framework.

does decrease it is not productive at all, it adds nothing to the output"; this rule implies that we should use only those combinations of factors which lie in the region (in the diagram) where both factors experience diminishing average product.

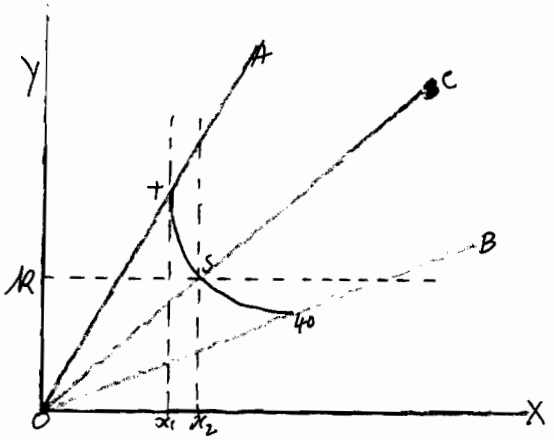
This rule is easily complied with in the case of our first function (hereafter called function No. 1); however it cannot be applied to function No. 2 because, according to Knight, all regions are now forbidden. The Knight's rule can be temporarily rescued by formulating it in a weaker form: a factor should not be used in the range where its average productivity is increasing. With this, line OC (figure p. 232) could satisfy a weaker formulation of the Knight's rule.

Case II: Factor X costs something and factor Y costs nothing: A fortiori here, regions of negative total product or marginal products must be ruled out. The Knight's rule as used today (in such a case) stipulates that the "firm will achieve greatest economic efficiency...at the ratio of labor to land which maximizes product per unit of labor" (122, p. 155), i.e. at maximum AP_X. In the case of function No. 1 this occurs on OA (the ridge line). Here the cheapest way to produce any quantity of product, for example, 40 units, is to use



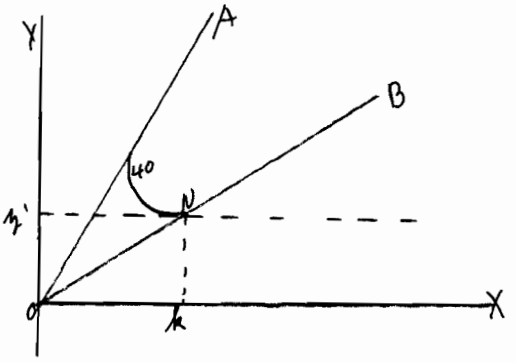
k of Y and x_1 of X. All the other ways to produce need greater quantity of X and thus are more costly.

However Knight's rule meets difficulty in function No. 2. Suppose we want to produce any amount, as previously, for example 40; Knight's rule will require a combination of X and Y on OC, (where AP_X is maximum). Yet this is not the cheapest way to produce 40 units.



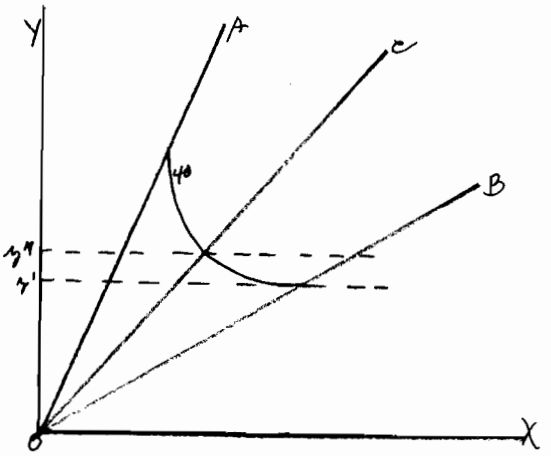
To prove our point we must introduce the notion of least-cost combination, a rule known today through the works of Frisch and Sune Carlson; hereafter called the Frisch-Carlson rule (34, pp. 31-41). The Frisch-Carlson rule determines factor utilization through the use of iso-costs. Since factor Y costs nothing, the slope of isocosts is infinite, thus the lowest isocost that permits the production of 40 units of product is tangent at T, requiring combination x_1 of X which is cheaper than combination S (on OC) as required by the Knight's rule.

Case III; Factor X costs nothing, and factor Y costs something: This is the converse of case II. Knight's rule demands that production occur when AP_Y is maximum. In the case of function

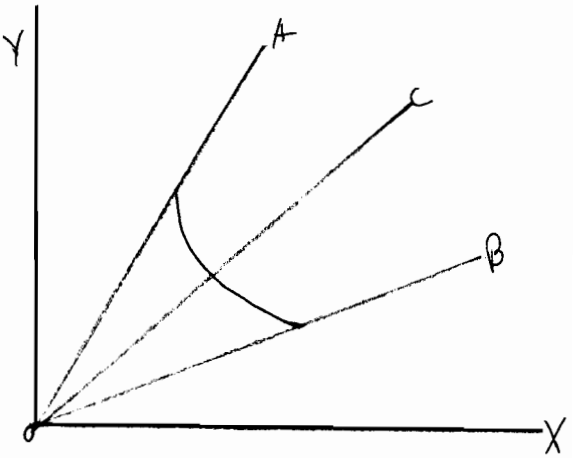


No. 1 the Knight's rule gives the same solution as the Frisch-Carlson rule. The isocosts here have a slope of zero, and are tangents to the isoquants on the OB ridge line. However in the case of function No. 2, as before Knight's rule leads to a false solution. In the first diagram of page 236, Knight asks for a solution on the OC line, while Frisch-Carlson isocosts demand a solution on OB.

Case IV: Both Factors Y and X cost something:



Here in the case of function No. 1 Knight's rule asks for both APs to be decreasing and this is realized in area AOB. The Frisch-Carlson in no way contradicts this, but simply gives a more precise solution. This must always be so because the Frisch-Carlson solution utilizes the tangency of the isocost and the isoquants. Since the slope of the isocosts is always negative, they can be tangent to the isoquants only in the range inside the ridge lines.



However, in the case of function No. 2, only by accident would the Knight solution agree with the Frisch-Carlson solution. Knight's in its weakest formulation asks for

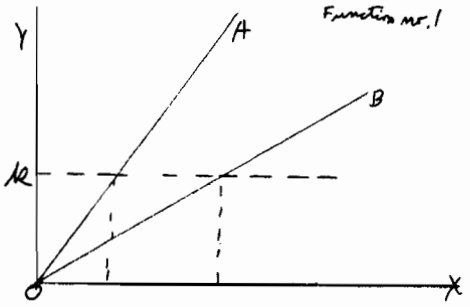
expansion path OC, while the expansion path of Frisch-Carlson can be anywhere inside AOB.

THE AREA OF FACTOR UTILIZATION--SHORT-RUN

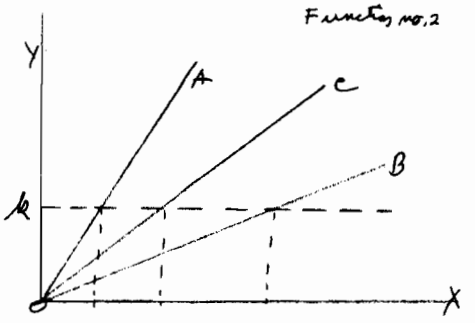
In the short-run one of the factors of production is fixed. Not being free to choose the optimum combinations of factors, this has the effect of rendering the Frisch-Carlson rule inoperative. The problem consists now in maximizing the returns, or minimizing costs of production.[26]

Factor x is the variable factor, and factor y the fixed factor (this will be true of all the cases in this section).

Case I: Factor X costs nothing; factor Y costs nothing:



In the case of function No. 1 production must be extended until TP has reached a maximum. (On OB). Knight's rule (in its weak formulation) is not violated; AP_X is diminishing, and AP_y is not increasing anymore.



The same solution applies to function No. 2; but since AP_y is still increasing in that range, Knight's rule is violated.

---

[26]In this context we do not take the demand factor into consideration.

Case II: Factor X costs something, and factor Y costs nothing:

The minimum cost of production is clearly achieved when AP_x is maximum; on OA for function No. 1 but on OC in the case of function No. 2. Knight's rule is violated.

Case III: Factor X costs nothing and factor Y costs something:

The minimum cost of production is achieved when AP_y is maximum; that is on OB for function No. 1 and on OC in the second function. (This also partly does not comply with Knight's rule).

Case IV: Factor X costs something and factor Y costs something:

For the first function the cheapest average costs (AVC + AFC) is at least not before OA because up to this point both AVC and AFC diminish. After that point AVC increases while AFC continues to decrease; consequently minimum cost is further than OA. However, this point is before OB, because by hypothesis, there is no way to produce more product after OB. Knight's rule is thus not violated.

Similarly for function No. 2 the least cost must be between OC and OB, because up to OC both AVC and AFC diminishes, and there is no way to increase production after OB.

SECTION V: LAW OF RETURNS--SECOND BRANCH

LAWS OF RETURNS TO SCALE

This is also a subject which had to be faced by the economists of the period when they discussed the most economical way to produce in the Long Run, for instance, when all the factors can be varied. The theoretical framework to study returns to scale is basically the same as the one used to study the laws of returns to the use of a particular factor; both utilize the same production function. However, the first branch studies the behaviour along cross-sections, while the second branch studies the behaviour of total production along radius-sections (or radius from the origin).

The purpose of studying the returns to scale is to discover the relations between changes in utilization of all the factors and changes in total output. This is basic to building later the long-run cost curve of the firm, because among those relations, one set of these will be chosen to establish the rate of utilization of the various factors.

This has an important bearing on many questions. Among others, the size of the firm. Although this is a loose term often used as synonymous of scale of production of the firm, when we are reasoning on the long-run cost curve. By custom we say that a firm changes its size (or scale of operations) when it changes the utilization of all factors in the same direction (plus or minus) although not necessarily in the

same proportion.[27] Actually three circumstances must be distinguished where the proportion between factors used is changed although only one is recognized as change in size of the firm.

(1) Along an isoquant the utilization of both factors is changed (one is increased, while the other is decreased) while simultaneously their proportion of utilization varies. This is not considered as a change in the size of the firm.

(2) Along any parallel to the axes (MN or MP figure p.218) the relative proportion between factors is changing as total output changes, but this is not considered a change in the size of the firm when we move in either directions, because there is at least, in each case one factor constant. Translated into costs terms we are moving along a short-run cost curve, where by hypothesis the size of the firm does not change.

(3) The third circumstance is moving in the northeast direction in the first quadrant. There are two methods to move here: moving on a radius (straight line from the origin)

---

[27]See Boulding (24, p. 677). One can distinguish between changes in size and changes in scale. Strictly speaking we could use the expression change in scale in situations when we are moving in the northeast direction on a diagonal only. Because only in such a circumstance are the proportions kept constant. Once we change the proportion we are no more changing the size according to the scale. For this reason we prefer to use the term change in the size of plant when we are moving in the northeast direction but not along a diagonal, while we will use change in scale when we are moving along a diagonal.

thus keeping the proportions between the factors constant; or moving on an isocline which might not be a straight line; this involving changing the proportions in the utilization of factors.

In both cases it is customary to call this, a change in the size of the firm, since presumably the size of the former fixed factor, for example, plant, is changed.

We have previously encountered a law of eventually diminishing returns to the use of one factor. No such law exists (a priori) when all the factors are variable. There is no definite pattern of returns to change in the size (scale) of the firm (plant). However, all the returns to change in size can be classified into three types:

(1) constant returns to scale: when by changing the use of the factors in a certain proportion we change the output in the same proportion;

(2) increasing returns to scale, when a given proportional increase in factors increase product by a still greater proportion, and

(3) decreasing returns to scale when the increase in product is proportionally smaller than the proportional increase in factor utilization.

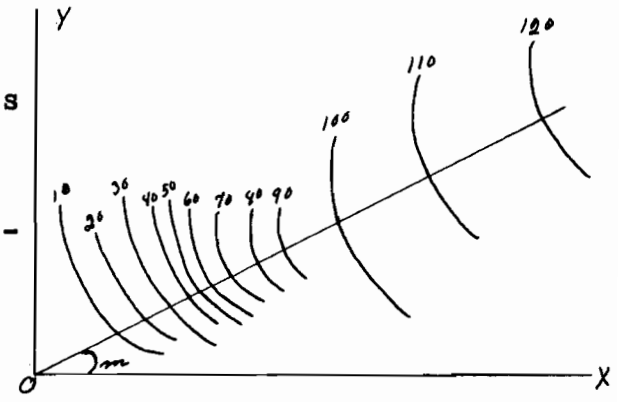
Let us remark that this classification applies only in the case where expansion in the size of the firm is along one given radius (or proportion), the other case (when many radii are used) will be dealt with later. In the case

where expansion is along one radius the concept of elasticity of production can be used as a short-cut to determine those relations between factor utilization and output. This is done as follows: (34, p. 17; 63, p. 59 ff.).

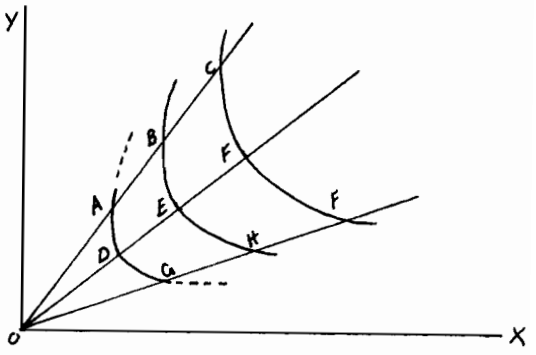
$$\text{Let } e = \frac{dP}{P} : \frac{dx}{x} = \frac{dP}{P} : \frac{dy}{y} \quad \text{where } \frac{dy}{y} = \frac{dx}{x} = m$$

(P = product; x, y, are factors or production and m is the proportional change in the factor utilization).

When all factor utilization is changed by the proportion m, the total change in P is given by the total differential: $dP = MP_x.dx + MP_y.dy$; now from the equation of e above, we can write $dP = ePm$ then: $ePm = MP_x.d_x + MP_y.dy$, which can be rewritten as: $ePm = MP_x.mx + MP_y.my$ and simplifying the m, we obtain: $eP = MP_x.x + MP_y.y$. Now since e can be $\gtrless$ 1, it is possible that a production function can experience either increasing returns to scale if $e > 1$; or constant returns to scale, if $e = 1$, and decreasing returns to scale if $e < 1$. To determine whether a function is increasing, constant or decreasing, at a certain point of the production surface, we solve for e. It is possible for some functions to experience successively all three values at different points on the production surface. The diagram at right, (a non-homogeneous function), experiences all three types of returns, along one proportion radius m. Up to isoquant 60 we experience increasing returns to scale,



constant returns from 70 to 90, and decreasing returns hereafter. The case of homogeneous functions is much simpler. Their general function is $t^n P = f(tx, ty)$. Constant returns to scale (at every point of the production surface occurs when $n = 1$, increasing returns when $n > 1$, and decreasing returns when $n < 1$. Graphically, constant returns to scale requires that: AB = BC = ...
DE = EF = ...
GH = HI = ...



Only one type of function can experience constant returns to scale at all points of its surface: they are linear homogeneous functions. The other functions (the non-homogeneous ones) can experience it only at certain points.

EXPANSION PATH:

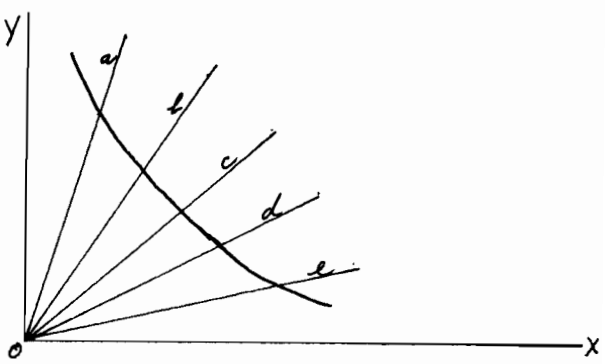
Returns to scale have a close relation to the expansion path. We have seen previously that the Frisch-Carlson formula determines the expansion path by connecting the points of tangency between isocosts and isoquants. The expansion path being an isocline, it will be a straight line in the case of homogeneous functions of whatever degree, (provided the prices of the factors of production do not change) and it will likely be a curved line for all other functions. The non-homogeneous case requires that production is not always increased by changing factors utilization in the same proportion. Staying

on the expansion path might require $\psi P = (\lambda x. \phi y)$; that is changing x by λ, and y by ϕ, and consequently changing P by proportion ψ. Here it is difficult to talk in terms of returns to scale because there is no way to reduce (λ) and (ϕ) to a common index. (Yet this can be done when working in terms of returns to outlay.)

What governs Marginal Productivity?

We already know from the previous section that proportion is important in determining MP_x or MP_y. When we varied only one factor, its MP changed as we changed its employment. Now since the other factor was fixed the proportion of utilization of X with respect to Y was simultaneously changed. This is why we say that proportion is important in determining the productivity of a factor. An alternative illustration is provided with the use of an isoquant:

Suppose various proportions a, b, c, d, e, ... It can easily be found that the productivity of a factor depends upon the relative proportion of its use. If proportion e is used, MP_x is very low while MP_y is very high, its the opposite in proportion a, where MP_x is high and MP_y is low. etc.... This is a general statement which is true for any type of function: homogeneous, or non-homogeneous. Yet when we take the size of the plant



into consideration we might now try to determine if the size of the plant also has an influence on the productivity of a factor. In other words do we have a "scale" effect on the marginal productivity besides the "proportion effect" on marginal productivity? If so, total effect will be equal to "proportion effect plus scale effect".

Let us take the simplest case: the one of a homogeneous productions function of degree one: here the Euler formula gives: $P = MP_x.X + MP_y.Y$; and also $mP = MP_x(mX) + MP_y(mY)$. Consequently along a given radius representing proportion k, changing X and Y by m will not change their marginal productivity. This can be verified by the fact that mathematically speaking the isocline is a straight line. Consequently here the "scale" effect is nil, because marginal productivity of any factor does not change when only the "size" of the firm is changed.

This is not what will happen when the function is no longer homogeneous of degree one. For instance if the function is homogeneous of degree 2, marginal productivity will change with changes in size (keeping the proportion constant). In this case doubling the factors means a fourfold increase in the product. Consequently if originally: $P = MP_x.X + MP_y.Y$; doubling the factors entails $4P = (2MP_x).2X + (2MP_y).2Y$. This clearly shows that doubling the size also doubles the marginal productivities of the factors.

Finally, in the case of non-homogeneous functions, they

will in most cases have isoclines that are not straight lines; that is any radius from the origin will meet successive isoquants in points where the slopes of their tangents are not parallel. Here changing the size (along one radius, not along one isocline) has a "scale" effect on productivity. In general then, the change in productivity of a factor is the sum of size effect and the proportion effect.

SECTION VI: LAW OF RETURNS--FIRST BRANCH

HISTORICAL SECTION 1870-1920

On this topic so rich and so important for the theory of the FIRM, it is surprising to find that most of the authors of the period are far from clear. The only exception is Edgeworth who made a decisive contribution.

JEVONS:

Jevons (113, p. 212), like most of the people of his time, begins by formulating the law of diminishing returns to the use of a composite factor (labour and capital) in proportional[28] or average terms, and later in incremental terms. This happens very often in his chapter on rent. He buttresses his reasoning by offering the classical proof of the likelihood of the existence of the Law in these words:

---

[28] By proportional Stigler (202, p. 19) means average. We shall see below p. 248 how unsuitable this appellation can sometimes be.

"It is quite impossible that we could go on constantly increasing the yield of one farm without limit, otherwise we might feed the whole country upon a single farm" (p. 212). In our theoretical framework this means that the total product curve is bounded, or, in terms of his symbolic statement of p. 217, that dx/dl never becomes negative, but simply tends to zero. Consequently Stigler (203, p. 36) erroneously criticized Jevons for having said that "x (the product) will never decrease when w (the labour) increases". To prove Jevons wrong amounts to showing that the TP curve he had in mind has a decreasing phase. But this is not so; he simply postulated a TP curve which is an asymptote to an upper boundary.

Later, (113, p. 214) he changes the formulation of the law from an incremental form to a calculus form: "so that the doses considered are infinitely small and indefinitely numerous."[29]

Finally much of what Stigler (202, p. 36) calls formulations in proportionate (average) terms may not be so at all. The Jevonian way to express himself is thus: "dx/dl represents the ratio of produce". Therefore each time we meet expressions worded in ratio terms, this might not be

---

[29]The incremental form is also found in his exposition of Rent, which he borrows completely from James Mill. See Jevons (113, p. 213) when he states that further doses of capital (or composite doses of capital and labour) will be less and less productive.

at all an average term language. For instance in an earlier statement he said "that if more or less labour or capital be applied to the same portion of land, the produce will not increase proportionally to the amount of labour." Paradoxically, this statement is not an average term formulation, because proportionally means here that $\Delta X/\Delta L = 1$, and more than proportionally would mean $\Delta X/\Delta L > 1$, and less than proportionally means $\Delta X/\Delta L < 1$. These are incremental formulations, not average. This is not the only case where Stigler is lead astray by the word "proportional", Prof. Jaffe scolds him also in the case of Walras. Yet since Jevons' formulation is sometimes in incremental terms, his 'flower pot' argument is liable to the criticisms of Karl Menger.

WICKSTEED

When one starts to read Wicksteed (228, pp. 527 ff.) one has the impression, which unfortunately does not last long, of meeting a clear-cut exposition of the Laws of Returns. In the first pages, Wicksteed enunciates the fundamental distinction between the first branch and the second branch: in the first branch at least one factor of production is fixed, and in the second branch all the factors can be varied. This approach leads him to confirm the statement of his classical predecessors that since agriculture is the place where "par excellence" one factor (land) is fixed, then the law of diminishing returns applies to agriculture; while he applies the

second branch of the Laws (increasing returns version) to industry. This is in itself an unsatisfactory statement because increasing and diminishing returns can be found in both branches of the Law. In another connection (p. 535 ff.) he assimilates the working of the first branch to short-run situation, while the second branch is assimilated to long-run situation. The working of the first branch[30] in Wicksteed besides being sometimes expressed in incremental form and in other times in proportional (average) form has also strange second branch overtones.[31] Furthermore these second branch overtones are implicitly based on homogeneous production functions of degree one; they are true only in this particular case. However this does not prevent Wicksteed from concluding falsely that "the law of diminishing returns, then, is really no more than an axiomatic statement of a universal principle" (p. 529). In other words he raises to the rank of an axiom a statement which is true only in a particular case. He is then left with no worthy proof of the Law of diminishing returns.

Finally, forgetting about homogeneous production functions of degree

---

[30] As noticed by Stigler (202, p. 49).

[31] Wicksteed (228, p. 527) "we find that at any rate after a certain point, doubling the amount of labour does not double the product." Also later p. 528, "when a number of factors co-operate to produce a result, you double some of them, without doubling others of course you cannot expect to double the results."

one, he undertakes to study "increasing returns" (in a second branch approach). He wrongly assimilates diminishing average costs with increasing returns. Without any proof except a vague reference to factors of production which are indivisible in short-run but divisible in long-run, he proceeds to depict a case of increasing returns (figure 36, 228, p. 535). There is no mention of the rising portion of this average costs curve although he says that "there is a limit, however to the decline of xc...(average cost)" (p. 535). This is as far as Wicksteed went.

WALRAS

The treatment of the Law of Diminishing Returns by Walras is not simple to summarize. In his chapter on Production (lesson 20) we would expect it; yet he uses fixed coefficients and consequently he does not have to face the law. Properly speaking, since this is a scale approach (2nd Branch), he does not have to mention it. Yet this approach is only provisory and has been used only for simplicity, like the use of a homogeneous production function of the first degree.[32] It is only when he studies the "marginal Productivity Theorem", Lesson 36, that he has to use the law implicitly. Yet there is no explicit reference to the law of diminishing returns.

---

[32]Walras (220, p. 385): "I preferred, however not to introduce the theory of marginal productivity into my general theory of economic equilibrium for fear that the general theory, which was already complicated enough, might then, be too difficult to grasp in its entirety."

Nevertheless, he is first successful in isolating the marginal product of each factor: $\frac{\partial \phi}{\partial T}$ (equation 3, p. 385). Then, he shows that in equilibrium, the utilization of each factor T, P, K, must be according to what is known today the least cost rule: $\frac{p_t}{\frac{\partial \phi}{\partial t}} = \frac{p_p}{\frac{\partial \phi}{\partial t}} = \frac{p_k}{\frac{\partial \phi}{\partial t}} = \ldots\ldots$ (No. 4, p. 385)

The difficulty with Walras is that he is mainly interested with equilibrium situations and that he does not show the path to equilibrium. For example, in this case he does not explicitly state that if starting from disequilibrium situation, equilibrium is reached, through the marginal productivity approach only if $\frac{\partial \phi}{\partial t}$ eventually is diminishing. Otherwise if $\frac{\partial \phi}{\partial t}$ is forever increasing we will always hire more and more of T without passing to P, K, ... In his chapter on English Rent, (another place where the writers have to meet the law)pp. 406-409, he expressed the law in terms of diminishing average product; for example, he is talking about rents per hectare as more capital is used. (He has no initial rising phase for the curves).

Yet one cannot say that he was not aware of the law. He is the first to draw a parallel between the theory of consumption and the theory of production. So that as the theory of final utility is the principle upon which demand for goods is established, also the marginal productivity theory "shows the underlying motive for the demand for services"

(p. 385). Paraphrasing we can continue by saying that since marginal utility must eventually diminish to have downward sloping demand curves, then MP must eventually diminish also to have downward sloping demand curves for the services of the factors of production. Yet it must be concluded with Hicks that his exposition is far from satisfactory.[33]

MARSHALL

The analytical progress is very slow. Marshall's exposition reflects some lingering with the old classical fetishes like the statement referring to the applicability of the laws of diminishing returns to agriculture and extractive industries, and increasing returns to manufacturing (see his representative firm). In his first definition of the law of diminishing returns (136, p. 125) he refers to agriculture and it is worded in the first branch approach. In his study of increasing returns (pp. 265-266) the law is worded in such an ambiguous way that it is impossible to determine which branch of the law is used, even if change in scale is implied. Marshall distinguished between the forces at work in the first branch and those at work in the second branch, and consequently was able to show that the branches

---

[33]Hicks (99, p. 234) says: "Walras expressed himself in a so crabbed and obscure manner that it is doubtful if he conveyed his point to anyone who did not possess some further assistance. Anyone who knows the answer can see that Walras got it, but anyone who does not must find it almost impossible to get it from Walras."

of the law depend upon quite different factors.

Marshall (136, p. 141) saw the reason for diminishing returns: he correctly states that diminishing returns are due to an "excessive proportion to the other agents." In other words the relative marginal product of factors of production depends (among other things) upon the proportion of use of the factors. However he does not mention that relative size is also a factor in determining marginal productivity.

Many people (225, p. 401; 202, p. 66) are of the opinion that Marshall was confused by the average and incremental formulation, and Stigler holds that any definition of the law in average terms is inappropriate, (Edgeworth is also scolded for the same "sin").

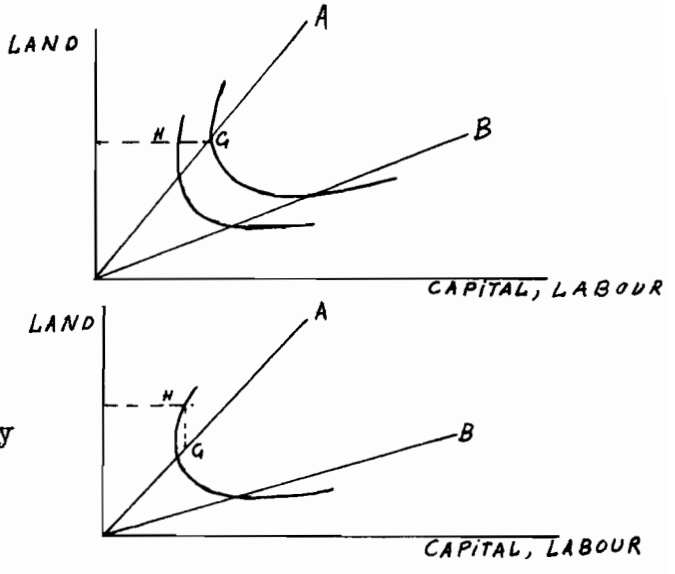
First it must be said that the average formulation is no more inferior than the marginal, simply because neither when formulated alone, are insufficient to determine the area of factor utilization (except if a money factor analysis is used).

Secondly, these various ways to define the law do not lead Marshall to any error as far as factor utilization is concerned.

Thirdly, as in Jevons, some statements even though they use the word "proportionate", besides having a meaning of proportionality, (average formulation), it can also be interpreted in incremental terms. Marshall (136, p. 125)

says: "An increase in the capital and labour applied in the cultivation of land causes in general a less proportionate increase in the amount of product raised...." This might mean that $\Delta X/X > \Delta P/Tp$, or $TP/X > \Delta P/\Delta X$ or $\Delta X/\Delta P > 1$. (where P = product; X = factor).

In these terms a proportionate formulation, would then limit the exposition of the Law of diminishing returns to the decreasing portion of AP curve. This is not a bad idea if one has in mind the area of factor utilization, where stage one (the rising portion of AP) is irrelevant. Yet the marginal formulation is necessary to determine the limit of stage two, and this is not seen in Marshall. Although p. 125 he correctly determines in the case where factor X cost something that production should be carried on OA (of our figure p. 218).



Increasing returns in Marshall means rising AP_X. In diagram at right, his reasoning amounts to recommending that if the firm is at H it should move to G. On the other hand, he may be cultivating too much land (see second diagram) so that it would be worthwhile to cultivate less. Until now he was utilizing an average formulation, however later p. 127 he utilizes the incremental formulation. In all these he was

very careful to limit the exposition of the law to conditions as things are (p. 127) so that improvements in arts are extraneous to the law.

THE AUSTRIANS, WICKSELL, CLARK

Menger and Wieser fail to mention the law explicitly (202, pp. 151, 161). Wicksell, Bowhm-Bawerk and J. B. Clark dealing with diminishing returns are quite elaborate but marred by their confusion between average and marginal formulations. Furthermore Wicksell and Bohm-Bawark were haunted by the desire to furnish a proof for the necessity of diminishing returns. Their ambition was to prove that the law is a "theorem of mathematical necessity." Karl Menger (see above p. 209) showed that they did not succeed although they thought they had. Schumpeter says that "the theorists have almost unanimously displayed an aversion to admitting that the "law" of diminishing returns is of course an empirical statement--a generalization from observed facts that only further observation can either verify or refute" (185, p. 588). Logic cannot prove or disprove anything here.

EDGEWORTH:

He does not spend time on philosophical matters and simply dismisses the axiomatic basis of the law of diminishing returns (202, p. 588). His contribution amounts to clearing many significant features of the Law, one of these being the distinction between average and marginal returns.

In Edgeworth, the law of diminishing returns is expressed in two senses:

(1) Primary sense: synonymous with diminishing marginal product (73, p. 63).

(2) Secondary sense: synonymous with diminishing average product (73, p. 67).

The confusion between these two formulations was nearly universal before him. He himself was often guilty of this sin, at least up to writing his articles around 1911 (202, p. 113). Analytically these two formulations are not equivalent, and many authors (185, p. 1036) (Edgeworth included) prefer the marginal formulation "as the more related to the theory of maxima," (202, p. 112).

It should be remembered that Marshall did not have difficulty with his average formulation because he did not pin point the exact level of employment of a costly factor. It is true that when only the area of factor utilization is concerned both formulations (even if used one at a time) are adequate: (at least in the case of a homogeneous function of degree one, and pure competition).

Average terms: We should not use <u>any</u> factor (in a many factors case) in the range where its average product is increasing.

Marginal terms: We should not use any factor (in a many factors case) where its marginal product is negative.

When one factor is costly, the above rules are sufficient

to determine the exact degree of utilization. But if each costs something marginal formulation is necessary. In the short-run (if factor X is the variable factor),

$$SRMC = p_X/MP_X.$$

Since the firm is in short-run equilibrium when MR = SRMC, then employment of X is determined by the marginal formulation.

Finally Edgeworth was the first to incorporate the revenue factor into the problem of area of factor utilization when he showed that the blind following of what we call today the Knight's rule could lead a monopolist not to maximize profits in the short-run.[34] He also mentioned (73, p. 71)

---

[34]Edgeworth (73, p. 73): "If the demand for passenger-service is very inelastic it might be the interest of the Railway to restrict the supply within such limits that the increase of carriage room would present Increasing Returns (in the secondary sense contemplated)" p. 73.

In modern terminology that would mean that the firm is operating in the first stage (with respect to its variable factor) of the total product curve. Furthermore, this sentence of Edgeworth is taken out of a page where all his reasoning is in short-run terms. He starts the reasoning with words like those:
"with a given road-bed and with a given equipment in the way of depots, offices, machine shops, etc." A quick survey of his Section I, Chap. 6 shows that his analysis from the beginning (p. 61) up to p. 75 is all worked out in short-run terms, e.g., one factor is fixed, except for one sentence p. 72. Later, pp. 75-76 he studies long-run problems (returns to scale, size of plant, etc.,) yet he hastens to say on p. 76 that he is not interested in this type of analysis. After this interlude, he returns to a short-run approach to production problems. One must agree with Stigler (202, pp. 118-122) that Edgeworth sometimes mixed the laws of variable proportions (one factor is fixed) with returns to scale (all factors are variable); that is he mixed short-run and long-run problems.

that in the case of perfect competition a firm would never operate in the range where average product is increasing. Yet here his reasoning is curiously in terms of long-run exclusively, while there is no need for it. (see below).

However, as we are prepared to show now, that the statement of Edgeworth holds only in the case where the production function is homogeneous of degree one. This is something which the modern authors utilizing this point[35] fail to mention. To prove Edgeworth's point requires proving that a monopolist might operate outside the ridge lines:

Suppose we are dealing with a monopolist using a homogeneous function of degree one similar to the one established previously in our theoretical section. We can draw the following diagrams:

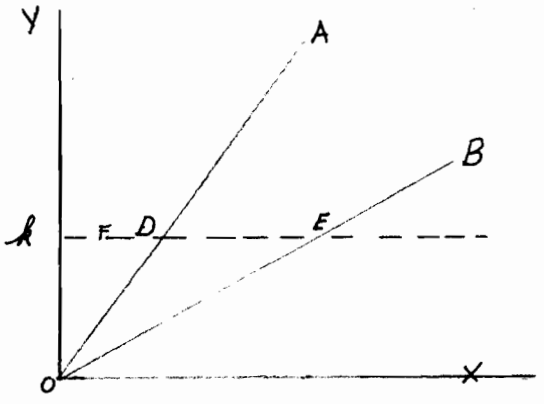


Fig. 1

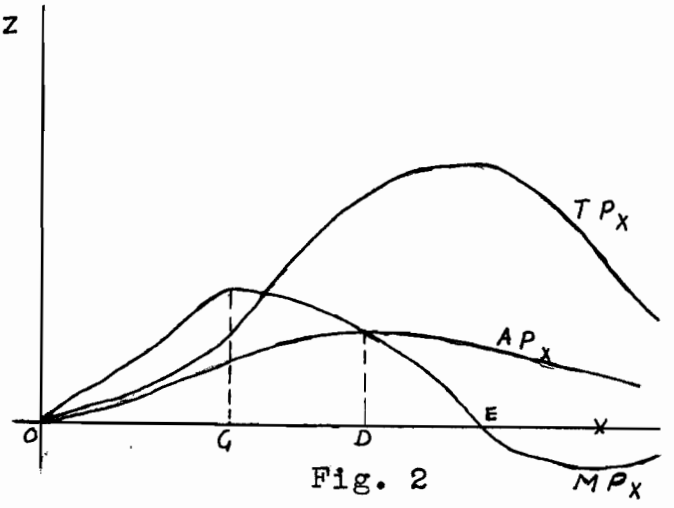


Fig. 2

[35] Stigler (202, pp. 116-117); Scitovsky (187, p. 117), Coppock (53, p. 296); Zeuthen (230, p. 116); Lerner (124, p. 154); Fellner (81, p. 67).

On OA: $MP_x = AP_x$; and $MP_y = 0$

OB: $MP_y = AP_y$; and $MP_x = 0$

Consequently points D and E in fig. 2 correspond to points D and E of Fig. 1 where in the short-run the plant size (governed by factor y) is k. Let us now transpose these data in costs terms. From fig. 2 we can draw fig. 3 with the critical values in terms of inputs (G,D,E) connected with output values.

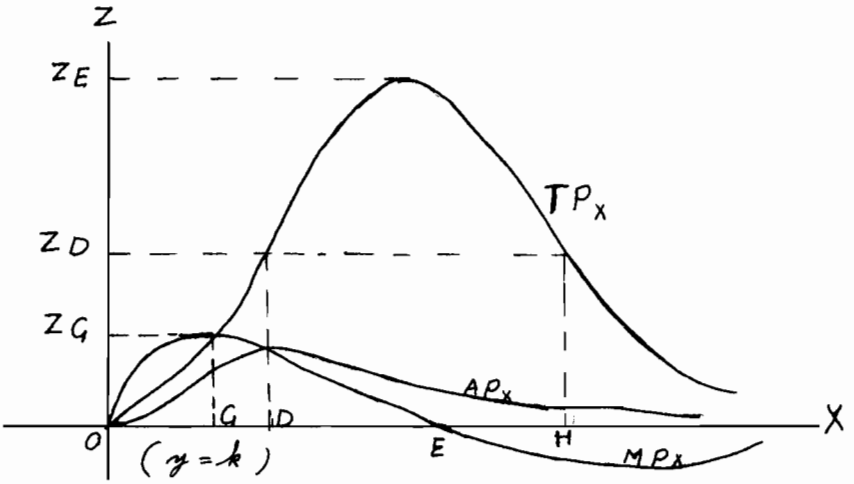


Fig. 3

Here at output:

Z_G, MP_x is maximum, Z_D, AP_x is maximum, and $AP_x = MP_x$,

Z_E, TP_x is maximum, and $MP_x = 0$.

In cost terms, output Z_G will be produced at the lowest marginal cost possible; Z_D at the lowest average cost; and no greater output than Z_E can be produced. Consequently at Z_E, MC is infinite. These properties are pictured below in fig. 4: In the terminology of fig. 2 or 3, outputs up to Z_D are produced in stage one (with respect to factor X) and

outputs from Z_D to Z_E in stage two of factor X. Consequently, point D on the ridge line OA of fig. 1 represents the cheapest way (so far as variable costs are concerned) of producing an output. The isoquant that passes through D in fig. 1 could be labelled Z_D. Similarly point E in fig. 1 corresponds to an output Z_E; in cost terms (fig. 3) it corresponds to a vertical MC.

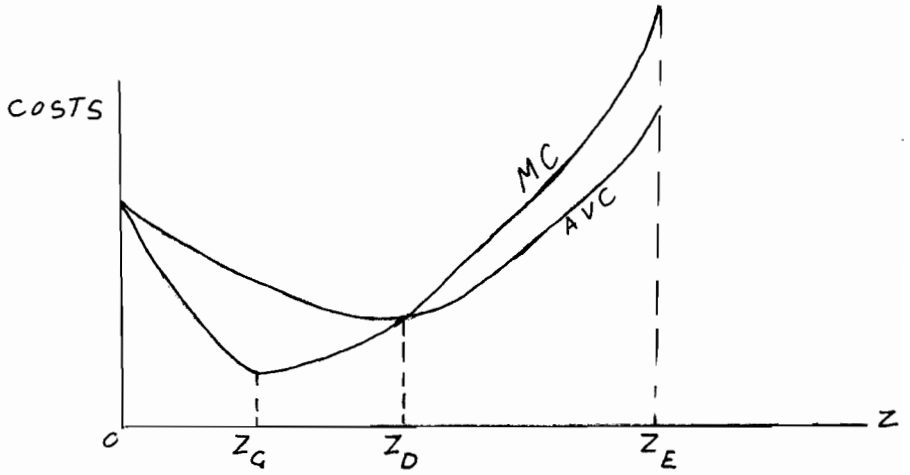


Fig. 4

Coming back to our initial assumption, that the firm in the short run is saddled with a certain plant of size k, it is nevertheless true that the firm will operate whatever its total revenue, so long as total revenue covers variable costs. It is possible for a monopolist to face a demand that will warrant operations in stage one (see fig. 5 below): In the short-run the firm will produce OF of output; since $OF < Z_D$, the firm is operating its plant in stage one. In terms of fig. 1, it is operating outside the ridge lines, (at point F). This is what we meant by the possible violation

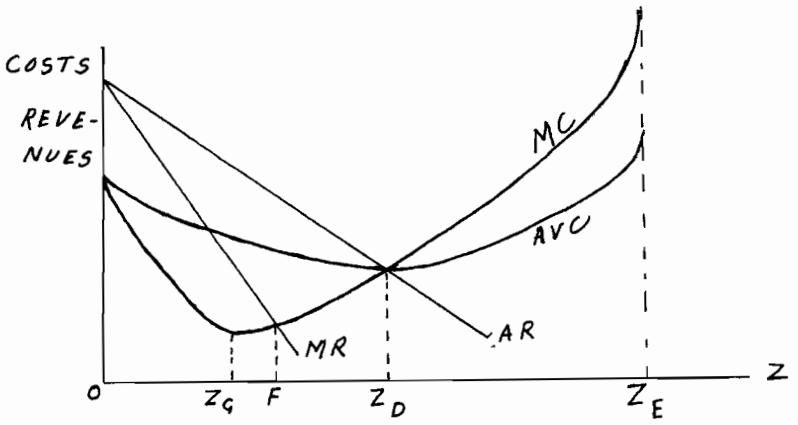


Fig. 5

of the law of area of factor utilization. We should now realize that point F in fig. 1 involves operation in the positively sloped section of the isoquant where the marginal product of factor Y is negative. Many writers, for example, Lerner, Zeuthen, Scitovsky, Stigler, dispose of this unconventional situation by stipulating that although point F is a possible solution, it will not in fact be used because we can go down to the ridge line OA (see fig. 6). By "throwing

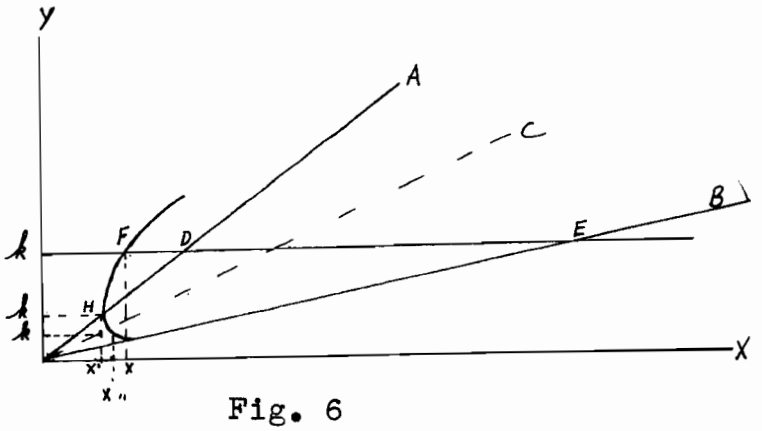


Fig. 6

away...units of the fixed service we can save...units of the variable service, still obtaining (the same amount of product)" (204, p. 117).

This is illustrated in fig. 6 where x'x of variable factor X could be saved by getting rid of kk' of the fixed factor. However, this is not a permissible move, because if the firm in the short-run is able to adjust its plant size from k to k', why not have it readjust its size to k'' which is the optimum combination of factors (because J is on the scale line OC). In other words, this would mean that in the short-run any change in demand would be instantaneously matched by the proper change in the size of the plant. This is inconsistent with the usual definition of the short-run period, which stipulates that the time period is too short to change the size of the plant.

Consequently, in the case of a sudden shift in demand in the short-run, the monopolist might be obliged to operate his plant at F; gradually, however, he might move to J. The preceding argument is worded in terms of a fixed plant where all the parts (units of Y) are heterogeneous; consequently, factor Y is imperfectly divisible and adaptable in the short-run, although it may be divisible in the long-run. This is the type of situation which prevails in most fields of economic activity, for example, manufacturing, railway transport, etc. However, in the field of agriculture, Edgeworth showed (73, pp. 72-73) that American farmers, facing superabundant land, could avoid utilizing land in its range of negative marginal productivity simply by leaving it idle. Consequently, the possibility (in fig. 6) of moving from

point F to points H or J depends upon the nature of the fixed factor. If it must be used as a unit (for example, a blast furnace, a railway, etc.) then the monopolistic firm might operate in the short-run outside the ridge lines (in stage one). If there is no organic connection between the parts of the fixed factor--if, for example, all the parts are homogeneous, then the "size" of the plant might be instantaneously changed. In other words, a firm with a homogeneous fixed factor, instantaneously divisible, does not have to pass through intermediate short-run periods to operate a readjustment in the size of its plant.

Many authors get rid of the difficulty by getting rid of the short-run period. Quite a sacrifice!

However, the above analysis proves only that the monopolist could operate in the short run in stage one. He will never operate (except if there is a constraint) in stage three because, as can be seen in fig. 4 and 5, in the short-run, whatever the demand curve, he cannot produce more than Z_E (where MC becomes vertical). Z_E corresponding to a point on the ridge line OB in fig. 1 limits stage two.

This landmark cannot be surpassed because when y is fixed at k, the firm cannot by varying x produce more than Z_E units of the product (see fig. 3) whatever the amount of X used. Similarly, in a cost diagram (like fig. 4) with output as an axis, point Z_E cannot be passed; the cost of producing a further unit of product is infinite. There is no demand curve which will induce an output greater than Z_E.

We must take account of the possibility of producing fewer than Z_E units by using more than OE of X (fig. 3)--that is, by operating in stage three (with respect to X). This is pictured in cost terms in fig. 7. In such a perspective there are two ways of producing Z_D of product: first, by using OD of X at a cost C_1 and second, by using OH of X (see fig. 3) at a cost C_2 (see fig. 7). Since X is variable and there is no constraint going beyond stage two (for factor X) is completely irrational behaviour and should be dismissed as such. In the short-run the monopolist will never operate in stage three of the variable factor. Now it could be pointed out that this might involve a contradiction because on a production diagram (at least for the type of production function we use in this paper) stage three for X is the same region as stage one for Y. In other words, if a monopolist might operate in stage one (as shown previously) he should operate in stage three because stage three for one factor is simultaneously stage one for the other factor. To take care of possible misunderstanding we must reword our statement thus:[36]

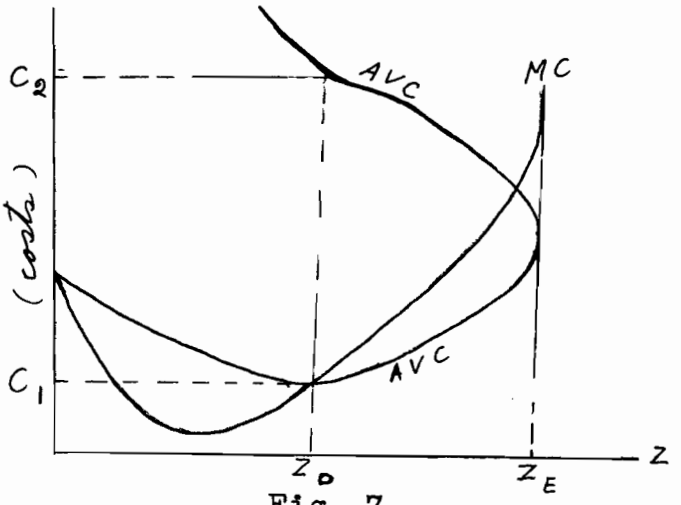


Fig. 7

[36]I owe this point to Dr. E. F. Beach.

A monopolist in the short-run might operate in stage one of his variable factor, but never in stage three of his variable factor; he can operate in stage three of his fixed factor, but not in stage one of his fixed factor. The monopolist will never use (in the short-run) combinations of factors belonging to stage three of factor X.

A purely competitive firm, since it faces a horizontal demand curve, cannot operate in stage one even in the short period. Fig. 8 shows that although p_1 would induce the firm to use a combination of factors belonging to stage one, since its shut down point is G, it will not operate in this range. Similarly, whatever the price, it cannot produce an output greater than OE. The purely competitive firm has to operate within the ridge lines even in the short-run.

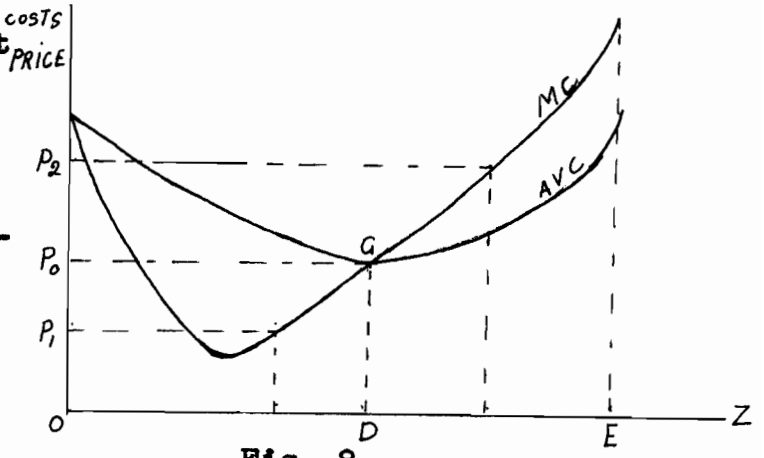


Fig. 8

In the long-run period, since both types of firms choose combinations of factors which are on the expansion path (or scale line) whatever the type of production function and whatever the market situation, they both must restrict their use of factors to combinations inside the ridge lines. The contrary situation would imply that the scale line can be outside the ridge lines.

The expansion path is always within the ridge lines because it is only in this range that the slope of the isoquants is negative; since the isocosts also have a negative slope, (as long as there is pure competition in the hiring of the factors, a hypothesis underlying the whole analysis of this paper) then it is only in this range that the isocosts and the isoquants can be tangent. Now it is well known that the combinations of x and y on the expansion path are those representing the cheapest method of producing any output. In other words, each point of expansion path corresponds to a point of the long run average cost curve and vice versa, because the long run cost curve is also defined as the locus of the points that account for the cheapest way to produce any output. Whatever the demand curve, a monopolist in the long-run will always use that size of plant whose SRAC curve is tangent to the LRAC (for a given output); and he will use that size of plant at an output which is such that SRAC for that output equals the LRAC of producing that output. This is easily seen if we draw the LRAC curve for a firm having a homogeneous production function of degree one. Such a firm experiences constant returns to scale at any output. Therefore its LRAC curve will look like the one of fig. 9. The curve will be perfectly elastic and consequently will also represent the LRMC. The effect is to have all the SRAC curves tangent to this LRAC at their minimum point. In other words, whatever the demand curve, in the

long run, the firm will always operate at the minimum of a SRAC curve. According to the reasoning of the previous paragraph, output Oq is produced by a certain plant size (suppose k). We can now go back to fig. 1 (redrawn here as fig. 10) and draw attention to a few points of interest. In the long-run, the monopolist has decided to produce Oq, and must use a plant of size k; in fig. 10 this output is produced by using x' of the X factor, and k of the y factor, where H is on the expansion path. With plant size k, we can also draw a complete cost diagram (here in fig. 11) where we know that the firm is producing Oq. Now we can establish that this point is inside the ridge lines because if we draw the corresponding SRAVC (at size k) we know that its minimum will be

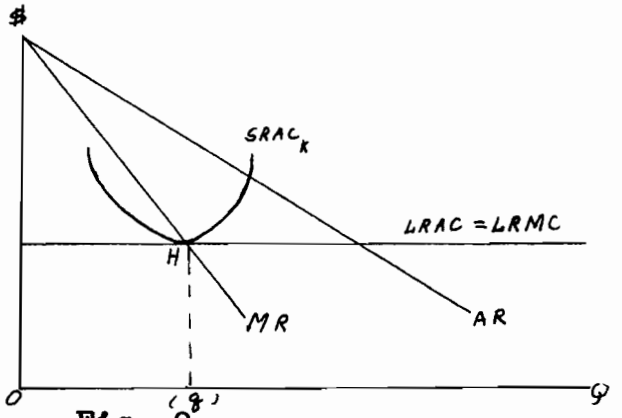


Fig. 9

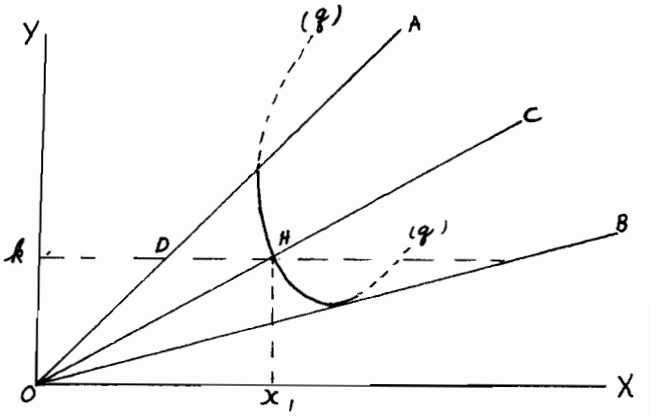
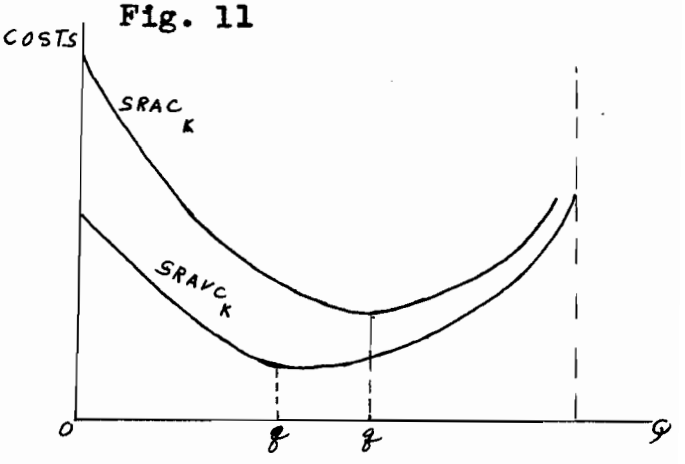


Fig. 10

Fig. 11



at an output smaller than q, (suppose q'). Having previously identified the minimum of AVC as corresponding to a point D on the ridge line OA, it is then agreed that H (in fig. 10) lies to the right of D.

Edgeworth's reasoning is of course less rigorous than the above. In the case of the monopolist, even if his reasoning has short-run implications, (73, p. 73) it is still fuzzy because he has no clear recourse to the short-run shut down point of the firm, for instance minimum of A.V.C. It was shown above that this constraint is essential to determining the area of factor utilization of the firm. This enables us to say that the 'theory of the maximum' (primary sense) is not the only rule that should be followed. In the short-run, the primary formulation should be used as long as it does not violate the secondary formulation which determines the minimum of A.V.C. We need both. Consequently we cannot follow Stigler (202, pp. 118-119) when he chides Edgeworth for "nowhere...(having made)...the flat assertion ...that average returns should not be used in short-run analysis." Our position is quite different from such an attitude. We hold that in order not to incur unnecessary losses "average returns" (meaning A.V.C.) should also be integrated into the analysis, for example, as a constraint.

Besides this important contribution, Edgeworth ventured to study increasing returns (73, pp. 81-84). But like most people he hopelessly mixes returns to scale and returns

to the use of a factor (202, p. 121). Most of his argument is worded in division of labor terms, yet is in terms of indivisibilities and as such can be a "cause" of increasing returns in the short-run.

SECTION VII: LAW OF RETURNS--SECOND BRANCH

HISTORICAL SECTION 1870-1920

Here the material is less abundant. The preoccupation with the second branch is only indirectly connected with the analysis of the individual firm. Primarily, the problem of returns is solved here at something resembling a macro level.[37] The returns to scale have been touched upon by Wicksell, Wicksteed, Walras, and J. B. Clark mainly to solve the general problem of Distribution. This might explain their initial insistence in using homogeneous functions of degree one because they wanted to be able to exhaust the total product at the distribution level.[38]

Originally Walras worked with fixed coefficient of production and this naturally prevented him from bringing in varying returns to scale. In general, Edgeworth failed to differentiate sufficiently between the two branches of the Law to be able to differentiate between "changes in the scale

---

[37]Hicks (101, p. 693) calls that the "public sector" level.

[38]Later Menger, Wicksell and Walras found out that the cost curves could be U shaped. In such a case exhaustion of the product is satisfied only at the minimum of the cost curve. To this we must add that pure competition must prevail. And the industry must be in equilibrium.

of output with given plant, from changes in the size of the plant" (202, p. 121).

With Wicksteed we meet a class of writers where the second branch is better stated. Carried away by his "proof" of the ubiquity of diminishing returns he denies that the Euler Theorem can be applied when a firm changes the size of production. This on the ground of the presence of indivisibilities. Furthermore, he restricts his analysis to the industry level; never did he use the individual firm as a subject of analysis. This is very unsatisfactory for determining the shape of the cost curves of the firm. Yet, he thought that most industries (at the industry level) experience varying returns to scale, for example, decreasing costs, while a few experience increasing costs. Since he does that without the use of the concept of external economies, this is very inadequate.

Although Wicksell does not have much on the second Branch, as far as the analysis of the firm is concerned, we are making progress. He is aware of what is implied in the change of size of plant. In the manner of Wicksteed he began by picturing the industries as experiencing increasing, decreasing and constant returns to scale, but later he considers that these "three laws" are different phases of the cost curve of a firm, rather than mutually exclusive alternatives (202, p. 375). With him "a firm may be subject first to increasing, then to decreasing returns, and the

individual manufacturing enterprises typically experiences both." (202, p. 376).

One thing is sure, at least in manufacturing; decreasing returns to scale eventually set in. The cost curve of any firm eventually goes up and consequently none of these firms will be able to secure a monopoly. He then states this condition as one essential to ensuring the existence of competitive industries. In these industries he shows that each firm will operate at its minimum cost point and furthermore that such an industry can operate (from the macro point of view) under constant returns to scale. He also shows that at this minimum point (on the cost curve of the firm) marginal productivity theory can be used (202, p. 380). This does not provide for a very wide use of this property because he does not think that linear homogeneous production functions are to be encountered very often, except in agriculture where the Euler theorem is stated to hold.

The whole thing is explained by specialization of labor which engenders decreasing costs, while "the later stage of increasing costs arises because "the advantages of centralization are outweighed by the increasing costs which are encountered when larger areas must be exploited for the provision of raw or auxiliary materials, or else for the marketing of products" (202, p. 379).

MARSHALL

In that period only Marshall has an individual firm's analysis in both branches of the law. He tackled the second branch (returns to scale) with two new concepts: internal and external economies. However, since our point of view is from the production function we cannot take into account the external economies which cannot be included in the production function (185, p. 1046).

The second branch of the law was found to be more difficult to explain than the first branch where even the classicists had scored with reasonable success. However, the classicists never faced (at the micro level) the situation of simultaneous variation of all the factors, except for West and Cournot who rightly explained increasing returns by bringing in economies due to the division of labor. Marshall was able to streamline this reasoning when he brought his representative firm. In his own words: "the law of increasing returns may be worded thus: An increase of labour and capital leads generally to improved organization which increases the efficiency of the work of labour and capital" (136, p. 265). This, according to Marshall, is true only if there is no fixed factor. In the case of raw produce (where there is a fixed factor: nature) Marshall thinks that for some time "improved organization tends to diminish or even override an increased resistance which nature may offer..." (p. 265). However, the position is reversed if the process is

continued.

Marshall's position leads him to say (p. 266) that decreasing returns to scale can set in only if some third factor (land in his case) is fixed or hidden like the entrepreneur. In the same page he abandons the dichotomy of the Laws of Returns in two branches to study two phases of the Laws (increasing and decreasing phases), which he labels wrongly Laws of Returns. These two phases according to him depend upon two different set of forces:

(1) Law of increasing returns:

It depends upon greater division of labor, better organization, buying power, and here he says (p. 266) that "the law of increasing returns acts almost unopposed."

(2) Law of decreasing returns:

It depends upon the scarcity (fixity) of a factor of production. Consequently Marshall implies that if it would be possible to increase all factors, we would have increasing returns to scale. He never contemplates a decreasing phase of returns to scale except through bringing in the first branch of the Law, for example, some fixity in one factor. He is quite prepared then not to envisage forever increasing returns to scale because any production is more or less remotely connected with land. In other words he implies that it is unrealistic to postulate forever changes in scale.[39]

---

[39]People are still arguing today about this point, that is, that decreasing returns to scale implies some fixity in at least one factor. See Peston (162) and Swartzman (211).

Marshall's world is a dynamic one, and has the effect of bringing in internal economies and external economies so that the cost curve of the firm becomes a histogram (185, p. 1046), and as such is irreversible. It is easily agreed that external economies cannot be incorporated into the production function (and thus into the cost curve of the individual firm), internal economies, (in their division of labor aspect) could. Yet this might also be problematic because division of labor in Smith's sense includes minor inventions and these are surely difficult to predict. It might then be impossible to build an a priori planning cost curve including all the internal economies.

Finally, Marshall might have been right in emphasizing the first branch because from a wide point of view, even in the so-called long-run there is always one or some factors that cannot be varied at will. In this respect he mentioned Nature and the Entrepreneur. Today other factors, for example, money capital, increasing risks, are factors limiting the growth of firms.

Yet Marshall can be criticized for having gone too far in blending the two laws, and for having recognized exclusive territories to the two branches: first branch is applied to agriculture, and the second branch to industry.

CHAPTER VI

THE FIRM 1870-1920

MARGINAL PRODUCTIVITY, COSTS CURVES, AND SUPPLY CURVES

SECTION I: MARGINAL PRODUCTIVITY THEORY

So much time has been spent on the laws of returns because they are the pillars (with the substitution principle) of marginal productivity theory, which as far as the firm is concerned has a great role. Among other things it enables the entrepreneur to draw the demand curve of the firm for any factor of production.

Under this heading, the entrepreneur decides how much to pay a factor, or what quantity of this factor should be hired if the price is given. As seen previously (p. 138), optimum hiring occurs when $p_j = MRP^i_j$, when $MRP^i_j = MR \cdot MP^i_j$. If the price is not given hiring should proceed until $MRC_j = MRP^i_j$. To build the demand curve for a factor two things are necessary:

(i) an evaluation of the contribution in physical terms of each unit of factor,

(ii) a method of evaluating this production for the firm.

This is where the laws of returns (First Branch) enter the scene. Through them a MPP curve is built for a particular

factor. For each output the particular MP is multiplied by the corresponding MR, and this yields an MRP curve which measures the worthiness to the firm of that particular unit of factor. That is the demand of the firm for this factor. We should note in passing that in the case of pure competition (in hiring and in selling) the law of diminishing returns to the use of a factor is essential to each equilibrium in factor hiring.

What precedes is the skeleton of the theory which as such cannot be applied directly because of "technical" difficulties; many of these were seen by the authors of the present period, while other difficulties formed the core of controversies in the 1930s.[1]

The first difficulty encountered stems from the marginal basis of the whole analysis: that is, the possibility of introducing minute changes in the employment of a factor. In the long-run, this type of substitution has been, in some quarters, held as possible, but it is generally denied in short-run. Substitution requires adaptation of the former factors' arrangement set up to a new arrangement. Because of the specialized character of the tasks performed by the factors, it usually takes time to readapt them to new functions.

---

[1]Here is a sample of this well-known literature. Our notes in the text outline their main points:
Knight (117), Pigou (163), Robertson (170), Hicks (99, and 100), Robinson (177), Machlup (129), Rolph (179), Lester (125).

If substitution is impossible in short-run, then the field of marginal productivity is restricted to long-run conditions.

Secondly, it has been held that it might be very difficult to disentangle the specific marginal product of a particular factor. Even in long run, since a rearrangement is compulsory, for example, providing a new unit of labor with the necessary tools; to whom should the change in production be imputed? To the new unit of labor or to the new unit of machinery, or to both? If to both, how is the share of each determined? Various panacae has been proposed, the most famous being to use, especially in short-run, marginal net productivity instead of straight marginal productivity theory. It consists in subtracting from the MRP of labor the cost of the accessories that have to be used in conjunction with labor.

Thirdly, it requires that the employers are constantly on their toes to compute the minute changes in product and vary employment accordingly. This method of operation has been denied on empirical grounds (see Lester). This put to question the whole methodology of reasoning in marginal terms.

Fourthly, even if marginalism reasoning is granted, the MRP has to be a discounted value because the production process takes time. Labor spent on a dam today will be used by the firm only in five years time. Computing MRP of labor then involves using a discount rate which has to be determined. It involves also speculating on, for example,

the price of electricity in five years. The troublesome notions of risk, uncertainty, interest, have to be introduced into the problem, blurring the issues still further.

Fifthly, the very nature of the units of factors themselves has to be brought in. To draw the marginal productivity diagram, homogeneity of the units is mathematically required. Now, it has been argued that labor is essentially a heterogeneous commodity, and that pieces of land differ in productivity, and that there is no common denominator for machines and plants. Various ways out of this dilemma have been put forward, for example devising efficiency units, wage-units, or translating the analysis in monetary units, none of which is satisfactory.

Sixthly, even if the "physical side" is settled, there remains the problem of the "price" of the product that has to be used in computing the MRP. In other words, the market structure has an influence of its own on the demand curve of the firm for factors of production.

Most of these things were more or less seen by the authors of the period, but they settled nothing. We have to wait for the 1930s for a renewed interest in the matter.

The anticipations of the theory are of course numerous. Schumpeter mentions two writers, Longfield and Thunen, who published in the 1830s and who "proved to be above their time by their writings that developed the marginal productivity principle" (185, p. 464). We could even go back to Thomas

Munn who is said to have used this doctrine in the conduct of the East India Company (21, p. 32).

However nothing systematic was done before the discovery of marginal utility principle of value. Surely the classicists, with their preoccupation with distribution at the macro-level, for example social classes, "never confronted the problem of how a given product may be imputed to the resources which cooperate in its production nor did they consider distribution as a value problem or discuss the pricing of productive services" (202, p. 152). This is because their formulation of the laws of returns was inadequate; confusion between average and marginal return prevents the correct imputation of the contribution of each unit of a factor of production (185, p. 587). Later even when the laws of returns were in better shape it was still a big step to work out a theory of imputation in money terms. The value side had to be settled. With the marginal utility theory developed by Jevons, Walras, and Menger it seems now that only one step remained to be made to change theory of value into a theory of imputation. This took 21 years, from 1870 to 1891.

Jevons was totally incapable of making it; Menger outlined the principle. With the Austrians, in the beginning it is a particular type of theory; "it was not physical marginal productivity multiplied by any price but physical marginal productivity multiplied by some consumers' marginal utility"

(185, p. 915). The theory has then some way to go before being applicable to the firm.

Walras, after freeing himself from fixed coefficients finally used market prices for goods and physical marginal product for the factors, but this was only in 1896. The theory emerges simultaneously then in the 1890s in many places; with Walras in Switzerland, Marshall, Wicksteed, Edgeworth, in England, Wicksell in Sweden, Clark in the United States and Barone in Italy (21, p. 36).

Curiously, the Austrians, after originating the idea, (with Menger) could not go very far in applying their theory of value to the productive services. For instance, Wieser and Bohm-Bawerk vitiated their analysis by using the method of fixed coefficients (110, p. 156), thus closing the door to the possibility of developing a general theory of imputation.

The theory was gradually developed, according to the inclinations of the authors and this explains the lack of unanimity in formulation and acceptance. The main objection to a complete surrender by Marshall and Pareto to the marginal productivity theory is their contention that for practical purposes (and especially in short-run), the coefficients of production are more or less fixed. Consequently it is not certain that small changes in relative prices of factors of production will entail immediate substitution among them.

Marshall with his "well-known dislike of the rigours

of static analysis" (100, p. 87) begins by talking in terms of "net product" instead of marginal product; for example, you cannot simply hire one more laborer, but you must furnish him with a spade (136, p. 337). But the quantity of factors used is clearly determined by the marginal productivity approach when he says that the entrepreneur "endeavours to employ each agent up to the margin at which its net product would no longer exceed the price he would have to pay for it" (136, p. 337). The controversy as to whether wages are determined by the margin or at the margin, does not concern us here because this is a general problem of Distribution (theory of Wages) and not primarily one of the firm. For us it suffices to say that Marshall had the "correct" approach as far as the firm is concerned. More important is his reluctance to admit instantaneous substitution between factors, because of the durability of machinery and plants.[2]

In the same vein, another dissident was Pareto, and he does not limit his opposition to variable coefficients to the short-run as Marshall does, but even to long-run since he had "difficulty in accepting the Principle of Variation" (100, p. 85). The machines that are on the market are specialized and "require a given amount of other factors to be used with them. If the amounts of cooperating factors can

---

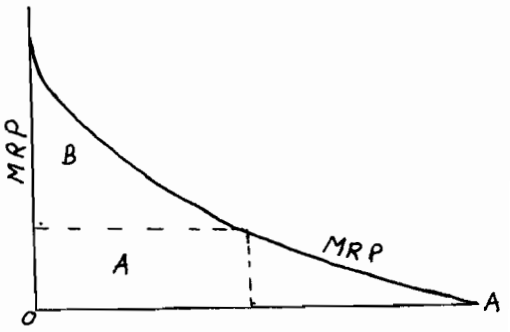
[2]"The indications seem to be that Marshall was chiefly troubled by the durability of material equipment." Hicks 100, p. 86).

be changed at all it is only within narrow limits" (100, p. 85). This opinion of Hicks is not representative of the whole work of Pareto. In the Manuel (157, p. 637) he takes great care to point out that the coefficients of production are not determined simply by technique but also by the relative prices of the factors and "en général de toutes les circonstances de l'équilibre économique."

Hereafter, the theory of imputation received more attention, notably with Wicksteed who is responsible for the first detailed and satisfactory exposition of the general theory of marginal productivity, (in his Coordination of the Laws of Distribution, 1894) (21, p. 37). He is noteworthy in that the significant measure of the marginal productivity of the factors is clearly not in physical, or in terms of social utility, as with the Austrians, but in revenue terms, and this is what is significant for the theory of the firm.

Later Wicksell (in 1900) extensively treated the subject in his article "On Marginal Productivity as the Basis of Distribution in Economics" (227, pp. 93-121). While in the United States, J. B. Clark was the outstanding propagandist. The whole work of Clark is permeated with the study of distribution within the firm through the marginal productivity theory (151, p. 285). His favorite tool is this diagram: Where A is the determinate share of Factor A, (A is bought on the market) and B accrues to the other cooperating factor

whose share is determined by the residual method (43, p. 194).



As implied previously, except for Walras, Marshall, Clark, the main purpose of the authors in working out a method of imputation is to isolate a system of Laws of Distribution. Thus they develop a micro-economic method to solve a problem at the aggregate level: for example, the share of National Income accruing to labor. Naturally they asked themselves whether their method accounted for everything, for example, whether total product is exhausted by the various shares determined by the marginal productivity theory. This lead to the famous controversy of the Euler Theorem. It was quickly conceded that as long as a homogeneous production function of degree one is used, there is no difficulty in showing that total product is exhausted. However some people, for example, Pareto, questioned the ubiquity of this type of function; Wicksell also acknowledged U shaped cost curves. But being a macro-problem, this does not concern the theory of the firm directly. The Clark diagram is much more reassuring to the entrepreneur. Simply because in such a framework, the product is always exhausted; because one agent receives a residual, for example, rent, profits. In the long run usually the remuneration of the entrepreneur

is residual, while in short-run all the agents in fixed supply receive a residual remuneration, (the quasi-rent of Marshall).

When the firm has the opportunity of hiring many factors, the least cost rule must be followed. This formula was elaborated by Walras (220, p. 385). As such, this is a major contribution to the internal equilibrium of the firm. Marshall also used a similar method when he said that the entrepreneur "if his business extends...will extend his uses of each requisite of production in due proportion; ...but not, as has sometimes been said, proportionately" (136, p. 141). In other words the production function may not be homogeneous, and substitution possible, such that the best way to increase output is to move along a sinuous expansion path. The expansion path itself is neither in Walras nor in Marshall, it had been incorporated in modern analysis recently by Frisch (34, p. 34), but it is strongly suggested by Edgeworth (73, p. 77) when he studies how an entrepreneur should behave in the case of plural factors....

Graphically some suggestive material is also found in Pareto (Manuel pp. 345-356) when he uses the box diagram.[3] However Walras and Pareto being mainly interested in the

---

[3]The box diagram itself and the contract curve are an invention of Edgeworth (Mathematical Psychics) to solve problems in consumer behaviour, for example, exchange between two people.

conditions of equilibrium, they do not, as Marshall and Edgeworth did, explore the path to equilibrium.

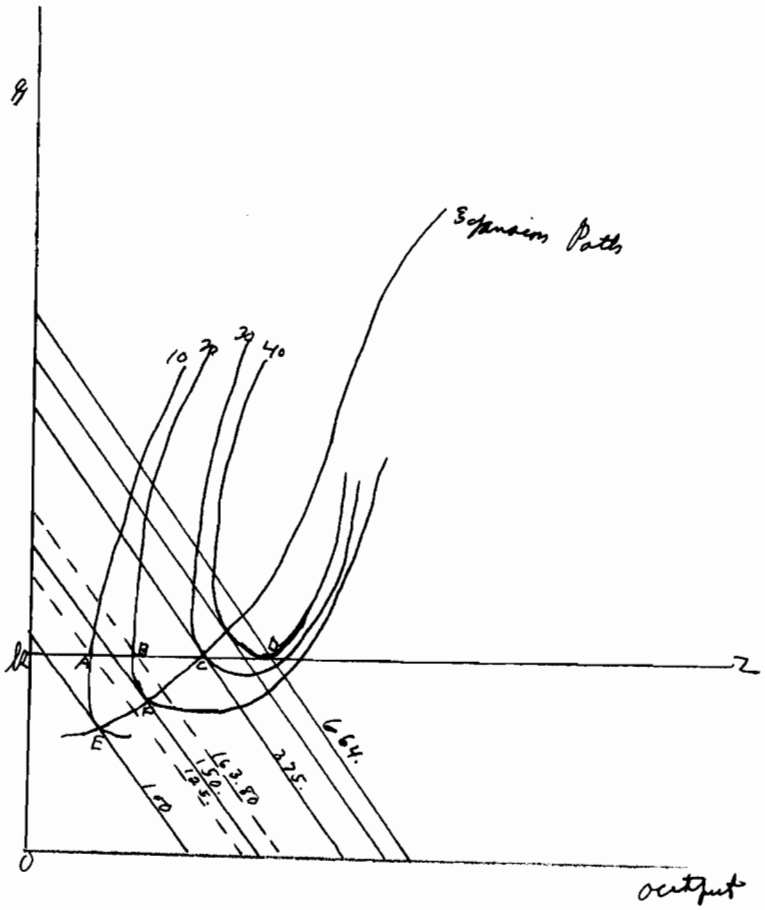
SECTION II: COST CURVES

Theoretically we are now ready to move to the establishment of the cost curves of the firm. There exist many methods of deriving them. Since 1931 (218) it is the custom to picture the derivation of the average long-run cost curve as an envelope curve which is the tangent to particular points of the short-run cost curves corresponding to the various sizes of the firm or is made of portions of these short-run curves. An alternative method is to derive them directly from the production function. The first method can be used ex post and therefore can take into account external economies, while the second is strictly ex ante and cannot include external economies. The first method is well known and need not be outlined. We will use the second one because it is a logical development of our previous work on the production function.

In the long-run, the firm is free to adjust the relative amounts of each factor in such a way as to produce any output at the lowest possible cost. Given the production surface and the price of the two factors (to simplify the diagrams we use only two factors and one output), the expansion path determines the various lowest total costs of producing various quantities of output. This information permits

the construction of a total cost curve, showing the total cost of producing each quantity of output. From this we can easily derive the Long Run average total cost curve.

Similarly we can construct the complete series of short-run cost curves by determining how cost varies along each parallel to the axis of the variable factor (in a production diagram). Suppose a quantity k of the fixed factor is available. The short-run cost curve corresponding to that size of plant can be derived by drawing a parallel kZ to the axis and establish the schedule relating the various total costs corresponding to the various outputs. This is done by finding, at the meeting point of the kZ curve and an isoquant, the value of the isocost passing by that point. In the diagram, these points are A, B, C, D, corresponding to the outputs 10, 20, 30, 40. The average total cost of these outputs are 12.5, 8.19, 12.5, 16.6 etc. Let us note that the resulting short-run average cost



curve (of size k) will be tangent to the long-run cost curve at output 30. (This corresponds to point C in the diagram above when both the kX line and the expansion path meet on the same isoquant). For all other outputs, i.e. 10, 20, 40, the $SRAC_k$ is superior to the LRAC. For instance output 10 could be produced with long-run combination E which is less costly than combination A etc. Graphically, the $SRAC_k$ curve would lie over the LRAC curve for all outputs except output 30 where the two curves will be tangent. We could similarly (by varying the parameter k) construct all the SRAC curves that could be read in this production function. The production function used above produced a U shaped curve both for the long-run and the short-run; and the S.R. curves were tangent to the L.R. curve at a point other than their respective minimum points. Only at the output corresponding to the minimum of the LRAC curve would the SRAC curve be tangent to the LRAC when both curves are at their minimum. A homogeneous production function of degree one would have produced SRAC curves tangent to the LRAC curve all at their minimum points.

The distinctions made above (long-run vs. short-run) must be traced to Marshall who was a pioneer in this matter. This was also the first time that time periods of various length were integrated in the analysis. Marshall also added a third time period: the market period where the rate of hiring of all the factors of production is fixed. There is

no cost curve in this case, because by definition the rate of output cannot be varied. Marshall is therefore quite different from his contemporaries, especially Walras and Pareto, who were mainly utilizing situations of long-run equilibrium. He also separated variable costs and fixed costs (or in his own words: prime costs and supplementary costs) (136, p. 299). This enables him to determine the short-run shut down point which is below average total cost but above average variable costs (136, p. 311). In the long-run this distinction is not pertinent (136, p. 301. n.1).

Marshall never drew these cost curves for the individual firm (except in the situation of a monopoly where he draws a L.R. cost-curve). Page 291 (Principles) he begins by utilizing the money costs for the individual firm. The probable shape of the various cost curves is difficult to establish because he used many. On page 290 a horizontal L.R. cost curve is positioned while on pages 238-239 it seems that he is operating with a J cost curve.[4] For the industry as a whole the matter is different. It seems that it can operate under constant costs, increasing costs, and decreasing costs (136, p. 388, n.1), while the cost curve of the representative firm will shift downward to the right with

---

[4]"in short there remains not very great difference between the economies available by a large and by a very large firm" Marshall (136, p.239).

increasing output of the industry. Marshall never considered the possibility of diseconomies (for the representative firm) in this last respect. The external economies seem inexhaustible as far as the industry is concerned.

In various places,[5] the short-run marginal cost curve is properly isolated. In the long run, decreasing costs in Marshall (136, pp. 285-286) are the product of a better organization, increasing ability of the personnel and the entrepreneur, and minor inventions which, in the manner of Smith, are bound to come from the exercise of a profession or from outside. Also the methods of production change (see p. 266) with changes in the size of the firm. Our derivation of the L.R. cost curve above is then not Marshallian, for example, no external economies are included in ours. All this, in Marshall's writing, accounts for a complicated production function which may not possibly be studied in a two factor framework. Also, there, time must be integrated because it may take 20 years, according to Marshall, to obtain the various combinations of the productions which are more efficient. Furthermore, uncertainty must be taken into account when drawing the cost curves (136, p. 332). On page 330, he distinguished between insurable risks and non-insurable risks, thus anticipating Knight. Finally joint costs are studied on page 323.

---

[5]Marshall (136, pp. 301, 334, 381).

The other writers are much poorer in details. For many, there is a complete identification between the behaviour of costs and the behaviour of returns to scale.[6] Most of the authors are impeded from distinguishing successfully the short-run period from the long-run period because, as we have seen earlier, they are hopelessly mixed up in the two branches of the laws of returns.[7] An example of lack of clarity on this topic is furnished by Wicksteed (228, pp. 527-529). When using the first branch he builds short-run cost curves (p. 531). At some other point (p. 531) he talks about the "scale effect", for example, "doubling any of the factors of production." Yet his exposition is so unsystematic that it is hard to find out which curve he is building. He finally settles for rising costs in agriculture and in other extractive industries, and for decreasing costs in manufacturing. Others such as Wicksell (226) come to study returns to scale (for example, second branch of the laws of returns) and translate this scale effect on the costs. On page 129 (Lectures) he says that homogeneous linear functions are an unlikely property for the individual firm and he agrees

---

[6] Although this is not strictly correct. See Schumpeter (185, p. 1045 n). Yet this identification did not lead to any error.

[7] Our approach in deriving these cost curves shows how important is a clear distinction between the two branches of the laws of returns to permit the construction of the various cost curves.

(227, pp. 122-123) to a U shaped cost curve. Yet he realizes that a homogeneous function of degree one is necessary for the application (at any scale of output) of the Euler theorem and he therefore adopts this type of function in all his deductive reasoning, for example, in his proof of the existence of diminishing returns (first branch).[8] Some other writers, such as Clark, apparently did not contribute at all (202, p. 299). Edgeworth's acquaintance with the cost curves (which he drew) stems from his translation of the laws of returns into cost terms. His reasoning is in short-run terms, for example, only one factor is variable (labor), and is accounted for. He draws (73, pp. 64-65) a total variable cost curve whose shape is adequately explained by the laws of returns. It is still an imperfect animal, for example, it is not long run, it does not take all costs into consideration, but technically, it is the best drawing of the period under survey.

SECTION III: THE SUPPLY CURVE

Even if he did not draw the U shaped cost curve for the individual firm[9] Marshall proceeded to build the supply

---

[8]Curiously enough, when he discusses the aggregate production function for a country (Lectures, p. 135) he thinks it will not be homogeneous, in that he contradicts later works by Douglas who came out with opposite results.

[9]This was done by Pigou. See Robertson (171, p. 20).

curve. By and large, Marshall furnished the outstanding contribution of the period on this topic. Continuing the English tradition, Marshall establishes the industry supply curve in two main market morphologies: (1) Free or open competition and (2) in monopoly.

(1) Free or Open Competition:

The free competition case of Marshall is defined simply as a situation of sufficient knowledge (136, p. 284) and presumably of many firms. This is more a definition of free market than a description of a real world market morphology or mode of behaviour. Hereafter, Marshall proceeds to build the supply curve of a competitive industry made of a large number of firms producing a commodity which is not completely homogeneous if it is a manufactured product, and quite homogeneous if it is a raw produce. This is done under three time periods:

(a) Market period:

The supply curve is the sum of individual supply curves of individual firms. This supply curve will be perfectly inelastic if there are no possibility of storage (137, p. 183), or not perfectly inelastic if storage is possible and expectations about the future can be formed (136, p. 277).

(b) Short-Run period:

Here also the industry supply curve is the sum of individual supply schedules of the particular firms (88, p. 506), where any particular supply curve is identical to

the marginal cost curve of the firm, at least for the portion which is superior to average total cost curve. For the portion which lies below the ATC curve, the supply curve of the firm proceeds for some distance along the MC curve, but at a certain critical price (always superior to the minimum of AVC) it heads for the vertical axis, no supply is offered below this price, because the seller does not want to spoil the market (88, p. 501). Let us note here that this is not compatible with simple quantity adjustment which is the mode of behaviour usually associated with pure and perfect competition.

(c) Long-Run period:

The long-run supply curve for the industry is, in Marshall, determined quite differently.[10] It is not arrived at by the arithmetical sum of the supply curves of the individual firms. Marshall does not use directly the individual firms as the basis to establish the industry supply curve.[11]

---

[10]His invention (the representative firm) to deal with this situation will become the subject of a long controversy in the period 1920-1930, which stopped momentarily with the appearance of Chamberlin's and Mrs. Robinson's works but which was revived by a series of articles on Marshall: Frisch (88); Robertson (171); Wolfe (229); Hague (94) and Maxwell (140).

[11]Marshall (136, p. 380): "Thus the history of the individual firm cannot be made into the history of an industry any more than the history of an individual man can be made into the history of mankind."

In general markets for manufactured products, Marshall thought that any increase in demand would bring a bigger supply. On the other hand he believed that increasing the output of the industry would bring external economies to the firms member of this industry. All this adds up to the possibility (at least for sometimes) of having a downward sloping curve for the industry. He does not establish it directly but through a landmark: the representative firm. This firm (which does not necessarily exist in real life) is a model depicting what happens to the key variables in the industry analysis. On this point, the model can be "representative" according to the size of the output or to the cost of production. Marshall in many occasions said that the representative firms would be representative of both[12] size and cost.

Furthermore Marshall distinguished two types of free markets; (i) General markets where every firm can more or less compete simultaneously with all other firms. (ii) Special markets where the industry market is the sum of all the special

---

[12] Re Size: see (136, p. 381). "We also expect a gradual increase in demand to increase gradually the size and the efficiency of the representative firms; and to increase the economies both internal and external which are at its disposal."
Re Size and Costs: see (137, p. 509): "These tendencies have constantly increased the size of the Representative business unit--that is, of the unit whose costs of production exercise so great an influence on the amount that will be forthcoming to meet any given market demand that they play a prominent part in the regulation of price under competitive conditions."

markets of individual firms.[13]

(i) General Markets:

(A) Size:

When it is representative as to size, the representative firm has an output which is a certain proportion of the total output of the industry. If the output of the industry changes the output of the representative firm changes proportionally. Thus far there is no quarrel, since the representative remains a statistical model. However from this imaginary firm, Marshall moves to a model which would possess all the characteristics of a real firm, for example, mode of behaviour. The representative firm is no more a statistical device, but a hypothetical firm which would have attained a stage in its development such that it "has normal access to the economies external and internal which belong to that aggregate volume of production" (136, p. 265). Therefore "if we want to know how total supply will react, we may simply study how the representative firm will react."[14]

---

[13]See Hague: (94, pp. 679-680) for various quotations from Marshall on this topic.

[14]Frisch: (88, p. 512). Marshall did not say that but says that if we want to analyze the normal cost of producing a commodity "we shall have to study the expenses of a representative producer" (Principles p. 264).

This personification leads the concept of the representative firm into difficulty.[15] In the case of general markets and perfect homogeneity of the factors of production (or if there is non-homogeneity if they are successful in collecting their rent), the concept of the representative firm is redundant because all the firms will be of the same size. There is also a second line of objections: it is not certain that the size of the representative firm will change the same direction as the output of the industry changes.

Let us take the case (which is not Marshallian) of a constant cost industry. Starting from an equilibrium position, if there is a change in demand, in the short-run the representative firm will increase its output along its MC curve, it will earn extraordinary profits. It is agreed that its output will increase in the same proportion as the industry output will increase. However in the long-run new firms will come in, their output will increase the industry supply and gradually the price will drop to its former level. In the meantime the representative firm is decreasing its output to the former level. In the long run then, the industry output will increase while the size of the representative firm will remain stationary. (Note that in industry equilibrium, the representative firm is also in equilibrium as

---

[15]Robbins, L. (175). He labels it as redundant and misleading.

far as internal economies are concerned).

Let us take the more Marshallian decreasing cost industry. With a change in demand, as before the representative firm will for sometimes be representative as to size but time will bring two changes: new firms in the industry and a change in all or in some of the prices of the factors of production (economies of scale). In this case there will be a downward shift of the long-run cost curve of the representative firm but it is not a certainty that it will always be a downward shift to the right as Marshall supposed. If the economies are in the form of cheaper and bigger machines not available before (as Marshall thought) the representative firm will then grow in size. But it has happened, in some fields, for example, gas filling stations, that technical progress and increase in demand lead to a greater number of these units, and not necessarily to bigger units. In the long-run then there is no assurance that changes in industry output will change the size of the representative firm in a particular direction. In this, we then cannot agree with either Frisch[16] or Wolfe (229, p. 346) who repeated Marshall by stipulating that "the size of the representative firm depends on the output of the industry and the price at which its products are sold. For each level of output there exists

---

[16]Frisch (88, p. 520): who shifts the AC curve of the representative firm, downward and to the right.

one price at which the size of the representative firm is in equilibrium and for each price one output. The locus of these price-output combinations is the supply curve."

Contrary to that, the present consensus of opinion is to build the supply curve of the industry outside data of this representative firm. To build the industry supply curve, one needs the concept of equilibrium of the industry which is that no firm wants to enter or leave because of extraordinary profits or losses experienced by the typical firms. This is compatible in some respects with Marshall because this formulation permits the entry of new firms to replace the old firms which are dying. When the industry is in equilibrium in this fashion, the industry output is matched with the ruling price and we have one of the points of the industry supply curve. Utilizing Wolfe and Frisch approach would produce a perfectly inelastic supply curve in the case of a constant cost industry.

(B) Unit Cost:

If the representative firm is representative as to unit cost of production, its lowest point on its long-ruh average cost curve must coincide with the price. This feature was better received than the preceding one (140). Indeed it is essential to determining the equilibrium of the industry. We have defined above the equilibrium of the industry as the situation where the representative firm does not make any extraordinary profits or losses. The cost of production of

the representative firm is then essential to determining industry equilibrium and thus to build the industry supply curve. The representative firm is a way to introduce normal profits while simultaneously maintaining diversity in the sizes and ages, and profitability of the firms that compose the industry.

We have an idea here of the burden that Marshall imposed on himself when he chose to work with the realist hypothesis of non-homogeneous firms. Non-homogeneity in firms renders meaningless the concept of normal profits, therefore a standard of reference is needed to give them any meaning. This is done by taking one type of firm (representative) as the standard. However, this entails two subsidiary postulates: (1) prospective newcomers or leavers (other than those who must naturally leave because of age) base their calculations on this representative firm, (2) there is an independent flow of newborn firms which exactly offset the output lost through firms that die through senility. Here the new entrants are attracted not by extraordinary profits but simply to replace outgoers. In other words the size distribution of firms is constant at a particular price (229, p. 347).

(ii) Special Markets:

Special markets means some differentiation, although it may be slight (firms can invade each others markets). An example might be found in retailing where differentiation

might be only in spatial terms. What has to be determined is the possibility of constructing an industry supply curve. One necessary element is the concept of industry. This concept was made fashionable by Marshall; it was used extensively by Chamberlin [illegible], but thereafter came under severe attack, and was finally dismantled by Triffin (214, pp. 85-90). In Marshall, industry is a rather loose concept that covers all sellers which are in competition through a single commodity and sell at a common price. Marshall talks about the retailing of shoes in a city such as London. As in the case of general markets, the industry supply curve can probably be drawn by adding the particular outputs of all the firms (for every possible price). However we must note that the demand curve for any of the firms is no more perfectly elastic. Furthermore, is it possible to draw the industry supply curve by using a short-cut, for instance, the representative firm?

Usually an important degree of differentiation renders meaningless any "collectivization" of individual demand curves. However in the case of Marshall differentiation is very light; for example, it may make sense to talk about the quantity of shoes of a certain model that can be sold at a certain price in a city such as London. Therefore, it might be possible for a firm to be representative of the industry supply conditions, but not always on both counts: size (output) and price (cost of production).

In the short-run, even if the demand curves are slightly sloping downward, it is easy to see that the representative firm will be representative on both counts if demand changes equally over the whole territory covered by the industry.

In the long-run a few cases must be distinguished. To begin with, let us establish the equilibrium condition. The representative firm would be one serving a certain area containing enough customers to enable it to exploit almost all the internal and external economies available. Since it makes only normal profits, it is operating at an output where its AR equals its AC. However, it has been said that in the case of special markets, the AR curve is sloping downward.[17] This can be obtained in two ways: (i) the AR crosses the AC curve, (ii) the AR curve is tangent to the AC curve. A solution in terms of (i) is not acceptable because it violates the Cournotian condition of equilibrium of the individual firm, (MC = MR); in other words, the industry could be in equilibrium while the individual firm would not. Yet this is the solution put forward by Frisch.[18] Alternatively (ii) meets the Cournot condition but the representative firm will not operate at its lowest cost output. This difficulty can be reduced in importance if the demand

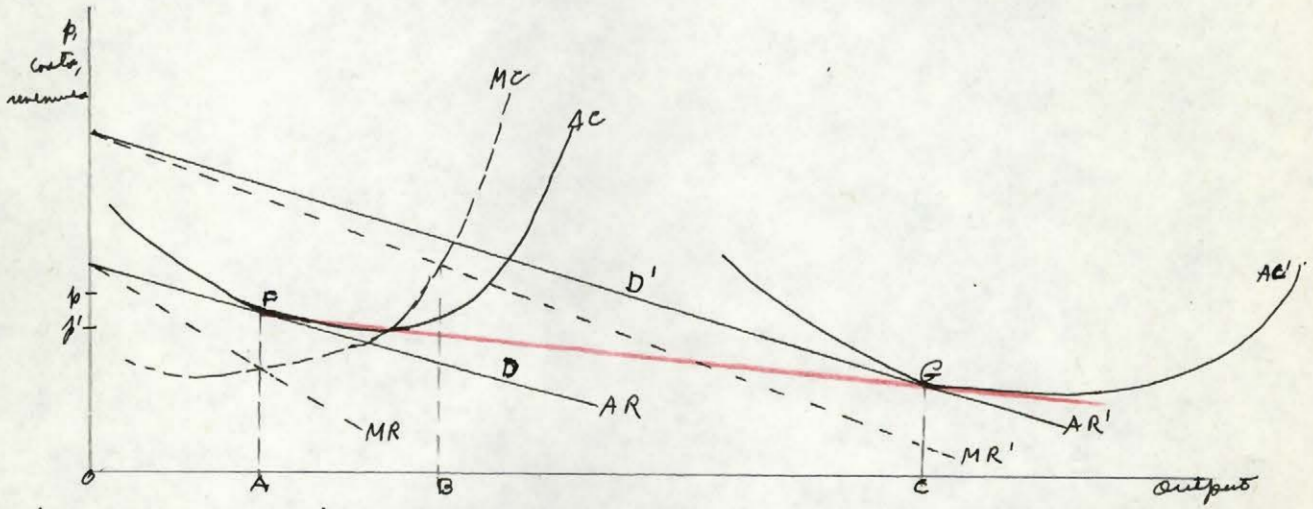
---

[17]This AR curve corresponds to the DD' of Chamberlin. This is a situation where every price cut by the firm is met by the competitors. This is also in accord with Marshall postulate of a free market where only one price can rule over the industry.

[18]Frisch: op. cit. See his fig. 7 p. 520. He makes the demand curve cross the AC curve when this AC curve is at a minimum.

curve facing the individual firm is made very elastic (as opposed to Frisch who exaggerated the degree of differentiation), as Marshall supposed it. And if the cost curve is flat bottomed like the J curve to which he referred. In this case both the price and the output of the representative firm are not appreciably different from the solution obtained in pure and perfect competition. So that we may be justified in saying that also in the case of special markets the representative firm enjoys almost all of the economies.

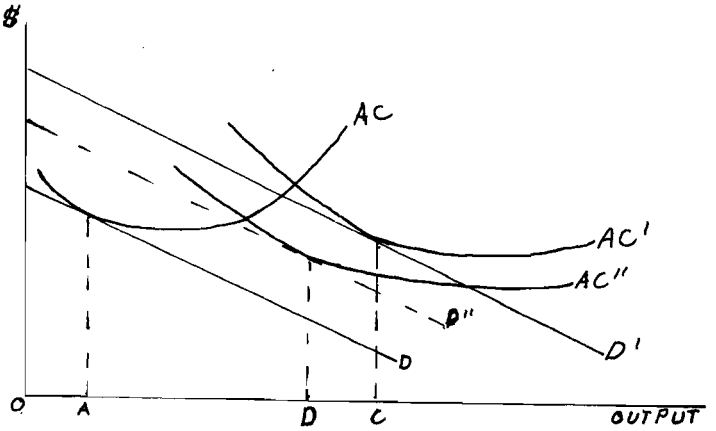
Let us now introduce an increase in demand (each customer has its demand curve shifted to the right) under no new entry. Each firm will continue to serve the same territory



(special market) as before, but its demand curve will shift to the right. (To D' in the diagram). In the short run the output of the representative firm will increase to OB. Yet, as all firms increase their output in this fashion (Marshall, p. 381) external economies are produced, therefore the AC curve is replaced by a new one having a minimum lower and to

the right of the previous one.[19] The former cost curve being no more available (the new one being AC') the representative firm will have to reduce the price to meet its costs and make normal profits. OC is the new equilibrium output and Op' the new equilibrium price of the representative firm. In this model, the representative firm is truly representative of both cost and output and price of the industry. Using the Frisch method, (joining points F and G), a miniature of the supply curve of the industry can be drawn from the data of the representative firm.

We may now study increase in demand coupled with free entry. Here the representative firm loses its representativeness because under new entry the market area of the representative firm is reduced so that it must operate with a different demand curve which stands between D and D' of the previous diagram. Also, because both the number of firms and their output are different, the external economies will be different and so will be the cost curve. The representative firm output is now OD which does not bear the



[19] As said above there is no compelling reason to envisage exclusively this case, but the whole Marshallian analysis seems oriented toward that single alternative.

same relation to the output of the industry as before because the industry output increased not only because the former firms increased their output in a certain proportion but because new firms came in and tapped a new demand.[20] Now, since Marshall admitted both slight differentiation and free entry, his concept of the representative is not completely operational (it is representative only as to price and cost, not for the size).

2. The Second Market Situation Envisaged by Marshall is Monopoly.

In Marshall the internal mechanism of the individual firm is better illustrated in this chapter on Monopoly. However in Marshall it comes after much has been said about the firm (especially the representative firm) and it is too late to incorporate the new mechanisms into the former analysis. The firm and the industry are now identical; the industry demand curve (which is downward sloping) becomes the demand curve of the firm. Similarly the average total cost curve (which does not include normal profits but an imputed interest charge on capital furnished by the shareholders) becomes the 'supply' schedule.[21] Finally the confrontation

---

[20]This is why it is hard to compare both situations. In a spatial model the aggregate output is a function both of the price and the number of firms. The explanation lies in the different c.i.f. prices which vary with the number of firms. See below footnote No. 11, p. 333.

[21]This is not a supply curve as ordinarily understood.

of these two schedules permits the determination of the output which will maximize profits. The procedure is not Cournotian, that is, no MC and MR are used, but Marshall's method leads to the same solution.

Duopoly and oligopoly are supposed to be inimical to Marshall's analysis.[22] Yet there are three instances where oligopoly analysis is implied: (i) In his market structure called "special markets" there is probably an element of oligopoly.[23] Strictly speaking any model of light differentiation which is akin to a spatial model is in reality an oligopoly (chain oligopoly) model. (ii) The reluctance of the entrepreneur to spoil the market (often mentioned by Marshall) is also another type of behaviour akin to oligopoly.[24] (iii) Finally interdependence with potential rivals is recognized (thus involving an oligopoly situation) when the entrepreneur sets his price in such a way as to forestall entry (137, pp. 396-397). Yet it remains that although somewhat implied, oligopoly is not systematically analyzed even if large sections of Industry and Trade are devoted to the

---

[22]"Marshall rules out this type of oligopoly--(oligopoly in the strict sense of really few sellers)--from the analysis of the Principles"(Hague, 94, p. 678); see also pp. 683-685).

[23]Robertson (171, p. 15) is not sure that this necessarily implies oligopoly.

[24]This has prompted people such as Hague (94, p. 681 ff.) to draw a dd' curve à la Chamberlin.

description of cartels and other oligopolistic devices. Similarly, in the Principles (p. 408 ff.) an esoteric case of duopoly is brought in, giving Marshall a chance to criticize Cournot's fundamental equations. In the Principles, he stipulates that the supply of the industry in the case of fewness, for example, "few large firms" (136, p. 329) cannot be drawn.

Characteristics of the Marshallian Firm:

After this short exposé, one may venture to list the characteristics of the Marshallian firm:

(1) In most cases it is a quantity-adjuster (171, p. 26), as opposed to a price-fixer. However, the exceptions are numerous, for example, monopoly, some instances of special markets, when spoiling the market occasions present themselves, price-fixing by cartels. Yet in the normal course of events the price is established at the industry level and is supposed to rule even when differentiation is slight. The shape of the demand curve is then slightly sloping downward for the monopolist and for the representative firm in the case of special markets. In the other case of general markets where there is no differentiation, the representative firm can be ruled out and the firms face a perfectly elastic demand curve.

(2) Another characteristic of Marshall's firm (especially the representative firm) is its biological nature. This characteristic is extremely important for any appraisal of contemporary controversies. Marshall accomplishes a break

with the Cournotian firm which is more akin to a mechanical model. The break is not complete because Marshall also dealt with mechanical problems. He justifies both types of model but he notes (136, p. 329) that the mechanical model "does not deal constructively with real problems." Another difference between the two models is that one (the mechanical) is concerned with static equilibrium, while the biological model is dynamic (deals with time) but is not concerned with equilibrium. Also, the firm is not individualized to a great extent. The end purpose of Marshall and of the English tradition is to use the firm merely as a canal for the forces of demand and supply. Continually, he takes its internal structure for granted. He treats numerous particular aspects of the firm but rarely gives a comprehensive view of the firm. He is so familiar with it that he does not feel the necessity of describing it systematically.

Marshall's industry model accounts for firms of various sizes and the explanation of this diversity of size of firms is not in terms of costs and revenues. For instance, in Cournot there is an optimum size of the firm which is determined by costs and revenues. In the case of free entry, the prospective entrant will build his plant according to this particular size. With Marshall, firms do not choose a particular size, but <u>grow</u> to a certain size (representative) passing more or less rapidly by the intermediary sizes, and sometimes dying before attaining that representative stage.

Once they have attained that stage they will not be able to stay there for long but inevitably they are drawn toward senility. In such a framework, the Cournotian equilibrium does not make much sense; this is why Marshall did not use much of it. In a Marshallian, firm it is almost impossible to draw ex ante the long run cost curve because it is impossible to know the extent of external economies which will appear gradually as in a growing process.

In addition, the Marshallian firm, contrary to Cournot, is not timeless. We have said that firms cannot immediately attain the optimum size. Presumably it takes time to hire the proper people and resources. It also takes time to coordinate these factors of production. Furthermore markets are exploited through time. Similarly the entrepreneurs' proficiency changes with time.

All these features of time permit Marshall to utilize simultaneously a decreasing cost curve for the firm and free competition. The increasing returns do not necessarily lead to monopoly because firms like the trees in the forest do not grow forever but in the end die (136, pp. 315-316). This is called today the life cycle approach to the theory of the firm, and it manifests itself in the height and position of the AC curves.[25] The biological analogy is not simply a

---

[25]See an illustration in Frisch (88, p. 503). Here the weak firm corresponds either to a newcomer or to a senile firm. The average firm corresponds to the representative firm, and the strong firms corresponds to an extraordinary firm.

metaphore in Marshall. He links the economic fate of the firm with the biological faith of the entrepreneur who runs it. To the same degree that the energy and productivity of a human being changes with age, so will be the efficiency of the firm this human being is running. In the last editions he began to realize that corporations with their unlimited life and their facility to replace aging managers, somewhat reduced the importance of this thesis.[26] Corporations are now permitted to stagnate instead of dying.

(3) A third characteristic is the small role played by marginalism in Marshall, although it is not absent. The marginal approach is more evident in the short period (136, pp. 309-314). Some long run MC is occassionally referred to, but the main use of the MC curve is in short run.[27] Yet he knows the existence of both MC curves (long-run and short-run).[28] In the long-run it is the average cost per unit which governs the supply price; yet here also the individual producer is pictured "as reckoning up whether it would be

---

[26]Marshall: (136, pp. 263-264). This had been signalled as early as in Robbins (175, p. 389).

[27]In short-run the shut-down point is based on marginal analysis. See Marshall Principles pp. 298, 289, 300.

[28]"the marginal costs of a particular producer at any time may stand in no close relation to marginal costs under normal conditions. Marshall (136, p. 334).

worthwhile to add a certain new line to his undertakings. We regard him as treating the output which would result from that change more or less as a unit, and weighing in his mind the cost against that gain." (136, p. 381). His marginalism remains, then, very rough.[29] Marshall as such probably does not qualify to become a member of the marginalist club of Machlup.

(4) Finally Marshall has no clear cut analysis of the individual firm in free and open markets. Nowhere (except in monopoly) does he draw the curves pertaining to the firm. His main model is the industry for which demand and supply curves are drawn. He does not spell out the relations of the individual firm with the industry except in the case of the representative firm. This lack of clarity enables him to include extraneous elements such as the flesh and blood entrepreneur which accounts for the diversity of sizes in firms. The integration of a flesh and blood entrepreneur is surely a distinctive feature of Marshall and contrary to the French tradition absorbed by Walras and Pareto, the entrepreneur is not simply maximizing profits. The entrepreneur of Marshall is a businessman who coordinates factors of production, takes risks, furnishes the capital or borrows part of it. The quality of the entrepreneur determines the

---

[29]See: The entrepreneur "regards an increase in his processes of production, rather than an individual parcel of his products as a unit in most transactions. And the analytical economist must follow suit, if he would keep in close touch with actual conditions. These considerations tend to blur the sharpness of outline of the theory of value: but they do not affect its substance" (136, p. 312).

status, length of life, and speed of development of the firm.[30]

Similarly, his market structures are not well defined and integrated into a general schema. In the Principles the dichotomy is between free competition and monopoly, with some inklings of digopoly. Already free competition is a mixture of competition and monopoly for example, special markets. On this topic Industry and Trade is quite rich.[31] Marshall appears to have anticipated Chamberlin and Mrs. Robinson when he speaks of the "interlacing of monopoly price policy and of competitive price policy" (126, p. 341). Besides, when studying Railways and Cartels he was brought to introduce price discrimination. He also pointed out the importance of product differentiation, advertising and other devices to shift the demand curve of a firm. However he mentions these things without incorporating them into an analytical schema.

SECTION IV: THE THEORY OF THE FIRM IN WRITERS OTHER THAN MARSHALL

(1) Internal Equilibrium

The alternative to the Marshallian firm in this period might be (among others) the Walrasian and Paretian firms.

---

[30] Marshall, (136, pp. 269, 238, 366, 250, 252, 262, 295).

[31] See a profusion of statements in Liebhafsky (126, pp. 339 ff.) where [redacted] the excerpts illustrate how Marshall was aware of many of the modern refinements in the fields of monopolistic competition and imperfect competition.

These firms are much different in that they are less individualized and their entrepreneur is completely "automatic".

Walras' macro-economics is a generalized micro-economic system. The individual firm is present in Walras mainly as a maximizing center automatically registering the various conditions of costs and revenues and reacting in an automatic manner to any variations in these data. To employ a modern comparison, it may be said that the firm of Walras is "computerized". Walras and Pareto had much difficulty in reconciling themselves with the individual firm per se because their approach was inimical to partial analysis (185, p. 991, n. 3). Walras (220, pp. 483-488) spends some time discrediting the attempts of Auspitz and Lieben to draw a demand curve and a supply (cost) curve for the individual firm. According to Walras, we cannot draw the demand curve because any change in price will put the system into disequilibrium, including the demand for the "numéraire".

Notwithstanding that, Walras was brought to make use of more rigorous models than Marshall on many particular points connected with the internal equilibrium of the firm. For instance, he was the first to study rigorously the influence of a change in price of the factors of production on their quantity bought by the firm (166, p. 247), and as mentioned previously, he also discovered and "understood that minimizing cost and maximizing profits were not the same thing; the first is implied in the second, but the second is not implied

in the first and may be considered independently" (220, p. 552).

With some other authors, for example, Jevons and Wicksteed, the individual firm is non-existent. Wicksteed denies the existence of a supply curve (110, p. 102). On the other hand, Mengoldt "draws and explains various differently shaped supply curves: a horizontal straight line represents constant costs; a horizontal line, rising abruptly vertically at a certain point is the case of constant costs up to the limit of a certain rigidly fixed supply. Finally, a rather flat U shaped curve falling over a certain output owing to economies of large scale, and then subsequently rising is explained" (110, p. 134). Clark (43, pp. 248-249) contrary to Marshall, introduces a detailed analysis of the plant.

The literature[32] also mentions the outstanding example of the analysis of the individual firm by Auspitz and Lieben, which goes as far as to determine the demand and cost (supply) curves of the individual firm (in total terms), study the structure of assets of the firm (an item which will be taken up by contemporary theory) and give an exposition of various market structures and modes of behaviour. In Pareto, however, the size of the firm is determined by the managerial limit. (166, p. 334).

---

[32]See Hutchison (110, pp. 188-191); Schumpeter (185, pp. 844, 849); and Walras (220, Appendix II).

(2) External Equilibrium

In the field of external equilibrium (market structures and modes of behaviour) the analysis is no better in the authors other than in Marshall. The emphasis is on a brand of competition which is more perfect than in Marshall. Also, as in Marshall, a chapter is usually reserved for Monopoly. Atomistic competition is usual in Wicksell (202, p. 267). This does not hamper him from discussing the solutions of Cournot, Bertrand and Edgeworth. Like Marshall, he says that there is no sharp distinction between competition and monopoly (110, p. 232). The same climate prevails in Walras and to a much lesser extent in Pareto. To obtain a solution to his general equilibrium system, Walras had to simplify tremendously: fixed coefficients of production, instantaneous production, perfect knowledge and free competition were the technical requirements of his models. The real world simply introduces friction and imperfection in such a context; to that extent, the final solution must be altered. This perspective is inimical to a Chamberlinian approach which would introduce these "imperfections" as integral components of an ideal model. Walras' goal is to establish scientific principles in a case of commodities produced under the condition of free competition to satisfy individual wants (as opposed to social wants). This theory has no predictive or even descriptive powers (220, pp. 225-257). That is why he studies monopoly in a separate Lesson (No. 41). Paradoxically it is in this lesson that the conditions of free

competition are clearly stated: (i) price equals cost of production, (ii) a uniform price prevails on the market. While the monopolist can vary the price but has no control over the quantity that can be sold at each price (220, pp. 436 ff.). Finally monopolistic situations permit price discrimination.

Pareto's contributions are more extensive. He (157, pp. 288 ff.) produces a systematic classification of modes of behaviour. His type I corresponds to our own polypolistic quantity-adjuster mode of behaviour, where the individual firm does not attempt to modify the ruling price. In Pareto this mode of behaviour is encountered in the market structure "libre concurrence". His second mode of behaviour, type II, corresponds to monopolistic price fixing. This mode of behaviour is encountered in the market structure of monopoly. Besides that, the firm may have many goals (he does not elaborate) but he says that we should confine ourselves to the normal case: maximization of monetary profits (157, pp. 288). He follows with an extensive study (157, p. 290) of the various means at the disposition of the firm to attain its goal. Although his apparatus, for example, "courbes de poursuites" is extremely awkward, his themes are refreshingly modern. Pareto is definitely reasoning while starting from the individual firm. Among the tools to modify market conditions, product differentiation is expressly mentioned (157, p. 602). Duopoly (157, p. 598 ff.) and cartels are also mentioned but

there is no contribution there. He simply says that there is no solution that can be arrived at through economic reasoning in the case of duopoly (157, p. 601).

Edgeworth's contributions are more technical and less rounded off. There is no attempt to build a complete theory of the firm. He is mainly remembered for his discussion of duopoly where he discusses the Cournot problem with a different behaviour hypothesis; for instance, the duopolist supposes that his rival will keep his price constant (73, p. 119). His solution as is well known, is in terms of an oscillating movement, whereby the price dwindles downward and suddenly jumps back upward. This unstable situation is compared (110, pp. 313-4) to a game situation. Allusions to this approach are also found in Fisher, Pigou, and Bowley. He similarly arrived at an indeterminate solution in the case of bilateral monopoly (a case also mentioned by Marshall). This is a situation which occurs when two isolated exchangers meet. For this purpose he devised what is called today the box diagram which depicts through indifference maps the preferences of the two parties, and enables the drawing of the contract curve. He also rigorously and extensively studied monopoly, improving on the classical economists who (he said) "rather anathemized than analyzed monopoly" (73, p. 172). His proficiency in monopoly analysis is demonstrated in his study of the effects of taxation on a monopolist, (a case also studied by Walras and Marshall) and in his exposition

of price discrimination which is indeed an advanced problem in pricing (73, pp. 172 ff.). Price discrimination is an example of real world commercial practices systematically incorporated rather early in formal economic analysis.[33]

One cannot leave the review of the state of the concepts regarding external equilibrium of this period (1870-1914) without mentioning the early criticisms on grounds of vagueness of the concept of competition by Moore (110, p. 309).

The Entrepreneur

The entrepreneur received much attention during this period, but not in a way as to make of him an operational factor in the theory of the firm. In Walras it is described but not integrated. His entrepreneur makes "ni bénéfice ni perte" (220, p. 225). The presence of the entrepreneur does not change the goals of the firm; it is then only a form of specialized managerial labor. We are once more confronted with the Cournotian automatic entrepreneur which simply makes sure that the firm is in equilibrium in the two markets where the firm operates, i.e. products and resources. Profits and losses are the signal that the firms and the industry are not in equilibrium. This signal will provoke a readjustment both in the number of firms and in their output. Where

---

[33]The pioneering writings on this topic are made by accountants and railway people. Price discrimination is also found in Walras.

profits are earned, the industry is not in equilibrium and new entrepreneurs will flock in (220, p. 225).

This approach is typical of all the writers of the period who cared to mention the entrepreneur in their competitive system. When we move to monopoly, the entrepreneur remains Cournotian, but profits are permitted. The birth and death of firms are not explained. One notable exception is Schumpeter (183) who reversed the trend completely and visualized the entrepreneur as the prime mover and catalysor of the phenomenon of economic development. For our purpose, this entrepreneur is less important than one may think, because the entrepreneur in Schumpeter is not presented as an organic component of the firm but as a factor in a wider frame of reference: the change in a whole economic system. Yet Schumpeter's entrepreneur may be the key to the explanation of the founding of new firms (either to carry an innovation or to imitate) and to explain why old firms die, for example, they are pushed aside by newcomers' innovations. Schumpeter also links the life of the firm to the performance of its entrepreneur in a Marshallian fashion. Yet it is not a fourth factor of production and seems not essential for the firm to exist. The reason is that when the circular flow is in equilibrium the Schumpeterian entrepreneur is nonexistent, yet (183, p. 45) Schumpeter admits that firms carry on production.

SECTION V: METHODOLOGICAL EVALUATION OF THE THEORY OF THE FIRM 1870-1920.

(a) If we make the exception of Marshall and Schumpeter, the theory of the firm in this period is still a minor accessory element in the theory of prices. It is still a mental construct adorned, it must be conceded, with more practical details than in previous versions. Still, the entrepreneur is automatic; its behaviour is only remotely connected with the real world. In fact, the entrepreneur of Walras and of others is the entrepreneur of Say and Cournot. The theory of the firm is so abstract that its premises or its conclusions can hardly be tested. Its abstraction is mainly caused by the requirements of the purely competitive hypothesis which is basic in that period.

(b) Marshall

It seems in the end that the tradition, which pictures Marshall as the founder of traditional theory of the individual firm, is a myth. Marshall is far from being the synthetist we are sometimes asked to believe. Far from being the outcome of all the previous works on the elements of the theory of the firm, he neglected or imperfectly presented many contributions of his predecessors and contemporaries. For instance, the internal mechanism of profit maximization is less rigorous than in Cournot; similarly his work on laws of returns and on cost curves is far from perfect. His interest in the individual firm is much smaller than the publicity he

receives in various books on history of economic thought.

A more exact perspective should present him as an important contributor to many tools in the study of the traditional individual firm; for example, he furnished the all important time periods. But he should be presented mainly as the originator of a different "brand" of theory of the individual firm; a theory which now incorporates many elements foreign to the traditional theory of the firm, for example, the flesh and blood entrepreneur.

An essential characteristic of the Marshallian firm "eludes capture in mathematical terms. That element is the complex of notions of time, friction, change, and decay... and which is not dealt simply by drawing long period curves instead of short-period ones" (171, p. 23). Therefore, as will be seen later, many people who claim to be Marshallian are probably not at all members of his club. Marshall broke away from the Cournotian model and introduced time, organization, the entrepreneur (in flesh and blood), all things which are taboo even today in the formulation of the traditional theory of the firm.[34]

Marshall was forced to move outside the bounds of the traditional framework because he had two incompatible postulates to reconcile: (1) perfect competition and (2) increasing

---

[34]On this topic the book of Mrs. Penrose (The Theory of Growth of Firms) is straight Marshall and it is not at all revolutionary or unorthodox as its author would like us to believe.

returns. In Marshall then, the theory of the firm is no more a pure exercise in economic logic.

Model building is no longer made out of thin air. The entrepreneur is no longer automatic, but an autonomous component, reacting in its own way, and modifying the goal attainment process of the firm.

All this produces a model of the firm which in many aspects is inferior to the traditional model of Cournot. Cournot's model is more compact, better articulated, and its parts are all rigorously connected. This 'traditional model' is not systematically incorporated into the model of the representative firm. But the model of the representative firm opens on a different road which probably leads to more useful models. In fact, this new "brand" of the theory of the individual firm is the ancestor not so much of the works of Chamberlin and Mrs. Robinson, but of contemporary attempts by Cyert, March, Simon, and others, to extend the theory along what will be called in the last chapter of this thesis unorthodox lines.

The explanative power of Marshall's theory may be diversely appreciated. Surely it accounts better for real world phenomena, although it is not strictly logical. Its predictive power, through the representative firm, is probably of not much value as far as total output is concerned. But it can still probably predict the size and the typical firm

if we know what happens to the relative prices of the various factors of production and the nature of external economies. The greatest value of the Marshallian apparatus is probably to point the way at the type of framework which further developments of the theory should take. His model is finally compatible with other partial theories dealing with the firm.

CHAPTER VII

THE PERIOD 1920-1939

SECTION I: MONOPOLISTIC COMPETITION

Space prevents us from studying in detail the contributions of the theory of the firm in that period. It might suffice to say that Marshall left a legacy which puzzled the generation which followed him. In the beginning, few people realized that he had really rocked the boat of the classical theory of the firm as found in Say, Cournot, Walras, Clark. He had actually introduced two foreign elements: special markets and an autonomous entrepreneur. Yet after him the discussions took two roads: the first consisted in discussing technical details of the classical traditional theory of the firm; for example, Pigou discusses what he calls 'monopolistic competition'; but it is in fact an examination of Cournot's duopoly (110, p. 314). A follower on this track was Bowley. The other road was more Marshallian; it consisted in examining critically the implications of perfect competition.[1] Later J. M. Clark in Economics of Overhead Costs (1923) produced the first modern straightforward discussion of imperfect

---

[1] Yet none of the roads are along the main Marshallian contribution which was to introduce a flesh and blood entrepreneur. This is why we hold that the period 1920-1939 formally expanded the theory of the firm mainly along Cournotian lines.

competition. In England what provided the spark to reorient formal analysis of markets on a new road was the celebrated article by Sraffra.[2] He showed that a complete reformulation was necessary. For some years there was some haggling over the matter; for example, the controversy on empty "economic boxes", but the appearance of Chamberlin's Theory of Monopolistic Competition and Mrs. Robinson's Economics of Imperfect Competition extinguished, for a while, most of the controversies. Mrs. Robinson adopted completely Cournot's monopoly approach, while Chamberlin seems to have built both on Cournot's monopoly and on Marshall's special markets analysis.

Mrs. Robinson's and Chamberlin's models are too well known to necessitate any detail exposition. What we will do is to isolate the implications of Chamberlin's approach for the theory of the firm. One generalization often encountered is that the theories of monopolistic and imperfect competition "is another development that objectively stems from Marshall."[3]

Contrary to that, we will say that the Marshallian

---

[2]Sraffra questioned Marshall's solution to the Cournot dilemma, which consisted in stating that increasing returns are logically incompatible with perfect competition. Sraffra suggested that the monopoly approach should be used to solve this riddle. He pointed out that the explanation of real phenomenon, for example, the limit to the size of the firm should not be looked upon from the cost side but from the revenue side. In other words, we must discard the classical competitive model.

[3]Schumpeter (185, p. 1150). See also Mrs. Robinson (178, p. v) and Chamberlin (36, pp. 69-70 n) where both refer to Marshall.

character of these theories has been overemphasized, especially in the case of Chamberlin. If the Marshallian concept of "special markets" reminds us of monopolistic competition (large group) the formal approach in Chamberlin is Cournotian. Contrary to Marshall whose firms are quantity adjusters (171, p. 26) and where the modes of behaviour are unconspicuous, the firms of Chamberlin are price-fixers and they operate under definite modes of behaviour and their entrepreneur is automatic. A more true perspective is the statement that, objectively speaking, Chamberlin has systemmatically integrated differentiation in a Cournotian model.

Some people are so carried away by this view that they go to the other extreme and do not see any formal difference between Cournot and Chamberlin. They establish their position by stipulating that: (1) the analytical tools (2) the equilibrium conditions and (3) the modes of behaviour, are the same in Cournot and in Chamberlin (5, pp. 104-109). We intend to show below that although we subscribe generally to this position, we do not agree with the method of establishing it.

(1) The Analytical Tools

Mr. Archibald (5, p. 108) sums up the results of his work on a Cournot-Chamberlin group with the statement: "the slope of the particular demand curve cannot be altered by changing numbers...." We intend to show immediately that if this is true for a Cournot duopoly group, it is not true for

a large (monopolistic) group of Chamberlin.

(a) Cournot's duopoly model:

We have seen previously (chap. No. 3 above) that in Cournot the mode of behaviour of the individual firm consists in assuming the output of its rival(s) as constant. Therefore the firm establishes its own particular demand curve by subtracting that constant from the market demand curve. Furthermore Cournot (54, p. 80) requires that all the firms, in this market, simultaneously adjust their prices to the new price. This must be so because the products offered for sale by the firms are homogeneous in the eyes of the consumers; consequently no firm can make a sale at a price higher than the one of its rivals.

The particular demand curve of the firm has then the same slope as the market demand curve because a constant is subtracted from each output that could be sold at each price on the market. As more firms enter this market, the only effect is to increase the size of that constant, and the individual demand curve of the original firm shifts to the left parallel to itself.

(b) Chamberlin's Monopolistic Competition Model:

Here the mode of behaviour of the individual firm is to assume that the f.o.b. price charged by its rivals remains constant.[4] This mode of behaviour yields the dd' curve of

---

[4]Note that this mode of behaviour is different from the one used by Cournot in the case of duopoly.

Chamberlin as a particular demand curve under the hypothesis of non-reaction on the part of the rivals. In the same market morphology, Chamberlin, using another mode of behaviour, constructed another demand curve for the firm (called DD'). Here the firm expects all its price changes to be instantaneously matched by its rivals. Both modes of behaviour are compatible with Cournot in the sense that each partially satisfies Cournot's approach.

Let us build these demand curves, and verify whether or not the slopes of the DD' and the dd' curves change under new entry.

One method to introduce some precision into a Chamberlin model is to use a spatial model. As a matter of fact differentiation means that the commodity dealt with by the various firms is otherwise similar except for some features like quality, location of the sellers, etc. There is no logical difficulty[5] that would prevent the use of economic space to represent differences in quality, and location. Chamberlin himself subscribed to this approach as a way of illustrating a monopolistic competitive model.[6]

---

[5]Although there are still some practical difficulties, for example, the problem of measurement.

[6]Chamberlin (37, p. 347) says: "we may speak of economic space instead of literal space, and conceive buyers distributed symbolically throughout a multi-dimensional area containing all the various aspects of the products in addition to convenience...." Other authors, for instance Abbott (2, p. 31) are equally optimistic as to the adequacy of a spatial model as being able to permit a systematic study of

To minimize the practical difficulties, let us suppose that the products of the firms are differentiated in only one aspect: literal location.[7] The large monopolistic group of Chamberlin will then consist in:

(1) a great number of firms distributed over an area,

(2) a great number of customers also distributed over an area,

(3) the product is homogeneous except for location,

(4) each customer has an individual demand curve for the otherwise homogeneous commodity of $p = a - bx$,

(5) transport costs are one unit of cost per unit of distance.

Take any sufficiently large area and start locating firms there. If only one firm is located at A (in Fig. I),

---

the problem of quality differentiation: "Differences in consumers' scales of preferences, due to differences in the environment in which they wish to use a given product, may be represented by differences in location of the population, while differences in the qualities of the varieties offered may be represented by differences in the location of sellers. The number of sellers is assumed to be less than the number of populated locations; each variety of product therefore is "perfectly adapted" only to those who live in that precise spot. The inconvenience of obtaining a product not perfectly suited to one's purposes may be represented by the incuring of transport costs."

[7] From here on we will speak mainly in terms of literal space, except when stated otherwise.

the particular demand curve of the firm will be a parabola.[8] In other words, if the area it can serve is sufficiently large, the firm, even if alone in the field will not serve the whole area, but only the smaller area enclosed in the dotted circle (fig. 1). As more firms make their entry in this market area, they will locate outside the area served by the original firm, at least as long as the whole market area is not filled. Monopolistic competition will begin suddenly when by introducing one new firm the areas served by the firms already there have to be compressed (and the firms have to relocate to make room for the newcomer). Losch (127, pp. 109-114) has shown that the circles in this case will be transformed into hexagons (shifting the

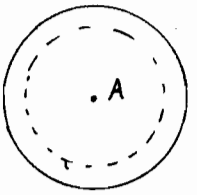
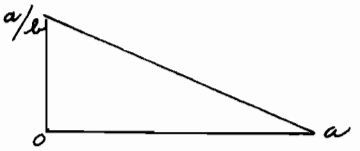


Fig. 1

---

[8] The equation of this demand curve is $x = \frac{1}{3}\pi\frac{a}{b}(a-p)^2$

It is derived as follows: Transport costs being one unit per unit of distance, the amount of sales of a firm that can be made at price zero, in a certain direction, is equal to the surface of the triangle of a height of Oa and of a base of Oa/b. The amount of sales over a circular area is obtained by rotating this triangle on its base. The amount sold at price zero over the area is equal to the volume of the cone engendered by rotating the above triangle. At $p = o$, $1/2\pi a^3/b$ is sold. Similarly at $p = 1$, $1/3\pi a/b(a-1)^2$ will be sold. The same for all other prices. Exploring all the possible prices will give the corresponding quantities sold and this schedule forms the demand curve of the firm serving this area. The equation of the demand curve is then as written above.



demand curve of each firm a bit to the left although not in a parallel way).[9] The market area is then a network of hexagons. Each firm is then in competition with at least six firms; any of its gains or losses in terms of customers are distributed equally among these six firms. If more features of differentiation are introduced, for example, quality, the space becomes multi-dimensional and the original firm comes into competition simultaneously with 6r firms, (if r = number of differentiating features). As more and more firms come into the area (other things being equal, for example, f.o.b. prices charged by all the firms are constant) the number of customers or the territory served by each firm will get smaller and smaller (see fig. 2).

We can now build the two demand curves in regard to the individual firm, DD' and dd', using the two modes of behaviour used by Chamberlin.

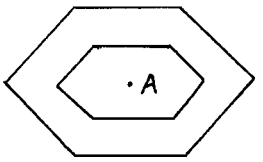


Fig. 2

(A) The DD'.

The DD' or the share of the market demand curve for the firm (it corresponds to the QC curve of Archibald (5, p. 106) is built under the mode of behaviour which specifies

---

[9] In this case the demand curve of the firm becomes: $x = \frac{\sqrt{3}}{2b}.a(a-p)^2$, where $\underline{a}$ is the length of the radius of the hexagon which is now the area served by a firm.

that the firms charge the same price. In the case of no differentiation, the share of the market demand curve shifts to the left by its base (exclusively) under new entry. This can be shown as follows:

The market demand curve is the sum of all the individual consumers' demand curves. In a spatial model with no differentiation all the consumers and all the firms are located on the same spot. Transport costs are then nil. If there are

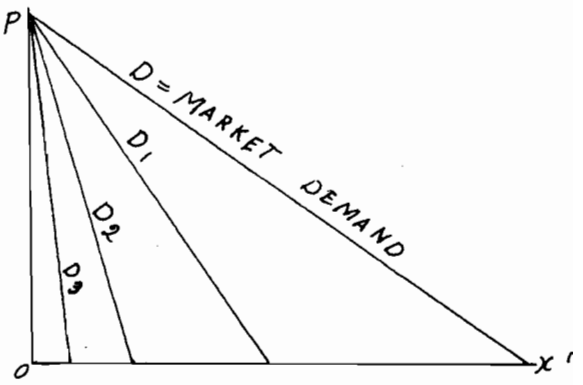


Fig. 3

$\underline{m}$ consumers, the market demand curve becomes $X = \frac{m(a-p)}{b}$ (if X represents the amount sold on the market at various prices and each customer has a particular demand $x = \frac{(a-p)}{b}$. If all the firms are located on the same spot and charge the same price, the sales of each firm at each price are 1/n times the market sales at a particular price (where n is the number of firms). The market demand curve could be written also as: $p = a - \frac{b}{m}X$. If there is only one firm in the market this equation becomes the equation of the demand curve of the individual firm. If there are n firms; the share of the market demand curve of each firm becomes:

$$p = a - \frac{bn}{m}.x'. \qquad (\text{where } x' = X/n)$$

It is easily seen that as the number of firms increases, the

slope of the share of the market demand curve of the firm increases.

In a Chamberlian model (with differentiation) there are some difficulties in fitting this apparatus.[10] The first difficulty concerns the size of the market demand curve. In the previous ponctiform model (no differentiation) the market demand curve was simply the arithmetical summation of all the consumers' demand curves, but in the case of differentiation, the amount sold over a whole market area does not depend exclusively upon the price charged by the firms but also upon the quality (variety of the locations of the suppliers) offered. In other words as more firms get into this area, many customers that were left out before (because the f.o.b. price plus transportation costs to their location from the nearest firm was greater than their maximum price a) are now buying, and those who were already served now buy more because the c.i.f. price (the only one that directly matters to them) has now decreased because some firms are now nearer to them. If now we want to translate this into an aggregate market demand curve, here is one of the many means by which to do it:

Suppose n firms in the area are selling, as a group,

---

[10]And this shows how artificial and almost impossible to the point of being a caricature is the attempt to fit into the same framework a Cournot duopoly model and the monopolistic Competition model of Chamberlin.

X_1 of the commodity at a price p_1. Without changing the price of the firms let us introduce one more firm; after readjustment in locations, if we sum up the outputs of each firm we obtain an aggregate amount sold in this market as X_2; where $X_2 > X_1$. In other words as more firms get in, the market demand curve shifts to the right.[11] However the share of the market demand curve of each firm will shift to the left as more firms make their entry; but not by as much as it would have been in a case of no differentiation.[12] The DD', then, does not shift parallel to itself. (See fig. 4).

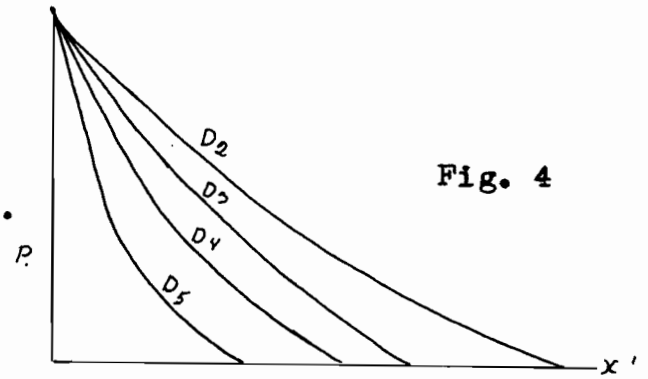


Fig. 4

(B) The dd' curve:

Let us now build the dd' curve or what Archibald calls the particular demand curve

---

[11] Among other things there is a common sense explanation which is expressed in these terms: A greater number of firms means a more intense cultivation of each customer. A Cournot undifferentiated model cannot show that, because here numbers do not matter. But a Chamberlinian model shows this truth which is not obvious in Cournot. Consequently numbers have an unsuspected influence in a differentiated model. We can apply this theorem to the following case:

It is likely that an economic system like the U.S. with a lot of variety in products (for example, various styles of automobiles), will sell more of these items (other things being equal) than an economy like Russia where the planning bureau might restrict to a few the number of models of automobiles put on the market. This is because some customers might not be tempted to trade-in their two year old cars because the style of the cars has not changed enough, while in the U.S. they might.

[12] When kn firms are introduced (thus reducing the area of the hexagons) the demand curve of each firm becomes: $x = \frac{3}{2}\sqrt{a}\left(\frac{a}{k} - p\right)^2$. This is still a parabola.

of the firm. In a Cournot model (as seen above p. 326) the firm sees its particular demand curve shift parallel to itself; this under the influence of new entry.

The difference between the DD' and the dd' is in terms of the basic Chamberlinian assumptions as to reactions on the part of rivals. In the DD' there is complete reaction, for example every price cut is matched. In the dd' there is no reaction, or retaliation. For instance in the case of Cournot, when one duopolist changes its price (and thus its output) the other duopolist does not retaliate. In Chamberlin, the non-retaliation is in terms of prices; one firm can change its price without incurring the risk of other firms retaliating by changing their price. The nature of what is at stake can be illustrated by the following small arithmetical model. This is a linear spatial model (the firms and customers are located along one road). What is to be determined is the shape of the particular demand curve of the individual firm when it can vary its price without the rivals doing so. (For simplicity we will use only one arm of the Y; see fig. 5). Each customer has a demand curve of the shape $p = a-bx$.

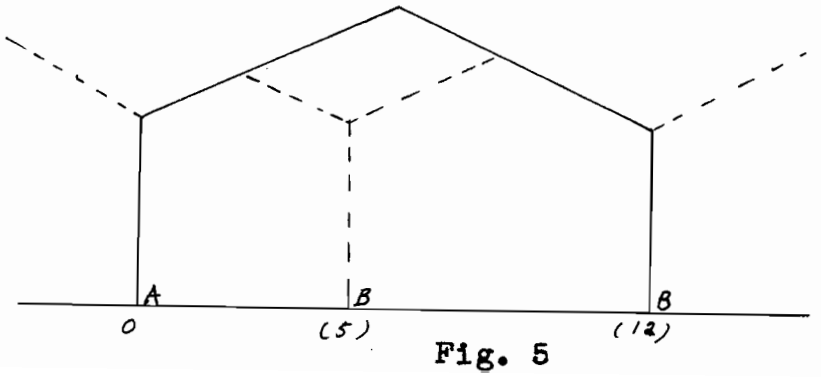


Fig. 5

In the first case firm B is located at 12 units of distance from firm A. There is a customer at each .5 unit of distance. Let us suppose also that a = 10; b = 1, and that the base price of B is constant at 4.[13] In the second case new firms come in such a way that now the next competitor B is now located at only 5 units of distance (the base price remaining the same) from A. We have then enough information to build two sets of DD' and dd' (one for location of B at 12 and one for location of B at 5). See the curves on fig. 6 page 336.

The dd' is built in Chamberlin simply by keeping the number of firms and their f.o.b. price constant, and having the firm A change its own f.o.b. price from 0 to a. By repeating the same experiment twice (in the first case the location of B is at 12 with n firms in the group, and in the second case there are $n + \Delta n$ firms in the group such that firm B now locates at 5 units of distance from A), we obtain the desired curves. Fig. 6 then illustrates what happens to dd' and DD' when new firms are introduced. When we have n firms we obtain DD' and dd'; and when we have $n + \Delta n$ firms we obtain DD'' and dd''. In both situations, (with

---

[13] In passing we might dispose of an objection of Mr. Archibald saying: "in the case of non-homogeneity the concept of an equal price change is ambiguous" (p. 104). In the above spatial model, at least as long as differentiation is restricted to location differences, equal price changes make sense if we are talking of f.o.b. prices.

K+E 10 X 10 TO THE INCH 359-5
KEUFFEL & ESSER CO. MADE IN U.S.A.

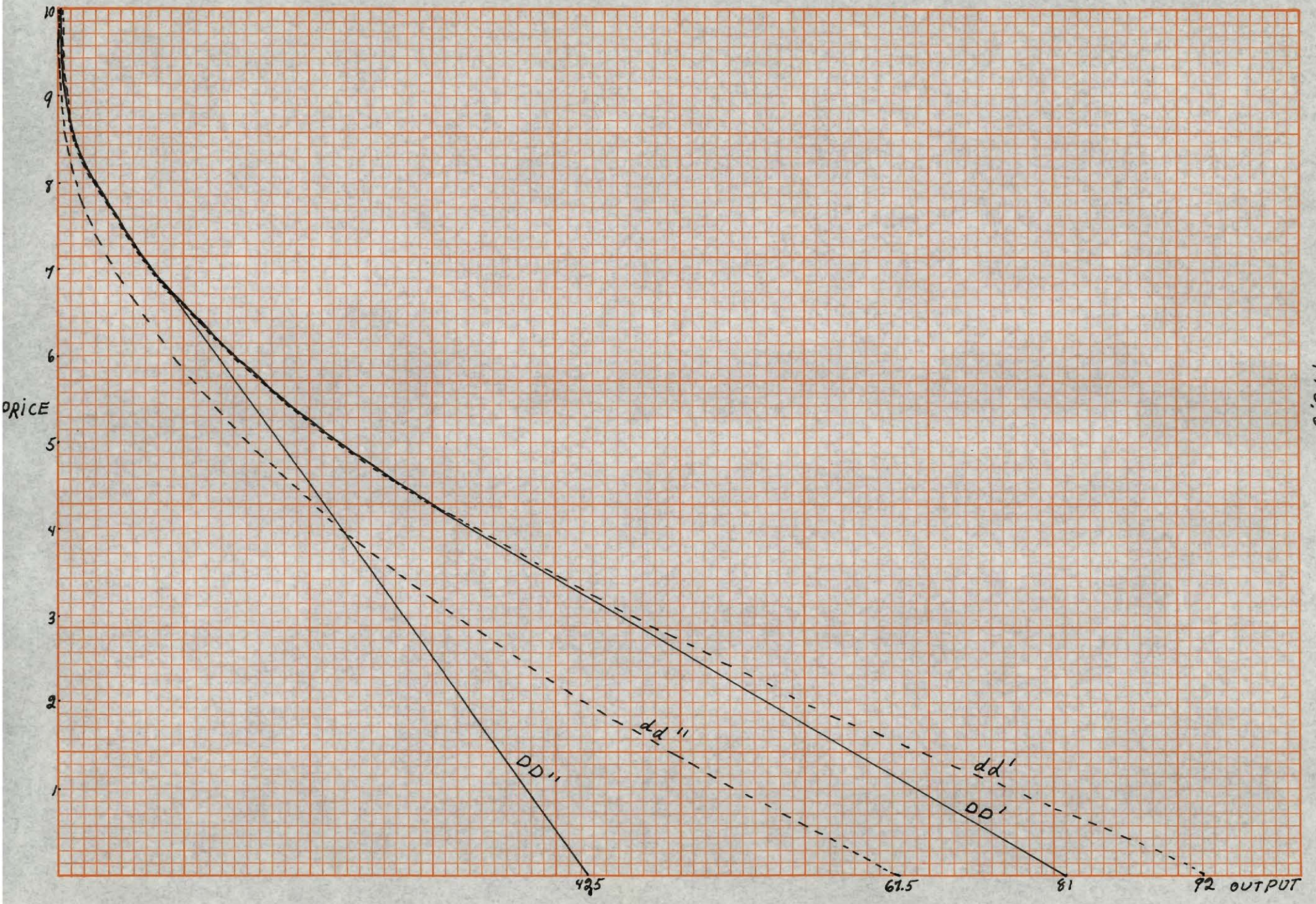


FIG. 6

respect to the dd curves) the f.o.b. price of the competitors are constant. Therefore the shift in the particular demand curves from dd' to dd'' can be attributed solely to changing the number of firms in the group.

It is clearly seen on fig. 6 that the particular demand curve (dd') does not shift parallel to itself as postulated by Archibald. Moreover we must note that the slope of the dd' curve of Chamberlin has no relation to the slope of the market demand curve. This slope is a function of the proximity of the rivals and of the rate of transport costs.

(2) Equilibrium Conditions:

Mr. Archibald adds to his previous statement by saying (5, p. 104) that "not only the analytical tools (are the same in a Cournot duopoly model and in a Chamberlin monopolistic competition model) but also the equilibrium conditions are identical."

We would like to show below that the complete equilibrium long run conditions are quite different in both:
(a) In a Cournot duopoly model, using the particular dd' curve, the individual firm will be in equilibrium, (and so will be the industry, when dd' is tangent to AC.)[14] See fig. No. 7.

---

[14]This might be a good occasion to discuss the statement of Mr. Archibald about costs: he says: (p. 105) that "non-homogeneity also makes the concept of equal costs ambiguous or even meaningless". Our model does not seem to come under this malediction because it is perfectly possible that over a certain area, the costs of producing the same basic commodity (see the example of Losch p. 105 where each farmer decides to produce beer) are identical. At least there is nothing in the basic assumptions that militates in favor of inherent differences in costs of production.

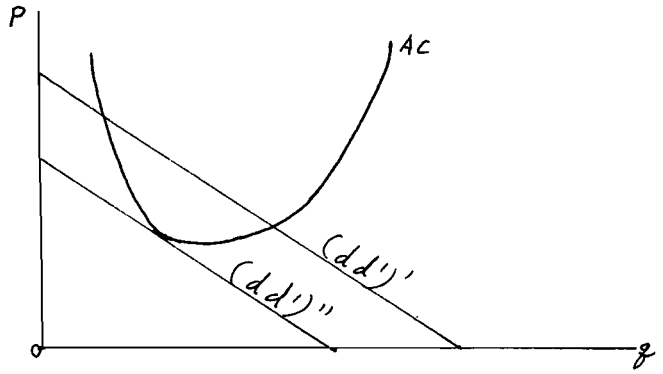


Fig. 7

If the original demand is $(dd')'$, profits are earned and new firms will come in. This will shift the $(dd')'$ parallel to itself. Entry will stop when $(dd')''$ is tangent to AC. That is, when all the firms are just breaking even. Note: we do not draw the DD' in Cournot because the DD' curve presupposes a mode of behaviour (interdependence recognized) not contemplated by Cournot.

(b) In Chamberlin, it is well known that we need two conditions: dd' must be tangent to the AC curve,

DD' must cross AC at the point where dd' is tangent to AC.

In other words, dd' being tangent to AC is not enough; see fig. 8. Suppose the ruling price in the industry is p_0; every firm is making losses at A on their DD'. Yet a particular firm, using its dd' can break even by reducing its price to p_1 (moving to B). However as the other

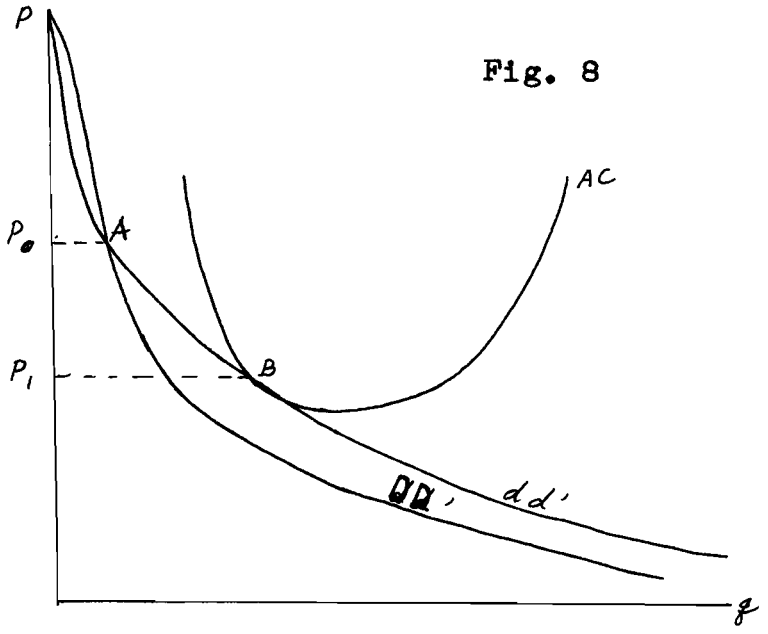
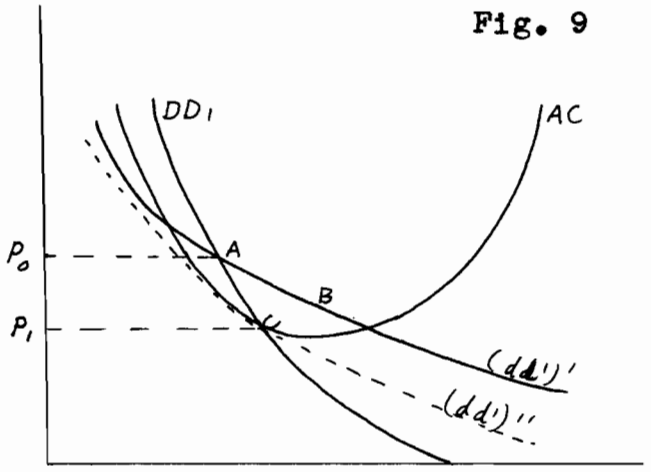


Fig. 8

firms imitate the original firm (they will do so not because they retaliate but simply because it is presumed that the firms are identical including management ability). This process will bring all the firms to point C which is not an equilibrium position for the industry. Consequently equilibrium can be achieved only if firms leave the industry; that is only if DD' is brought back to a point like B (fig. 8) or C in fig. 9: Starting from an original price p_o; all the firms are at A. However one firm by using its dd' can do better by cutting the price (moving in the direction of B). Yet as in fig. 8, competitors will imitate and this will have the effect of shifting downward the individual dd'. This process will stop when all the firms will be charging p_1. At C, no firm has any inducement to cut further even along its dd' (since dd' is tangent at that point).

Fig. 9



The above reasoning is probably enough to support our view that the equilibrium conditions are not identical in Cournot and in Chamberlin.

(3) Modes of Behaviour:

Mr. Archibald says (5, p. 104) that "there is no analytical distinction between the Cournot solution in the case

of many firms because the same beliefs about the reaction of the rival(s), and hence the same behaviour is attributed to the individual firm in each case."

We would like to show that the mode of behaviour in the analysis of Cournot is different from the one used by Chamberlin in his monopolistic competitive case:

(A) Cournot: We have seen in chapter three that Cournot used two modes of behaviour: price-fixing type A and quantity adjuster.

(B) Chamberlin's Monopolistic Competition: The large group of Chamberlin is characterized by:

(a) differentiation of products

(b) a large number of competitors.

Both characteristics are favorable to produce a polypolistic behaviour of type B, where no interdependence between firms is recognized. Numbers and differentiation are merely brought in to make plausible the eradication of any consciousness of being watched. Thus, "...there must be a large number of buyers and sellers so that the influence of any one or of several in combination is negligible." (36, p. 7). Price-fixing polypolistic type B behaviour requires that the firm: (a) take the price of the others as constant, (b) does not expect any reaction by its rival(s), (c) is not preoccupied with the effect of any change of its policy on the market price, and (d) thinks it can change its own price or 'product' to better its position. This mode of behaviour is postulated

by Chamberlin when he says: "that any adjustment of price or 'product' by a single producer spreads its influence over so many of his competitors that the impact felt by one is negligible and does not lead him to any readjustment in his own situation" (36, p. 83). It is then possible to find an "analytical" distinction between Chamberlin and Cournot.

Even if Chamberlin is different in the details from Cournot (as we have just seen) it remains that as a matter of general approach it is a continuation of Cournot. Chamberlin simply used another variety of the polypolistic mode of behaviour. However, even if the formal approach is Cournotian, many elements of Chamberlin's theory have Marshallian connotations. For instance, the special market of each firm and the emphasis on group equilibrium seem remnants of Marshallian preoccupations: the industry output, the supply price.

What is more important is the methodological evaluation of the Chamberlian theory for the theory of the firm. The Chamberlinian theory, more clearly than anywhere else before, definitely begins with the individual firm. This might be interpreted as a forward step towards more "realistic" models. Yet the Chamberlin's theory[15] is not more realistic than others, such as Marshall. The truth is probably

---

[15]We restrict our views on Chamberlin to his main contribution: the monopolistic competition (large group).

the opposite. Chamberlin's model accounts much less for real world phenomena than Marshall's model. The notions of special markets, such as industry, are systematized and more clearly defined in Chamberlin than in Marshall. But this increase in theoretical rigor is at the cost of less contact with reality. The Marshallian firm is much more "flesh and blood" than the Chamberlinian firm. The main reason is that the Chamberlinian firm does not integrate the autonomous entrepreneur (who is born and dies) of Marshall. Chamberlin's entrepreneur is automatic and its presence does not modify in any way the process of attaining the goal of the firm which remains profit maximization (as in Cournot).[16]

Chamberlin's large group has never been found to account, even theoretically, for real market situations. Most people who embarked on testing this possibility usually came out with another market morphology: chain oligopoly. Chamberlin's model is therefore at the same level of abstraction as Cournot's model. It is a logical construction out of premises, i.e. a given market morphology coupled with a given mode of behaviour.[17] If this method is followed, there is a

---

[16] We are quite aware that Chamberlin has spent a lot of energy to "differentiate" his product, for example, that it is reconciliable with full-cost, etc. Yet these attempts seem to be extensions of the original model rather than integral features of it. In the last chapter these orthodox extensions will be dealt with.

[17] While Chamberlin was constructing one model, Zeuthen, Stakelberg, and others achieved different results by utilizing different combinations of market structures and modes of behaviour.

great number of theoretical models that can be constructed, but nothing guarantees their relevance. The predictive power of this theory is then problematic. In terms of testing, as it stood in 1933, its possibilities are small because some of the key components, for example differences in quality, were not directly measurable;[18] the theory does not enable us to eliminate some vicious circles such as the one where quality is judged by price. Also there were no successful attempts in establishing the welfare implications of selling costs, product variation, and excess capacity. Summing up, the firm of Chamberlin is a mechanistic model; it is mainly a cost and revenue center which operates in the manner of a computer. The theory of the firm is therefore a portion of a wider price theory based on quite rudimentary psychology, without any feedback from the world it proposes to explain.

SECTION II: THE STATE OF THE TRADITIONAL THEORY OF THE FIRM

By 1939 then, for most people the theory of the firm is a portion of a sub-theory of the general theory of prices. It is no more than a specialized view of the theory of value.[19]

---

[18] In the last chapter we will give a short account of the tentatives at remedying this lacunae.

[19] "and its vitality is derived almost exclusively from its connection with this highly developed and still basically unchallenged general system of the economic analysis of the problems of price determination and resource allocation." Penrose (160, p. 11).

In the words of one of its main contemporary exponents, "what economists usually call the theory of the firm is a part of the theory of relative prices, inputs and outputs by business units under various competitive and monopolistic conditions" (131, p. 27). As far as economic theory is concerned it "permits the extension of supply and demand from the artificial pure exchange case to situations involving production" (78, p. 560).

Value theory is initially expressed in a model involving the supply and demand functions of a commodity. Marshalling the data in two separate kinds of forces is the usual method of procedure of economics.[20] Formulating the problem into two types of forces does not throw much light on the problem except if the economist is able to say a lot about what is underlying Demand and Supply. In other words the economist must 'explain' demand and supply. Since the price of a commodity is determined on the market, the work of the economist amounts to determining the market demand and the market supply. Therefore, the traditional theory of the firm requires two sub-models: (a) Demand, and (b) Supply.

(a) The Demand Model

In Cournot, and its followers, for example, Chamberlin,

---

[20]This method is encountered outside value theory, for example in monetary theory, employment theory.

market demand curves are established. In Cournot there is no theory as such, because demand is not explained, neither is its slope justified. As seen previously Cournot satisfied himself with a common sense basis for the law of demand, and curiously (see below), economists should have imitated him. The efforts of his followers to 'improve' this theory of demand are rated today as practically worthless. The approaches to establish a theory of demand have been numerous. Among the notable ones, one must mention the utility theory pioneered by Marshall, the indifference curve approach of Slutsky and the Hicks-Allen team, and more recently the revealed preference approach of Samuelson. All this is for methodological reasons[21] of dubious theoretical value.

This is not to say that economists should not accept the downward market demand curve; it is quite the opposite, but now they must do so for common sense reasons, not for theoretical reasons. Actually one way to establish the market demand curve for a product is to rely on pragmatic methods such as econometrics.[22] In other words, as far as demand

---

[21]See section I of Chapter VIII below where a detailed study of the methodological worthiness of the theory of demand is made pp. 363 ff.

[22]It would be too strong a statement to say (for the purpose of establishing the market demand curve) that econometrics does not rely on any theoretical reasoning put forward by the traditional economic theory. Actually two courses of action are available to the econometrician. He can (as in the case of macro-economic theories) proceed independently, that is, make statistical inquiries and try to isolate (through

theory is concerned, one can start in Cournot and jump straight to today without missing much.[23]

If the economists cannot establish theoretically the market demand curve this does not help them to establish the individual demand curve of the individual firm. Once we leave the pure and perfect competitive case we come into a no-man's land called monopolistic competition and oligopoly. In this case the individual firm's demand curve cannot theoretically be deduced from the market demand curve, even if it could have been established. In the words of Machlup "there is no such thing as an oligopoly demand curve..." (131, p. 355).

The traditional theory of the firm is then off to a bad start: it cannot establish theoretically that the market demand curve slopes downward; it cannot deduce the firm's demand curve from the market demand curve.[24]

---

the use of correlations) the important parameters that influence demand. (See Spencer and Seligman, (200, p. 136)). Or he can borrow from the traditional theory the demand determinants like price, income etc., and hereafter proceed on his own within the general framework of the traditional theory. However his models will not usually be tied down to a particular brand of traditional demand theory. This is why even if our micro-demand theories have been proved right for the wrong reasons, they are still useful to the econometrician because he did not use a detailed formulation of any of them.

[23]He would not even miss the concept of elasticity which is in Cournot.

[24]If we expept the case of pure and perfect competition which is not likely to be found very often.

(b) The Supply Model

Was the theory more successful in establishing a market supply curve? The theory works at the industry or market level. The purpose is to eliminate the vagaries of individual data in the case of particular firms.

The process of building a theory of market supply goes as follows: The market supply function is postulated as $S = f(p)$. Quantity supplied is a function of the price. The hypothesis is that if p increases, S will increase.

There are many implications in this theory. It requires that on the average firms increase output when price is increased. Therefore, there is another postulate implied: each firm (or most of them) increases output accordingly. In effect this is predicting the output of the individual unit and as such difficult to reconcile with the statement that theory is not interested in the individual firm (131, p. 33). The second implied postulate is that firms maximize profits, for example, that firms on the average determine their output by comparing changes in revenues with changes in costs.

Costs being brought in, the economist needs another sub-theory: the theory of costs, a theory which has to describe and predict how costs vary with changes in output. Now if the theory insists in not wanting to predict the output of a single firm, care must be taken to insure that the industry remains competitive. This can be obtained, as

pointed out by Cournot, only if costs are eventually increasing. Marshall, the unorthodox, kept both decreasing returns and the competitive industry by introducing a *deus ex machina* which has been rejected by his followers.[25]

Up to now the theory of supply rests on two basic postulates about the behaviour of the individual firm:

(a) The output of firms is increased until MC = MR.

(b) Costs must eventually increase.

The first postulate implies:

1) that on the average firms are rational enough to maximize profits,

2) if some firms do not follow the rule, this idiosyncracy must be overcompensated by others. Yet there is nothing mentioned in the theory that this will surely happen.

3) It requires that the firms 'know' their MC and MR or at least that they act as if they know about these data.

This requires some comments.[26]

Firstly, in the traditional theory the postulate of

---

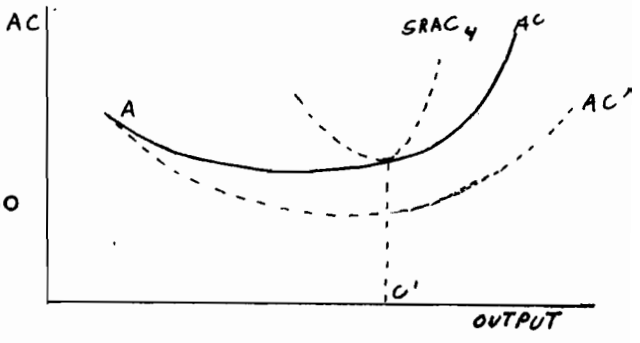
[25]The *deus ex machina* is from the real world and consists in the supposition that firms take time to expand; that also they are born and die as their entrepreneurs do. If these conditions prevail, the firms may not have the time to expand enough as to become monopolists.

[26]The purpose of these remarks is simply to point out obvious implications of the postulates of the traditional theory. At this moment no exhaustive criticisms of the traditional theory is attempted.

rationality is identical with profit maximization. This postulate is justified by the axiom that managers prefer more profits to less profits. Yet as in many axiomatic systems, this does not lead very far. For instance, the axiomatic formulation was not successful in the field of demand theory. Actually, managers will prefer more profits to less profits only if this procedure does not entail any 'cost', subjective or else, like loss of control of the firm, censure from governments, or the public. The hypothesis of rationality remains then a postulate which is bound to suffer many an exception. Postulating profit maximization does not in any way make entrepreneurs more rational, and any theory based on this is a case of implicit theorizing.

Secondly, the postulate of perfect knowledge is even more troublesome and there is the perennial difficulty about measurement. It might be remembered that various theories of demand were dismissed precisely because of the impracticability of their measurement units, for example, the utility theory. Similarly in the theory of the firm, it might be difficult for managers to know their MC and MR. Putting the reasoning in another form, MC and MR may be known by the managers only through a process of learning. A process of learning means 'tâtonnements', oscillations (with the possibility of instability), shifts of the curves. Limiting ourselves to the cost side, we might realize that it might be impossible, at any point of time, for the managers, to draw

their long run cost curve. This cost curve to be really of any service to the individual firm must include all sorts of economies and diseconomies, for example, external which are reaped only if the industry grows, and internal (in the form of minor inventions à la Adam Smith) and which are discovered only through the process of growing by the individual firm. This is why the straight forward analysis picturing the firm as choosing a single equilibrium position may not account for the situation at all. Many firms are never confronted (or never figure) with a multiple choice of different positions. In other words, the location of the equilibrium position itself is influenced by the path for reaching equilibrium. This situation can be illustrated as follows. Because of inherited capital structure, organization structure, knowledge, and a certain capacity to learn, a firm is at a position A on what it presumes is its long run cost curve AC. Suppose it decides to reach an output C', where size No. 4 seems the lowest cost one. However by the very fact that it will grow to attain this size No. 4, it may find itself moving along AC'.



This had been seen by Marshall when he noticed what he called the irreversibility of supply. This phenomenon makes any individual planning very awkward.

Furthermore one must note that the AC curve is built out of a production function supposedly 'given' to the economist. It has been already been pointed out[27] that very often the production function does not exist in a form directly usable by the economist. The isoquant might first have to be discovered through a linear programming approach; the whole production map may remain unknown to the manager. The only way to discover this function may be to grow. This type of difficulty is not easily dealt with in the traditional theory which sees equilibrium as a choice among many known possibilities.

Connected with ignorance is the problem of uncertainty. This element adds to ignorance. Traditionally this difficulty has been taken care of by forms of discounting the variables in question. Yet when the possibilities expected by the firm present themselves in the form of probability distributions of outcomes, cost curves and revenue curves are impossible to build.

The traditional theory of the firm is not bothered too much by these difficulties, because they are acute or difficult to solve at the individual level, and the traditional theory operates only at the industry level. Presumably, at the industry level, a consensus about the key variables develops. An average performance becomes expected with

---

[27] See above in Chapter V, p. 184 ff.

almost complete certainty, so that now the firms have stable curves. Exactly how this is achieved is not said. As in the case of demand theory, it seems that the rising supply curve is obtained not through theory but through the intuition of the theorists.

All this results in a curious theory of the firm. "The model of the individual firm (thus constructed) does not bear any close resemblance to anything that we in the real world know as a firm."[28] In fact, "it is not concerned with any 'particular' firm in the real world."[29] The firm is then something like an average or a representative model, and may be no more than a set of statistics. Its statistical substance makes it methodologically similar to the principle of insurance. The insurance principle predicts an average behaviour for the individual units. The assessment of this behaviour is inaccurate if compared with the actual behaviour of any individual units, but it will be right if the aggregate is considered. The difference between the operations of insurance companies and the operations of the traditional theory of the firm is that the former uses empirical data

---

[28]Machlup: (131, p. 8). Contrasts this with the definition of the firm by Joan Robinson (178, p. 17): "A firm is a concern very similar to the firms of the real world, but which produces only one commodity, and is controlled by a single independent interest. The controlling interest of the firm is an entrepreneur."

[29]Machlup: (131, p. 33). He continues p. 49; "The purpose of the analysis of the firm is not to explain all actions of each and every firm in existence; we are satisfied if we can explain strong tendencies in industry or in the whole economy."

while the theory of the firm does not do so, but uses postulates of modes of behaviour (rationality, knowledge) which are entirely the fabrication of the economist. The models of the firm are then shockingly unrealistic (131, p. 7). Such a model of the firm "does not attempt to give all the reasons why a given firm makes the type or quality of product it makes; why it produces the output it produces; why it employs the workers it employs; or why it charges the prices it charges" (131, p. 33). In other words, "it is not the purpose of the theory of the firm to predict the prices or outputs a particular firm would decide upon in an objectively described situation (131, pp. 54-55). Confronted with this refusal to deal with the individual firm, it is difficult to see how the traditional theory can justify its pretention of predicting the behaviour of the aggregate variables, i.e. market price and output. Especially since the statistical techniques (insurance principle) is not available to it. It seems that the necessary data are lacking, especially for short-run predictions. Statistical methods being unavailable, it could proceed by aggregating the particular modes of behaviour. Yet it has systematically refused this approach. Its pretention is then based on a postulate which has not been verified empirically.

Not only is the traditional theory of the firm likely to fail in its own goal, i.e. predicting the market price,

but its refusal to deal with the individual firm renders it completely unfit to explain other aspects of the firm such as its growth, decision-making process. Neither can it explain why and how firms are born and disappear.

The traditional theory of the firm, when stripped of all extraneous material, boils down to a theory of profit maximization under various market morphologies, themselves introduced in the model, with their corresponding modes of behaviour, on an axiom basis. As such it is liable to be labelled a "gigantic redundant tautology".[30] The theory of the firm is then akin to a Geometry, where every proposition is deduced from a basic postulate. Such a geometry (there are many geometries, for example, Euclid, Rieman, etc.) has no predictive power, except if a real world situation happens to fit exactly every component of a theorem, no more, no less. The traditional theory of the firm has then no predictive power except if it can be proved that the real world situation corresponds exactly to its characteristics, and that the elements of real world situations not accounted for do not matter.

The traditional theory of the firm can be characterized

---

[30]Schoeffler, S.: (189, p. 84). "The proposition, for example, that the firm will equalize the ratios of marginal cost of all the factors of production, tells us nothing that we do not already know from the proposition that 'the firm will maximize profits'."

as follows:

(1) it is not individualized,

(2) it does not integrate the organization factor,

(3) it has no entrepreneur, except an automatic one,

(4) time processes are not taken into consideration. Perfect knowledge is assumed; some types of uncertainty cannot be dealt with.

(5) It is more at ease with a single product,

(6) Its operational framework is marginal analysis and profit maximization.

These, may or may not be defects depending on which school of methodology one belongs. However, in terms of methodology, the following points are usually made:

(1) Friedman and other tenants of the traditional theory, though repeatedly promising a list of successful predictions never came forth with one (134, p. 179).

(2) The premises of the theory are not drawn from reality although some formulations are nearer than others, for example, monopolistic competition. However it has never been shown, for instance, that the large group of Chamberlin corresponds (even in theory) to a real market structure.

(3) Many traditional theories are logically inconsistent, for example, the large group of Chamberlin.

(4) Only a portion of the necessary data to build traditional models is available.

(5) Its predictions (at least those implied in its models) often contradict those of managerial economics.

Therefore the traditional theory has been criticized as follows:

(1) The theory, because of its characteristics listed above, for example, lack of an organization factor, does not properly describe the decision-making process within industrial firms. As such, it is acceptable. Economics is not managerialism. The "traditional theory was concerned with market behaviour (i.e. behaviour with respect to prices and quantities), not with decision-making procedures" (22, p. 168). However this may be a way to evade the issue. Undoubtedly the critics hold that to predict successfully the price or output, one must know about internal processes, and integrate them into the analysis. For instance, traditional theory could show that the integration of the administrative factor is not necessary if it could be established that the corporations do in fact maximize profits.

(2) The traditional theory makes incorrect and naive assumptions about the behaviour of firms. For instance, one of its a priori postulates is to stipulate that marginal analysis is used as a tool. Traditional methodologists, for example, Friedman, hold that 'realistic' postulates are not necessary for a theory, as long as it makes good predictions.

Another naive set of behaviour assumptions is the collection of modes of behaviour utilized under various oligopoly situations. It is not only that the modes themselves may be not comprehensive enough, but we have no rule to determine which one to use in a particular circumstance. As has been shown in the chapter on Cournot, a given morphology of market is compatible with many different modes of behaviour. Since the solution to an oligopoly problem depends on both the market morphology and the mode of behaviour, it is hard to find out how the 'traditional theory' can be of any value here. It is, to say the least, incomplete.

(3) The automatic 'entrepreneur' allows for only one type of motivation and behaviour: profit maximization. While it is notorious that many exist.

(4) Being so removed from 'reality' it cannot have any predictive value.

In general "deductive systems, divorced from empirical investigations, are not likely to lead to conclusions that are both valid and specific" (77, p. 48).

All this is summarized in the various epithets that have been used to qualify the traditional theory: tautology, deductive system, a priorism, and trivial solutions because:

"(a) it does not integrate into the main body of theory an explanation of the non-profit

maximizing goals,

(b) it does not integrate...the business rules observed in practice which arise in part because of uncertainty" (133, p. 189).

CHAPTER VIII

THE CONTEMPORARY THEORY OF THE FIRM

SECTION I: METHODOLOGY

To evaluate the merits of the criticism lodged against the traditional theory, we must equip ourselves with a methodological apparatus.

Up to this point in the inquiry, the theory of the firm has been explored through the general approach of economic science. Therefore, we have been building what can be called an economic theory of the firm. As such, the theory of the firm must simultaneously satisfy the requirements of a theory in general and of an economic theory in particular.

Theories are normative or positive. The normative ones determine "what should be". The positive ones describe "what is" (past, present or future). Traditionally, the type of economic theory with which we have been preoccupied above in the previous chapters has been positive economics. As in all theories, the hallmark of a good theory of the firm would be that it satisfies these criteria:

(1) It must be successful in making predictions in terms of prices and output or of other matters with which the theory of value is concerned with.

(2) Its postulates or premises, even if they are abstractions and consequently unreal, must be

directly and easily reconcilable with reality, actual patterns of behaviour.

(3) It must be logical, i.e. internal self-consistency is required.

(4) It should be in such a form as to be easily testable. That is:

(a) its concepts should correspond or be easily reconciled with those used by statisticians and accountants.

(b) facts must be available to test the predictions of the theory.

(5) The preciseness of its predictions or of the required computations must be commensurate with the problem at hand. This will directly influence the methods and behaviour of the firms in achieving their goal(s).

(6) If it is a partial theory, it must not contradict other successful partial theories which explain phenomena of the same class of those tentatively explained by the theory in question.

These various criteria can be justified as follows:

(1) Predictive Power:

This point does not need much elaboration. Theories that do not predict are sterile tautologies (189, pp. 82-83). This led some methodologists, for instance Friedman, to

stipulate that "a theory is to be judged by its predictive power" (86, p. 8). These people are brought to recognize only one criterion to justify the validity of a theory: "the proof of the pudding is in the eating". That is, only factual verification can uphold or disprove a theory (86, pp. 8-9). Some people carried away by this requirement are of the opinion that realism in the premises or postulates (i.e. market morphologies, modes of behaviour, goal setting and fulfilling,) is then not essentially required. Therefore, for Friedman et al, showing the lack of realism of the premises is not an argument against the theory.

(2) Reconcilability of the Premises with Reality:

Lack of realism in the premises as tolerated by Friedman (see above) is challenged on the basis that before accepting this view that obvious discrepancies between the basic postulates used in the model and directly observed behaviour (included in the basic postulates) do not matter, we have to prove that these discrepancies do not matter (119, p. 140). This criticism has never been directly answered.[1] What actually is objected to is not lack of realism in the premises per se, because we must agree with Friedman that being simplifications, "from the mass of complex and detailed circumstances

---

[1]On the contrary it seemed to have been embraced by some followers of Friedman. See Bodenhorn, (22, p. 165): "Our confidence in a theory should be based partly upon the number of opportunities we have had to refute it, but we should also consider the reasonableness of the assumptions."

surrounding the phenomena to be explained" (86, p. 14) the premises are forcefully unreal. We should not then dismiss a theory simply on the ground that its postulates are unreal or cannot be immediately tested against reality. What is objected to is a degree of abstraction such that reconciliation of the premises is impossible. For example:

Suppose we are planning a new model of an automobile. We may want beforehand to determine whether a certain motor will be satisfactory. Obviously, to do that, we do not need the whole body of the car including all the details from the cigarettes lighter to the door knobs. We may simplify the body of the car to a chassis and four wheels. This will be enough to determine all the necessary performances. However this chassis (which fills here the role of a premise in a theory) must anytime be proved to account theoretically and practically for the whole body as far as performances (not appearance) are concerned. Not any kind of body of a car would fit. For instance, using the body of a truck or tractor would not faithfully account for the characteristics of a car. Refraining from requiring reconcilability of the premises would amount to say that any chassis would be satisfactory.

Finally there is a law of logic which says that from false premises there is no assurance that a good conclusion will ensue. For our purposes "false" premises would be postulates not accounting faithfully for a real situation,

or postulates so remote from reality that there is a great danger of not finding our way back to the real things for which they are supposed to account. This point can be illustrated by what has happened in the theory of consumption.

The problems in the theory of consumption are similar to those encountered in the theory of the firm and this illustration will serve as a basis for examining the methodological requirements of a theory of the firm.

As in the field of the firm, we do not have as yet a general theory of consumption. Consumption has been studied at three levels: individual, market, and nation. Three types of demand must be determined: the individual demand, the market demand, and the aggregate demand. Successively three theories have been tried, at least to establish the first two types of demand curves. These theories are: utility theory, indifference curve, and revealed preference.

So far, economic theories have been unsuccessful in establishing the demand curve of a particular consumer. This may not be as deadly as it looks if economic analysis is concerned exclusively with predicting the market price of a commodity. Therefore, one must examine the performance of the theories with market demand before pronouncing a judgment.

Sometimes economic theory felt quite confident at this level (i.e. market level). Presumably its inability to predict the individual consumer behaviour was due to the presence of idiosyncracies which cancel out at the market level. Yet,

it has been impossible to 'prove' theoretically once for all that the market demand curve slopes downward.[2] The reasons are mainly because of the treacherous income effects at the individual level and because at the market level ceteris paribus conditions are not tenable (146, p. 10).

The market demand curve is one important example where people can be right for the wrong reasons, and this supports our contention against Friedman in rejecting success in predictions as the sole test of the adequacy of a theory.[3] In other words, all these theories (utility, indifference curve, revealed preference) predict a downward sloping demand curve, but this does not preclude the fact that they are all "false" in the sense that they are redundant tautologies. This situation is mainly imputable to the type of postulate inputs fed into the theories.

No theory of the consumer uses postulates (for example, modes of behaviour of the consumer) that would meet our criterion No. 2. On the contrary, the postulates use a consumer's behaviour which is a 100 per cent of the economist's fabrication. Essentially this fabricated mode of behaviour boils

---

[2]"this supposition (that the market demand curve slopes downward) in no way derives from our theorizing but from intuition and observation." Mishan, (146, p. 10).

[3]We have met a similar situation in the Cournot theory of duopoly. In Cournot the duopolists were finally right (in equilibrium) for the wrong reason; they still had (at the equilibrium point) false expectations about the behaviour of the rival.

down to attributing to the consumer a form of rational behaviour. This is so because all the theories mentioned above are based on postulates of: maximization, convexity, transitivity, and consistent choice (for revealed preference). In other words, logical consistency is postulated. But this does not "prove" that the real world actually maximizes or submits to the laws of transitivity, and convexity. Postulating logic does not make people more logical. The present theory of consumption in economic analysis is, then, a branch of logic, and therefore so are our theories of value and of the firm. Logical consistency does not guarantee relevance. Some people have suggested that if this mode of postulating maximization, transitivity etc. would have been adopted by other disciplines; for example, in psychiatry, the practice of psychiatrists would be restricted to normal people who obviously do not need them.

The present theory of the consumer is "wrong" mainly because it does not start from premises having been established through some kind of empirical research.

About testability, consumer theories do not score any better. When the theories are reduced to a testable form, the researcher usually ends up with results such as: "since the consumers like it, they buy it" or "a sign that they liked it was that they bought it".

Similarly, as a postscript one may deal briefly with the limited success encountered by economic theory in establishing the aggregate demand for consumption goods. Probably

here, as elsewhere, a good hunch (as the one made by Keynes) is as far as theory can go. Because once people started theorizing about the consumption function, chaos set in. After Keynes, people began to worry about such an inadequately expressed concept; today we are embarrassed by too many theories: "About the only thing which is now clear is the conclusion that we have no such precise, stable mathematically constant relation as many tried to derive out of Keynes" (121, p. 294). If some theories have not yet been knocked off the shelves, such as the permanent-income hypothesis of Friedman, it is because they are not measurable or testable (121, p. 323).

It must be noted that the present theories of consumption have not been crudely dismissed because their premises simply lacked realism, but because they are "<u>too</u>" unreal. Most of them are pure fantasy if the norms of current psychology are used to appreciate their value. Abstract postulates about the behaviour of the consumer which would be furnished by empirical studies might yield better theories.

(3) <u>Internal Self-Consistency</u>:

Although highly praised by the logicians, this item is curiously relegated to a subsidiary role by Friedman. Its "function is to assure that the hypothesis says what it is intended to say and does so alike for all users" (86, p. 10). Actually, a more middle-ground approach would put this criterion on a par with others. The advantages of using that rule lies in the economy of testing efforts. If

a theory is already 'wrong' in its conception, it is unlikely to work well; or if it works it will be for the wrong reasons. Tests of logical self-consistency play the role of wind tunnels in the elaboration of sound theories. An example of a contemporary theory of the firm which is internally inconsistent is the "large group" of Chamberlin. When this theory is translated into a spatial model (a method of representing differentiation), it is found that there is no such thing as a "large group" where everybody is equally equidistant from everybody else.[4] When a newcomer is introduced into a group, he cannot be introduced into the model at equal expense from everybody but only at the expense of those "near" the particular location of this newcomer. When this inconsistency is removed, only the "small group" remains in Chamberlin. Chain oligopoly is not a large group but a series of interconnected small groups.

(4) Testability of a Theory:

Generally speaking testability is an absolute requirement. There is no logical way to induce reasonable people to subscribe to theories which have never been, cannot be, or will never be tested. This does not mean that artificial experiments are easily set up. Social sciences, with many other sciences, such as astronomy, cannot usually set up

---

[4]See my M. A. thesis: Economics of Retailing, pp. 152-155.

experiments. But as in astronomy, theories can be checked against events that already happened. Therefore there are some possibilities of experimentation in economics. The requirements for testability are that:

(a) The concepts of the theory correspond to those devised by statisticians and accountants; or if not, its concepts should be easily translated into those used by these people.

(b) Facts must be available to uphold or disprove "predictions".

Much of the lamentations about the difficulty in testing in economics come from the fact that many economic theories are half thought out and are consequently "untestable". Or as mentioned by Friedman, facing this difficulty, some people retreated "into purely formal or tautological analysis" (86, p. 11) This is not to say that sometimes tautologies are not useful, but they become useful only as long as there is a possibility of breaking their circular reasoning. That is, the theory must, at least in one of its forms or extensions, use variables that can be taken out of the real world. Theories that will be classified as successful must then show that they have been "verified". This is what the traditional theory of the firm has never done in spite of many affirmations

that it could do so if it wished.[5]

The preceding statement must be understood as meaning that the traditional theory has not demonstrated its ability in predicting prices and outputs in specific circumstances. For instance, the theory cannot predict the profits of a particular firm for 1962. On the other hand, the theory has, up to now, made predictions in terms of tendencies, for example, monopoly prices tend to be higher than competitive prices, monopolistic competition leads to excess capacity. Yet these predictions have been labeled (not always rightly) as "tautological predictions" because they suppose the "product" unchanged by changing the morphology of the market. For instance, reducing a monopolistic competitive model to a purely competitive model would amount to standardizing both the product and the tastes of the consumers.

(5) Preciseness in the Data:

An economic theory being a theory of an aspect of human behaviour, does not always need to be numerically precise to the tenth decimal in its computations. Economics being a social science might (sometimes) simply predict

---

[5]See for instance Margolis who says: "Bodenhorn never states that the traditional theory of the firm has proved its usefulness by successful predictions. When he discusses this crucial point in his defense of the traditional model of the firm instead of evidence he cites Friedman's excuses for not providing the evidence" (134, p. 178). For Friedman repeated promises of verification of the theory see (86, pp. 22-23).

tendencies. For instance, in a given problem, such as the effect of a rise in the cost of factors of production, it might be sufficient to predict that there will be a rise in the price of the product. At some other times, yet, both the direction and the extent of the change might be needed.

Some years ago, there was a long controversy (Lester vs Machlup) about the "numerical definiteness of the estimates" (131, p. 42 ff.) and in general about the plausibility of marginal analysis as a 'realistic' mode of computation that could be postulated in the theory of the firm.

Cases where exact numerical solutions are not important has been illustrated by the "overtaking theory" of Machlup (131, pp. 44-47). It is notorious that many, although not all, important economic decisions are taken under rough approximations. But this should not carry us too far in the other direction, to which Machlup seems to be drawn. The firm can attain its goal(s) under three modes of computations:

(1) maximizing, (2) optimizing, (3) "satisficing".

(1) Maximizing

Maximizing a situation implies the use of calculus and, as such, marginal analysis. On this ground lumpiness, indivisibilities (if they are relatively important) may prevent the formulation of the problem in calculus terms. However, if, on the other hand, this factor of indivisibility is not relatively important, even if it exists (i.e. we cannot

hire three-quarters of a laborer) as long as management tries to formulate the problem in as precise a form as possible, this is usually sufficient in economics to describe the situation as one where maximization is practiced. Besides that, this maximizing mode of thinking involves some computations which, sometimes, are very difficult to make (excluding the factors of indivisibility and lumpiness this time). One of these difficulties is the isolation of the marginal product of an individual worker. Besides, in many situations, the management must resort to methods of computation that actually put the whole process into a new framework. For instance, in very short-run situations, where marginal analysis breaks down, such methods as linear programming must be used. This leads us to our second method of computing.

(2) Optimizing

Optimizing is a more general term than maximizing. It means that the "best" solution has been chosen with regard to a proposed criterion, (18, pp. 3-10). Optimization is encountered in two ways: full optimization and sub-optimization (103, p. 179).

(a) Full optimization. It requires that all the alternatives open to the firm are considered, that all outside events that may have an impact on the firm are considered, that the utility function of the firm is maximized.[6] In this

---

[6](103, p. 170). Full optimization considers operations of the firm simultaneously at all levels.

case optimizing is compatible with profit maximization, (if the goal of the firm has been set in this form) or with other goals, (if the entrepreneur maximizes under constraints). Full optimization is then compatible with marginal analysis. However, even here, this is true only under certain conditions. Once the data present themselves in the form of probability distributions, the maximizing methods of attaining goals give way to optimizing methods such as game theory which could hardly be reconciled with marginal analysis.

(b) Sub-optimization. When firms are complex organizations, the "over-all company management may find it difficult or impossible to establish 'explicitly' the relevance each problem has to the over-all objectives of the organization, to consider all alternatives, and to consider all influences and all side effects. That is, rather than truly optimizing from a company standpoint, the department must sub-optimize. It must optimize from its own level" (28, p. 25). That is, each level is provided with a criterion, and each level optimizes in view of this criterion; for example, the shipping department may try to minimize costs. Breaking up the general problem of the firm in this way, for instance by introducing lower levels of optimization, is not without danger; but rules may be devised to safeguard as much as possible the attainment of the over-all objectives (103, pp. 179-186).

However, once the problem is to optimize at lower

levels, this changes the type of data required and available, and consequently the modes of computation; thus the method of achieving subsidiary goals. It is likely that most departments must reason with more or less fixed amounts of certain resources. In other cases the department, such as the stock room, may have some data available only in the form of probability distributions (i.e. expected sales). In all these cases the sub-optimizing routines are non-maximizing routines (non-marginal) for example, queuing theory. The bearing of this on the type of data available is as follows.

Economic theory can be reshaped in sub-optimizing routines which can use data that could not formerly be used by the traditional formulation of the theory. Sometimes, for example, through the Monte-Carlo technique, data can even be manufactured if they were originally not available. Therefore the old controversies with respect to the type of data and methods of computations can be resolved. If the data suit the mode of computation assumed by the theory, there is no problem. However if the data do not fit well, as in the marginalist controversy, there might be ways of reformulating the theory in such a way as to enable it to use this type of data. In fact the tendency today is to build theories or reformulate older theories in such a way that they can directly absorb the data of the real world, for example in linear programming, which, can without too much manipulation use accounting data.

In the end, Machlup was quite prepared to admit into his marginalist camp businessmen who were not very proficient in calculus, not say arithmetic.[7] The only entrance requirement was the "marginalist way of thinking" which is a rough maximizing process not far from an optimizing process. It seems that he could today extend the membership to many other people.

(3) "Satisficing"

Finally there are 'satisficing' methods. The two previous methods of computation require reasonably (or made reasonably) precise data. Yet there are cases when the data are intolerably insufficient because of: (a) ignorance, or (b) the cost of securing information is prohibitive, (c) uncertainty, (d) a particular process of decision making in the organization. Then, 'satisficing' methods or processes will be put in use. These methods enable the firm to reach its goal(s), yet they do not entail primarily either calculus or operations research. For instance, the managers may perceive the alternatives exclusively through a learning process. Ignorance is gradually conquered through sequential decision-making "where the additional information generated as a consequence of actions provides the basis for future decisions" (133, p. 199). In such a situation, it is not so much that the managers are against marginal analysis, but

---

[7]Machlup (131, p. 43) where very rough calculations are accepted.

simply that they cannot use it. The alternatives do not present themselves as a collection of final positions for the firm out of which (if they are familiar with calculus) the managers can pick the one that maximizes profits. For instance, one bit of data may become available only after another has been dealt with, each case entailing a decision which is either to accept or reject only one alternative.

In such a model, uncertainty presents itself in such a manner that maximizing methods cannot take care of it. Machlup (131, pp. 54-56) gets out of this difficulty by retreating into what he himself calls "empty tautologies". He now deals with a 'firm' which is no more a firm. In a behaviour context, uncertainty leads to the establishment of safety margins and in general to a "pattern of behaviour which will protect itself from the unknown while seeking ever growing profits" (133, p. 189). This does not mean that it will tend toward maximum profits; but that the objective will now be "satisfactory profits".

Finally, the fact that the firm is not only confronted with ignorance but that it must deal with ignorance through its organization presents further complications in terms of conflicting goals within the decision-making unit. This second factor usually reinforces the implementation of "satisficing methods" (196, pp. 263 ff.). One goal, such as survival, usually summarizes other conflicting goals, and in the case of ignorance it is ideally achieved through "satisfactory

profits".

What is the effect of all this? It is important in terms of types of theories that will eventually be used. The reasons are thus:

(1) As seen above, computational methods alter or even determine the modes of behaviour of the firms. They may also alter their goal(s). Therefore, the type of data available will determine which type of theory will be used and which postulates inputs will be fed into the theory.

(2) New computational methods change the degree of reconciliability of the theories with reality and their testability. For instance, with linear programming it is possible to establish theories which will use less abstract postulates. Similarly, the new computational procedures will permit theories to state hypotheses (predictions) in a language more akin to the facts of the real world. Previously armchair economists in their ivory towers could coolly theorize about the firm which was 'owned and operated by an ever lasting bachelor who had no relative or associates, etc...', build a model out of this and let the business economist struggle to make both worlds meet. Today they don't have to do so.

(6) Partial Theories Should not Contradict other Successful Partial Theories Dealing with the same Phenomena.

This brings us to study the relative merits of a "general theory" embracing every aspect of behaviour versus special

theories which explain one aspect of the behaviour. This question (as far as the theory of the firm is concerned) is the subject of lively discussions in the literature.[8]

Economic theory, like any science, progresses through special theories or special sciences. Material phenomena were first studied by philosophers who built a special section of their science under the name of cosmology. Afterwards, their descendants (the scientists) divided this material philosophy into sections, for example, a chemical section where the color, and composition of the bodies change, and a physical section where only the location of a body is changed. For many years these two sciences (chemistry and physics), which are partial views of material phenomena, proceeded almost independently. Then parcelling occurred: biology got out of chemistry, and mineralogy from physics,

Nevertheless, nowadays, there is a tendency to reunite them into inter disciplinary sciences, for example, in biochemistry, atomic-physics, astro-physics, (the atom was originally studied under the scope of chemistry). This is justifiable, because if the phenomena are very complex, it is likely that a complete explanation of them can be given only by a pool of all these partial sciences.

---

[8]See for instance: Cohen (84a), Penrose (160, pp. 27-28), Mason (139, pp. 221-222), Papandreou (156), Boulding (26).

Similarly, human phenomena are very complex, and began to be studied by various disciplines, for example, sociology, psychology, economics, each, specializing in an aspect of the behaviour. Economics itself broke the problems into various aspects through what may be called partial economic theories: price theory, employment theory, monetary theory. Price theory includes market theory which is made of two sub-theories: consumer behaviour and the theory of the firm, (one explains the demand side and the other the supply side). That is why, as has often been said, the traditional theory of the firm is but a part of price theory (160, p. 11).

Nevertheless, in economics as in any other sciences, partial theories should not contradict other partial theories, for instance, atomic-physics should not contradict the Mechanics. If they come in conflict, they cannot both be right. Actually experience has shown that progress in astrophysics helped atomic-physics. This illustrates the well known adage that 'truth is indivisible' (119, p. 139). Yet not everybody subscribes to that. For instance (22, pp. 166-167) when studying two phenomena a and b, and judging of the value of a theory only by the criterion of successful prediction, if theory A predicts a but not b, Friedman would rate the theory as a good theory, while Koopmans would declare it false. This need not be so. In the dichotomy suggested above, (classifying theories in general and partial

theories), theory A is a good partial theory if it successfully predicts a (and satisfies the other criteria) and also does not contradict another partial theory which although not predicting a would predict b. The same criterion has to be applied to partial science B which predicts successfully b. Here both theories A and B are partial theories of a more general theory X (not yet organized) which would simultaneously predict a and b.

Applying this to the establishment of a theory of the firm a few points can be made. It has already been noted that the theory of the firm is a sub-theory of a price theory. This price theory is a sub-theory of general economic analysis which has the sole function of enabling the economist "to come to grips with the real world" (146, p. 1).

As such, as has already been noted, the economist did not think that his theory had to explain the individual behaviour; only the market values had to be explained and this was considered a sufficient grip on the real world. In this way, it was thought that most criticisms of the theory of the firm in terms of lack of realism were directed at a strawman.

On the other hand, the traditional theory of the firm is not the only partial theory dealing with the behaviour of the entrepreneur. For instance, decision-theory also tries to explain the behaviour of the entrepreneur. One begins to wonder when it is found that one partial theory uses as a

postulate (i.e. the maximizing hypothesis) something which has been notoriously found to be false in another partial theory. Until a reconciliation is worked out, by showing one of them as wrong or by showing that actually the concepts used by the two theories do not account for the same thing, grave doubts can be entertained as to the value of the theories.

This in no way suggests that we should immediately scrap the traditional theory of the firm, and work a more "general theory" which would engulf both traditional price theory and managerial economics. I am quite prepared to follow Machlup when he advocates a special theory for each problem, but not at the cost of theories which contradict each other. Also, before using a theory it should not be required that it satisfies equally well all six criteria, especially if no other theory is available. These criteria form a sort of ideal type. A theory should be appraised with regard to the degree with which it satisfies them.

SECTION II: LEGITIMATE EXTENSIONS OF THE TRADITIONAL THEORY OF THE FIRM (1939-1960)

The state of the traditional theory of the firm (dealt with in chapter VII) pertained to the elements of the theory as they were available in the 1940s. Hereafter the contributions to the contemporary theory of the firm are from two sources:

(1) From those whose task seems to polish up, meet the criticisms and remedy the insufficiencies of the traditional theory of the firm while preserving the former framework created by the fundamental postulates of profit maximization and marginal analysis. These people are called "traditionalists". In this section we will restrict the survey mainly to the contributions on the "demand side" of the theory of the firm.

(2) The other source of contributions is from those who want to obviate the already mentioned defects of the theory of the firm, but do it in a larger framework. These people have been called the revisionists and will be dealt with below in section IV.

In the 1940s the theory is not very Marshallian; at least it is not as much flesh and blood as Marshall would have liked it. The purely mechanical advances, along Cournot lines, of Chamberlin and Mrs. Robinson, contributed to bringing the economist's firm toward the managerial firm, but the distance between the two was still enormous. In Chamberlin, already, the competitive parameters are schematized in three main classes: price, product, selling activities. This opened one road of expansion, (along the traditional Cournotian approach of marginal analysis, profit maximization). It consisted mainly in exploring how more complete models of price variation (especially in the field of oligopoly), product variation, could be built.

(1) The Price Variable and Oligopoly Problems:

The models established by Mrs. Robinson, Chamberlin, Stakelberg, were quite complete, at least as far as profit maximization was concerned. However sometime later Hall and Hitch and Sweezy shook the walls of this edifice with their contention that the firms do not use marginal analysis to establish prices. This called for some interpretation, refutation, and assimilation by the traditionalists, notably from Stigler (206, pp. 410 ff.) and Machlup (131, pp. 471-474). The attitude of Stigler and Machlup was to minimize the importance of the kinked demand curve.[9] Yet an important exception is found in Chamberlin[10] who welcomed it as a "further development of the theory of monopolistic competition" (38, p. 319).

The Sweezy version was shown (by Stigler) to be compatible with marginal analysis and profit maximization. Assimilation of both versions of full cost pricing would have been useful because it was taking care of an important managerial practice. Yet these extensions of the theory were found defective on many grounds. The first one is that the methods have no built-in mechanism to explain and predict

---

[9](131, p. 353); also Stigler (206, p. 421) who does not find any empirical basis for it.

[10]Chamberlin (38, p. 319). This is followed by Chamberlin's claim that the principle of full cost pricing "has been from the first an integral part of "his analysis. (ibid., p. 319).

the "height" of the kink (102, pp. 19-34). Secondly, it is mainly restricted to the large groups or to the small groups under collusion. In the case of small groups where mutual interdependence is recognized indeterminacy creeps in; here both the traditional theory and its extension by full cost pricing are unable to provide for a solution.[11]

The critics[12] thought at the time that the way out of these difficulties was to drop the profit maximization hypothesis, especially since these indeterminacies meant uncertainty for the firms. Rigid prices, i.e. full cost pricing, kinked curves, can be explained they maintain, not by profit maximization, but by another goal: security.[13] The "security" goal can be obtained by the firms holding a certain position in the market structure. This can be achieved through various means: sales position, cost position, advertising. This leads to conflict between the various departments of the firm. In the end the price is no longer a function of costs and demand, but "tends to be the outcome of conflicting tendencies within the firm" (180, p. 456).

The above is a sample of how difficult it is to extend the traditional theory of the firm. When the theory

---

[11]Rothschild (180, p. 446): "the question cannot be solved within the framework of existing price theory."

[12]See, for instance, Rothschild (180, pp. 450 ff.).

[13]Since price fluctuations (required by profit maximization) might jeopardize the 'peace' between firms. See Rothschild (180, p. 452).

sets out to predict the price in the hinterland between competition and monopoly, it may do so only at the cost of having to drop one of its basic behaviour assumptions: profit maximization. The answer to the Rothchild's suggestions, that is against the introduction of security as a primary goal, has usually been in terms of long-run profits as a greater goal involving both security and short-run maximization of profits (131, pp. 424-428).

The source of the difficulties of the traditional theory is now identified as being the situations of oligopoly and duopoly. Extensions along orthodox lines had then to be pushed in this direction.

Here also the path of expansion had already been traced by Chamberlin in the 1930s,[14] when he subdivided the oligopoly problem in two branches:

(1) mutual dependence ignored

(2) mutual dependence recognized.

It seems that the first type corresponds to the Cournot-Bertrand-Edgeworth cases, while most modern extensions consists in exploring the ways and means through which the interdependence can be recognized. However as will be realized below, many of the extensions will not be able to be made exclusively through orthodox channels. Many extensions will

---

[14]See his memo about that in his paper "On the Origin of Oligopoly" in (40, pp. 37 ff.).

consist in devising warfare tactics, a political process of collusion, tacit agreements, leadership and various means of reducing risks or increasing knowledge.

The contributors (80) begin by questioning the stability of the classical reaction curves (80, p. 66). In their place is substituted the strategy of quasi-agreement. Note however that this course of action still permits joint profit maximization (80, chap. 4). This does not exist for long, because under oligopoly "profit maximization in the usual sense is not always the most effective way of acting under the economic incentive" (80, p. 145). Uncertainty under oligopoly may prompt some firms to work with safety margins. In practice this is translated by cost-plus methods of pricing. In Fellner (80, p. 157) these safety margins are not reconciliable with long-run profits maximization. When more real world oligopoly tactics such as price warfare, are introduced, the analysis becomes a ticket of admission to institutional economics (138, pp. 64-65).

Oligopoly rivalries are manifested not only through price competition but nowadays by non-price competition. That's why we must produce Brem's contribution on this topic. Brem's excursion into the field of oligopoly competition has not attracted much attention.[15] His model is built to solve

---

[15]His book was not even reviewed in the American Economic Review. Only lately (see 194, pp. 33-34) was it incorporated in the standard body of literature on duopoly and oligopoly.

a problem of differentiated duopoly; its distinguishing feature is that competition is in terms of quality variations, selling efforts; the price parameter is not used. When the analysis is worked within such a framework, "quality variations and selling efforts are usually the outcome of long processes" (30, p. 172) which produce time lags between action by one duopolist and reaction by the other. This time lag in turn produces the condition of imperfect information thus rendering other models, such as the one of Neumann and Morgenstern, inoperative.

In the Cournot tradition, his apparatus uses a set of reaction curves in a two dimensional diagram. This restricts his analysis to duopoly, although he claims his conclusions are valid for oligopoly (30, p. 188). Next the analysis is focussed on coalitions, and is claimed to be superior to the theory of games as expounded by Neumann and Morgenstern which cannot deal with coalitions under oligopoly (30, p. 184). This tendency to enter into coalitions is introduced, as in all traditional models of oligopoly, by endowing the participants (at least one of them) with a certain mode of behaviour: he does not like uncertainty (30, p. 185). The result is as follows: when a duopoly situation is restricted to quality competition and selling efforts, their time lags introduce uncertainty. Since "in the real world people dislike uncertainty" (30, p. 185) they enter into coalition. These coalitions can be entered

or induced by explicit bargaining (30, p. 195) or through strategies of intimidation, for example cutthroat competition in terms of quality and selling efforts (30, p. 221). In other words another behaviour postulate is thrown in: not only one duopolist is scared by uncertainty but he is also also influencable by bluffing. The leader can then induce his opponent to enter into a coalition under favorable terms. Yet compensation by one party to the other is not brought in.

Therefore Brems extends the traditional theory by integrating the time element as an essential component; for example quality competition takes time to operate. Furthermore the traditional theory is led away from simple price competition by a more realistic model where quality competition is the main weapon. The model is therefore compatible with price rigidity which was almost impossible to explain previously.[16]

Oligopoly problems can also be studied from a different point of view. Following Mason's dictum, the factor of entry should be examined. This topic is not foreign to traditional analysis. The traditional model presupposed not only free entry but something quite automatic about it. For

---

[16]Yet everything is not perfect. He continues by illustrating that quality competition proceeds by moves and countermoves which are part of an irreversible sequence. Contrary to price which can be increased after having been lowered for an interlude of economic warfare, quality cannot be reduced; for instance, Ford cannot revert to the Model T. See Brems (30, p. 244).

instance the group model of Chamberlin is essentially based upon that type of mechanism. Group equilibrium in Chamberlin is a situation where nobody gets in or nobody gets out because the profits are normal.

J. S. Bain sets out to examine this process of entry. He finds four types of barriers to entry: scale, product differentiation, unfavorable cost situation for any new entrant, and capital requirements (13, p. 169). Therefore "entry" is no more automatic. For instance, in the case of homogeneous oligopoly (undifferentiated product) and J shaped cost curves, a higher price than the "lowest cost level associated minimum optimal scale" can be maintained "without attracting further entry" (13, p. 93). This is explained by the factor of interdependence which is recognized not only by the firms already in, but also by prospective newcomers.[17] There is then a critical price (an entry-forestalling price) high enough to permit extraordinary profits, but not high enough to permit entry. More precisely (and this time we base ourselves on the work of Sylos)[18] the difference (premium) between this entry forestalling price and the lowest cost level price "tends to increase with the importance of economies

---

[17]In this case the prospective entrant does not figure with the present ruling price but with the price that will rule after he will have entered. And the present price although permitting extraordinary profits may not permit the viability of a new member. Bain (13, p. 93).

[18]See Paolo Sylos Labini, Oligopolio e progresso technico ("Oligopoly and technical progress") cited by F. Modigliani (147, p. 215).

of scale and to decrease with the size of the market and the elasticity of demand" (147, p. 220).

In the case of Bain, a by-product of his empirical researches was the confirmation of the J cost curve for many industries (13, pp. 81-82). This enables the models based on that type of cost curve to explain why full cost pricing is so prevalent. The novelty here is that the Sylos' and Bain's models explain the height of the kink and render this rule of thumb compatible with profit maximization.[19]

It seems, then, that except for Fellner and a few instances in other authors, the traditional theory, with the help of behavioural postulates whether obtained a priori or empirically (as in Bain), could be extended a long way while still keeping to rough approximations of profit maximization and marginal analysis.

(2) Extensions of the Theory of Product Variation and of Selling Costs:

A movement toward greater realism can be initiated if theory accounts for another managerial tool: changing quality and advertising.

This was already contained in Chamberlin's original

---

[19]"Now it can be verified that, at least for moderate variations in costs and well-behaved demand functions, a good approximation of the new critical price can be obtained precisely by adding to the new average cost the very same profit margin that prevailed before the change; and nearly as good an approximation can be applying to the new price cost the original total percentage markup" (147, p. 226).

work. However the original analysis was partial equilibrium. Chamberlin's technique amounted to drawing a cost curve for each version or quality of a product and finding in each case the optimum price and output and the corresponding profit. A comparison of the various profits corresponding to the various qualities is made; this permits the choice of the particular quality that will maximize profits. A similar technique can be used in the case of selling costs. According to Chamberlin[20] product variation is qualitative; the final solution requires a set of many diagrams.

The recourse to a series of diagrams was obviated (at least in the case of selling costs) by Boulding (23, pp. 581-587) when an indifference curve technique involving selling costs, price of the product, and output which optimum values can both be simultaneously determined. Some years later Stigler[21] devised a similar technique (this time involving two diagrams is taking care simultaneously of price, output, and amount spent on selling efforts. Yet, since these techniques are directed solely at solving problems of selling costs; therefore the imperfection of Chamberlin's

---

[20]Chamberlin (36, pp. 78-79): "Another peculiarity is that 'product' variations are in their essence qualitative, rather than quantitative; they cannot, therefore be measured along an axis and displayed in a single diagram. Resort must be had, instead, to the somewhat clumsy expedient of imagining a series of diagrams, one for each variety of product." Also ibid., pp. 80-81.

[21]Stigler, (203, pp. 260 ff.). Curiously this technique disappears in the 1952 edition.

model of product differentiation, for example, the non-measurability of production differentiation, remained. Remedying this deficiency will occupy many minds in the years to follow.

The first attempt to introduce more realism and measurability was made by Brems in 1948 (29). The schema of Brems consists in determining, at a given price, how the cost of producing and selling one commodity varies as its "quality" varies quantitatively according to some criteria. He then suggests that the quantity sold of a product depends upon the size of the criteria (or criterion). For instance, the greater the wheelbase (in inches) of an automobile the greater the amount sold, at least up to a certain point. He integrates various criteria and finally ends up with a total cost curve and a marginal cost curve both for making and for selling various quantities at a given price. Still, numerous diagrams must be used (one for each price) and the most profitable solution is picked by inspection. The claims of this approach are: (1) more elegance, the analysis is in terms of marginal cost curves, (2) it is more realistic: product variation is now measurable. Yet it is not completely successful in terms of introducing measurability, because some criteria, for example, colors, are still non-amenable to measurement.

In 1951, he extended (30) his schema to what now appears to be a problem of suboptimization. In the case of selling

costs, Boulding and Stigler had gone as far as determining what amount of money should be spent on advertising where presumably the amount is spent where it is most productive. Brems intended to solve the suboptimization problem which is prior to the above problem: "what media of selling should be applied? If more media should be applied at the same time, in what proportions, then?" (30, p. 8).

His apparatus includes now both a demand and a cost equation. The demand equation is a function involving a series of technical criteria: $q = q(p_1, q_1, q_2, q_3, q_4 \ldots q_n)$. (30, p. 27). On that topic his purpose is to determine what criteria should be used and to what extent. This amounts to giving to the manufacturer a theory of how to choose among competitive criteria. In this he does not succeed because of many difficulties. The most important one is that some criteria are not measurable, for example, tail fins. This renders almost impossible the use of mathematics, especially those elementary ones used by Brems.[22] The difficulty can be obviated by evaluating them in money terms. Yet this contention that increasing costs increase the quality (and thus sales) might be false; for example, car fenders with fins may not cost more than a different style of fender.

The construction of this demand curve is based also on some dubious postulates. For instance, it is assumed

---

[22]Brems turns around the difficulty by increasing the number of equations; at the cost of some cumbersome apparatus.

that we know how producers' criteria affect consumers' criteria, and how consumers' criteria affect quantity sold (30, p. 28). This faith that these connections hold is based on some statistical verification by Brems. In the case of automobiles he finally settles for "brake horsepower" as the crucial criterion that decided people to buy or not to buy a particular make of car. Unfortunately this mode of reasoning is not well regarded by specialists in the field (149). At this moment the discipline of marketing research does not pretend to know why people buy, but simply predicts what products will likely sell after having tried things on particular markets which play the role of samples in a statistical survey.[23]

Brems' analysis is then a case of too much extension of the theory. Actually his immense apparatus does not furnish new tools to the economies or the businessmen (42, p. 353), although his chapter V is successful in obtaining a solution where simultaneously price and product can be

---

[23]"The inadequacy of the list-of-motives approach has been noted by a few writers...A. W. Kornhouser advocated its abandonment more than 30 years ago".... "This kind of thinking (list-of-motives)...now is being increasingly recognized as an oversimplification"... "Generally, marketing research has been much more active and successful in finding out what has been going on rather than explaining behaviour in terms of buyer wants and their satisfaction in specific product"... Moreover a market research study in the sale of automobiles has shown: "that many people have no real interest in the complex technical features and do not give them detailed consideration in deciding which of several competitive and technically comparable makes to buy" Newman, (149).

varied.[24] The difference with Chamberlin is that product differentiation is now specified in terms of "measurable" criteria. His theory has of course no predictive power mainly because of faulty construction of the demand. This is not to say that the criteria of quality are not relevant. But "it is one thing...to describe those things which influence the entrepreneur, and another to show how they influence him" (199, p. 107).

In 1953, Chamberlin (39) joined the movement and welcomed a wider model of the firm where all the competitive variables (price, product and selling costs) would be on a par. The equilibrium of the firm is now defined as the firm having used all its weapons optimally. In Europe this new look of the theory of the firm bears the name of parameter theory.

At the same time, Abbott was spelling out in detail (1) the analytical equivalence of quality competition and price competition. His contribution has to be put in the following framework:[25] Ever since 1929, the theorists of location theory had shown (at least implicitly) that competition through location is equivalent to price competition,[26] and location

---

[24]The problem amounts to maximizing the function $Z = pq-C$; where Z is profits, p is the price, q the quantity, and $C = c(q, \alpha_1, \alpha_2, \alpha_3, \ldots \alpha_n)$. This formulation is more elegant than the one of Chamberlin.

[25]Abbott himself did not do that.

[26]See my M.A. thesis op. cit., p. 150.

is an aspect of product differentiation. In a model of economic space (as different from literal space)[27] a spatial model of location theory can illustrate and become a model of "quality" competition. Yet the spatial models of Hotelling or Lerner and Singer, and Smithies have one drawback: the change of location is assumed to be costless. This means that in a wider model involving quality, changes in quality would be costless. This is highly unrealistic in many circumstances but not all. The contribution of Abbott is to have worked out a Chamberlinian model of group competition in terms of price into a model of group competition in terms of quality, where changes in quality involve changes in costs. The difference between Chamberlin and Abbott is that in Chamberlin the large group model shows which price is likely to prevail when an industry is in equilibrium, while Abbott shows which "quality" will prevail, when firms use quality as a competitive weapon and where quality can be identified on a scale. In a sense this is not saying much more than did Chamberlin who also had a group equilibrium model under product differentiation only (36, pp. 94-100). In the end the contribution of Abbott is reduced to a method of integrating into the analysis the ways of building the various cost curves corresponding to the various "qualities". He simultaneously gives us a hint as to how the cost curves

---

[27]See footnote No. 6, Chapter VII.

corresponding to various products could be constructed in Chamberlin.

This is not without cost; while Chamberlin's analysis can simultaneously take care of changes in prices and product, Abbott's analysis cannot. The price is constant in Abbott. Furthermore, of all the possible types of changes in quality only those that are "vertical" (higher costs means automatically higher quality) can be taken care of. Therefore, the "realism" of this analysis is quite restricted. For instance not only does it not take care of horizontal and innovational (1, pp. 828-829) changes in quality, but it cannot account for such consumers' behaviours as those where the "quality" that influences consumers' buying is inferred not from the technical qualities (those connected with costs) but from the price they pay (186).

In his book (2) (which appeared in 1955) things are even more confusing. He spends nine chapters in exposing a theory of consumers and business behaviour which is too fantastic to be of any use. Like Brems he tries to explain too much. At the moment any 'Freudian' excursion, on the part of economists, into explaining or theorizing about the actual mechanism that govern the buying habits of the consumers, is bound to end in disaster. Consequently his "extension" of the theory has no predictive (or even normative) value. In passing we may note that there is an effort in Abbott to make the extension of the theory compatible with

a firm pursuing many goals. Under the circumstances it is compatible with both profit maximization and with sales maximization (when a satisfying return on investment has been earned) (1, p. 841).

A generalization of Abbott and Brems appeared in 1954 (66). Using elementary calculus instead of the geometrical apparatus of Abbott, Dorfman & Steiner solved elegantly the problems of the firm which competes in price, advertising, and quality. The equilibrium of the firm is attained when the function $q = f(p,s,x)$ is maximized.[28] The firm will then maximize profits when: $u = n = \frac{p}{c}, n_c$; where u is the marginal value product of advertising ($u = p\left(\frac{\partial f}{\partial s}\right)$); n is the price elasticity of the demand curve, $n = -\frac{p}{q} \cdot \frac{\partial f}{\partial p}$.; c is the average cost of production: $c = c(q,x)$, and n_c is the elasticity of demand with respect to quality variation. $n_c = \frac{q}{p} \cdot \frac{\partial f}{\partial x} \Big/ \frac{dc}{dx}$

In this perspective Abbott's case (optimal quality at a fixed price) is a special case of the general model. However it is clear that Dorman and Steiner are utilizing the basic postulates of both Abbott and Brems. They made Abbott a particular case, and from Brems they use the measurable criteria of quality. However they do not go into sub-

---

[28] Where q is quantity sold, p is the price of the product, s is the amount in dollars of selling costs, and x is a measurable index of quality.

optimizing problems, in the sense that they assume that the money is spent on advertising or product differentiation optimally. In contrast with Abbott, they emphasize the equilibrium of the individual firm and pay only lip service to the group equilibrium.

We have seen up to now that the theory of the firm has been extended from a price model to a general model where price, quality, and advertising expenditures are substitutes to determine the equilibrium position of the firm. However the whole thing is still based on measurable or unmeasurable criteria. The choice among competitive criteria of quality (a sub-optimizing problem) is assumed already solved, as for example in Dorman & Steiner. Brems (1957) proposes in this recent contribution (31) to devise the necessary sub-optimizing routines (which he had promised but failed to deliver in previous occasions) and also to have another try at minimizing the inherent weakness of the whole analysis for example, the question of measurability of the quality criteria. Brems proposes now to measure quality by the quantity of input necessary to produce one unit of input. (He uses a_{ij} as a symbol for that input coefficient). This presupposes that quality, as judged by the consumers, increases monotonically as the producer's inputs increase. This is far from being obvious. It may happen that quality can be increased only by a continuous change in the <u>type</u> of inputs. In some cases quality may be increased only by

using different raw material, for example the quality of a coat can be increased if fur is substituted for cotton. In other words, this model is still restricted to the case where quality increases only "vertically" (seen in Abbott). Furthermore this approach is compatible only with static conditions. If increasing quality means new ways of making the same thing but in such a way as to obviate to its present defects, supposing we want to increase durability, it is far from true that more inputs are necessary. Many increases in quality are curiously both labor and machinery saving. One outstanding example is the electronic industry where increasing quality of television sets (for example, increasing durability, decreasing the rate of failures, decreasing the size of the instrument) means simplifying. The first television sets had a lot of wiring, tubes, components etc., and the greater the number of parts, the greater the amount of sockets, and connections, the greater the likelihood of failures. Increasing quality meant simplifying diagrams, introducing printed circuits, transistors. This is a case where reducing labor and other inputs actually increased quality.[29]

The second drawback of Brems' method is that it is legitimate only at the industry level, while he uses it at

---

[29]See in 1952 a 17 inch television set which sold for $359.00 in Montreal could be obtained in 1959 in a better quality for $219.00.

the individual firm level. His method is borrowed from Leontief input-output analysis which postulates that a_{ij} is independent of the level of output. Brems is then sure that a change in a_{ij} is a sign of change in quality. In other words he wants to eliminate varying returns to scale. This is probably a good approximation at the aggregate level, as Leontief used it. But Brems forgot that what is legitimate at the aggregate level may not be so at the micro-level. For instance at the nation's level Douglas obtained an homogeneous production function of degree one. This does not require that each firm in the economy has a homogeneous production of degree one. Similarly at the industry level (the level of Leontief analysis) it is possible to have constant returns to the scale of the industry for instance, a constant cost industry, while all the firms composing this industry have production functions which are not homogeneous of degree one.

If production functions other than those that are homogeneous of degree one are permitted for the individual firms (as it should be, otherwise we must scrap the U shaped cost curves), a change in a_{ij} in the individual firm will not necessarily mean a change in quality.

Notwithstanding these difficulties if it is granted that quality increases monotonically as inputs do, and that the individual firm has a production function of degree one, then following Brems, the equilibrium of the firm is when:

$\frac{\pi_1}{\frac{\partial x_j}{\partial a_{1j}}} = \frac{\pi_2}{\frac{\partial x_j}{\partial a_{2j}}} = \cdots\cdots \frac{\pi_m}{\frac{\partial x_j}{\partial a_{mj}}}$ (31, p. 110) where $\pi_{1,\ldots,m}$ is the price of inputs 1, ...m; x_j the output of a firm of industry j; and $a_{1.....m,j}$ is an input 1 ..... m into the firm of industry j. Yet this complicated notation is simply the condition of equilibrium of the firm in hiring many factors of production;[30] and can be rewritten as :

$$\frac{p_1}{MP_{1j}} = \frac{p_2}{MP_{2j}} = \cdots\cdots \frac{p_m}{MP_{mj}} .$$

In the end this is not a big progress over Dorfman and Steiner or Scitovsky (187, p. 259), who also had come to similar results. However as far as Brems is concerned he has achieved his goal (at some costs!) of working the equilibrium of the firm in terms of product variation and selling efforts, when the changes in quality are expressed in measurable criteria.

This particular road of extension of the traditional theory of the firm is still popular today. As an illustration, a few words may be said about two recent contributions: those of Hahn (95) and Ozga (154).

Hahn, without noticing what has been written since 1942, sets out to remedy the lacuna which would consist in the fact that most equilibrium conditions of the firm which

---

[30] See above p. 117.

use the three avenues of competition mentioned previously, are set only as first order conditions. This was not necessary, Brems (31, p. 109) had already required that the second derivative be negative. However, since Hahn sets the second-order conditions in the context of an individual firm, it is still worthwhile to read. In another vein, Ozga deals with advertising, as a method of shifting demand curves of the firm by increasing knowledge, as distinguished from the persuasive effect of advertising. The novelty in Ozga is that advertising is not directed at changing the "tastes" of the people but simply at making the offers of the firm known to the prospective buyers. Ozga also shows the bearing of such an approach in terms of structures of imperfect markets.

(3) Extensions of the Theory of the Firm with Respect to Size and Growth of Firms:

As a final illustration of the contemporary extensions of the theory of the firm along traditional lines we might survey recent works on the size and growth of the firms.

The original theory of the firm of the 1890s (that of Walras) did not explore the size of the individual firm. The theory was more proficient in the size of the industry. When Marshall sets out also to build the supply curve, it is only incidently that there is a glimpse (through the representative firm) at the possible variations of the size

of the individual firm. Marshall admitted firms of different sizes; however, like the trees in the forest, their distribution in terms of sizes was normal around the size of the representative firm. This particular piece of analysis was weakened by many criticisms, notably those of Robbins (175).

In the 1930s, the industry supply curve is still the main center of interest (178, pp. 125-126). However now the size of the individual firm depends upon the cost situation (not on a biological analogy):

(1) In the case of constant costs (218, p. 211) the theory cannot determine an optimum size for the individual firm,

(2) with U shaped cost curves:

(a) Chamberlin through a "heroic" assumption builds a model where all the firms are of the same size (36, pp. 82 ff.).

(b) Mrs. Robinson (178, pp. 125-126) also obtains firms of the same size (the optimum size) when the rent of the scarce factors are included in the cost curve of the firm.[31]

---

[31]Yet this permits empirical verification to yield firms of different size because the data of the firm, as seen from an external observer, do not include these rents, at least they never include the rent of the entrepreneur.

The most recent (1958) contribution on orthodox lines can be found in the modernized version of E. A. G. Robinson: The Structure of Competitive Industry. The main purpose of the book is to isolate and analyse the factors that account for the size and growth of real life individual business units.[32] The results of the "extension" of the theory is to establish that in a given set of circumstances only a certain size of firm will be 'viable' or representative. Optimum size is further identified with the firm operating at the "lowest cost of production per unit" (176, p. 11), and as the one that is likely to survive in the long run and to be the preferred size in the case of starting a new firm.

Hence the logical deduction from such a statement is that the distribution of firms in terms of size should be normal around some critical value. The critical size is presumably the size that would come out of empirical researches because there are forces (competitive) that will tend "to establish (the) most efficient scale of production" implying "the assumption that all scales of production are increasingly more efficient up to the optimum size, and that a firm is therefore led on automatically from stage to stage" (176, p. 105).

---

[32]Robinson (176) isolates the following forces: technical, financial, marketing, and forces of risks and fluctuations.

However, empirical researches do not verify this prediction. First of all, the U shaped cost curve was replaced by the J cost curve in many cases (13, pp. 81-82). In this case, the single optimum of E. A. G. Robinson does not exist. Many sizes are optimum and there is no unique size which will permit the firm to survive or even at which a new firm will enter the industry. Robinson's apparatus amounts simply to predict that the forces of competition will oblige firms to clear the relatively short and unimportant initial decreasing phase of the cost curve. In other words, "the static analysis may predict the minimum size of firms in an industry, ...but it will not predict the size distribution of firms" (197, p. 608). In fact the empirical studies usually yield a Paretian distribution of firms in terms of sizes in most industries.

An extension of the traditional theory to explain the size of firms is then unlikely to succeed. Similarly it cannot explain the growth of these firms from one size to another. Very often the process of growing is through mergers or branching out into by-products or multiple products. It is then difficult to represent the size of the firm on a single axis representing many outputs. Traditional analysis can deal (when the shape of the cost curve is a U) with the size of the firm in terms of one product; it can also build a planning model in terms of long-run cost curves, to decide about multiple production; but it is inadequate to

deal with the size of a going concern where this size is accounted for by the simultaneous production of many products.

Finally, predicting the growth of firms, even in the case of single production, is awkward in the planning stage because of the difficulty of predicting the external economies, and some internal economies. Here the conventional industry model is the only place where traditional theory can predict output.

This concludes our exposé of the extensions of the theory of the firm along traditional lines. Yet this does not exhaust the subject, because extensions are also made in other aspects, such as returns to scale. Before leaving, one must note that the entrepreneur is still automatic in these extensions, and as such quite absent. We shall see below (section IV) that the entrepreneur reentered the picture in many ways recently.

SECTION III: EFFICIENCY METHODS AND THE TRADITIONAL THEORY OF THE FIRM

As extensions were being made into the formal economist's model of the firm, extremely interesting progress was being made in what used to be called management techniques. As such these techniques had never (up to World War II) interested the economist qua economist. However during the War, and immediately after, new techniques of management came in use, such as operations reaearch. Since most of the economists who became interested in them (sometimes not

as economists but as management consultants) are members of the school of "revisionists" (studied in section IV), many people were of the impression that these new tools of management science are not part of economics and especially are inimical to the traditional theory of the firm.

The present section will attempt to show that most of these tools are neutral (at least up to a certain point), and can be used in any framework.[33]

By efficiency methods is meant all these sub-optimizing routines which promote the "best" allocation of resources within the firm. Provided that the entrepreneur is successful in integrating all sub-optimizing centers in such a way that maximizing at low centers is compatible with maximizing at higher levels, the organization factor is then not an impediment to full optimization, and it does not render inoperative the traditional model of the firm. Note that the traditional theory of the firm considers only full optimization. However if there is interaction between lower centers of maximization, or if their goals interfere and become side conditions, the traditional theory of the firm does not work any more. We postpone until the next section the study of this type of situation.

The catalogue of efficiency tools is quite long. The

---

[33]Yet some play a greater role in a revisionist framework.

main ones are: operations research (for instance, linear programming), game theory, statistical decision theory, simulation, theory of inventory management, location theory, marketing theory. In general, they comprise what is included under management science which contains as subsidiary items of interest to economists: total value analysis, production scheduling, Monte Carlo analysis, economic lot-size model, queuing theory and various other methods to deal with uncertainty, seasonal variations, at low centers of maximization.

(1) Operations Research:

In the instance of operations research and more particularly of linear programming, when the production data satisfy the linearity requirement, they can be used to construct the map of isoquants of the firm. As such, the traditional theory of the firm does not bother with this problem. The traditional theory of the firm begins after this preliminary problem has been solved.

This mode of thinking involves making a distinction between economic problems and management problems. "Traditionally, economic problems have related to the allocation of resources among broad uses, taking the efficiency within firms for granted, while managerial problems have pertained to resource use within individual firms (or other organizations)" (103, p. 147). This thinking dates back to Pigou in his often quoted statement about the job of the economist.

The traditional framework restricts the economist to the field of allocation of resources through the market mechanism. Commodities or factors of production which are allocated through other means such as planning, would not then be studied by the economist.[34]

Since the distinguishing feature of the firm is planning, for example, "the suppression of the price mechanism" (47, p. 334) no wonder that traditional economics preferred to work at the industry level where the price mechanism could be used. The traditional field of economics covers economic phenomena up to the doorstep of the firm; from there, the manager takes over.

Yet, if one begins to think, there is no essential difference between the work of the economist and the work of the manager; both study the allocation of resources. What prevented the economist from going through the front door of the firm was that once inside, he was lost, having no price mechanism to guide him. Yet today linear programming permits the computation of the 'internal prices' of the resources already brought. Allocation of resources within

---

[34]This is definitely the impression we have when reading Coase (47, pp. 332-335. That is, the traditional theory went as far as determining what should be done within the firm, for instance that each factor should be hired until its marginal resource cost equals its marginal revenue product. Since Walras the internal optimum conditions of equilibrium of the firms are known. But the economist never ventured to show "how" this could be done.

the firm is then possible through a system of 'shadow prices'. Output can be optimized internally through a price system. All this sounds much like the economist's jargon, and the door is now open to let him in. He should go in and assess the efficiency of firms.

Furthermore, as noted by Koopmans (119, p. 185) this retreat of the economist from the internal problems of the firm was confined mainly to Anglo-Saxon countries. With the new tools, such as operations research, the economist might discard the traditional dichotomy between economic problems and managerial problems.[35] Not only the appraisal of the efficiency of the firms may be possible, but there may be a feedback effect in terms of the models of the firm usually constructed by the economist. For instance, modifications may be introduced in the cost curves: more "realistic" production functions will be isolated.

(2) Game Theory:

Game theory such as linear programming and calculus is a purely mathematical apparatus and as such has no economic content. Depending upon the type of employment of this mathematical tool, it may either be a sub-optimizing or a full-optimizing routine.

---

[35]Note also that "the programming man is a direct descendant of the economic man." (193, p. 105).

Mathematically speaking, there might sometimes be advantages in expressing linear programming problems in the form of a game (20, p. 136). Game theory is then a tool of sub-optimization.

It ceases to be a sub-optimizing tool when it is used as a method of deciding about the major variables of the firm, for example, pricing policies. For instance, in the opinion of the founders[36] it was supposed eventually to help solve some of the firm's problems, notably oligopoly pricing and other strategies. Typically it was supposed to describe 'rational behaviour' under conditions of limited information in such a way that our ability to predict patterns of behaviours in concrete situations would be improved (69, p. 274). Up to now these promises have not been fulfilled. The latest major contribution (194) using game theory has provided a fresh reformulation of the classical oligopoly problems and solutions, from Cournot to Fellner, but no precise solutions. That is, the theory of games has not yet been successful in predicting the behaviour of firms where numbers are few. The main reason is that "the full scope of potential outcomes cannot always be ascertained."[37]

According to our former criteria for a good theory,

---

[36]Game theory has been known for sometime before it was officially founded in economics by Neumann and Morgenstern (150).

[37]Wagner (219, p. 358). See also same pronouncement in Kaysen (115, p. 1039).

among other things, a theory should use postulates which without being necessarily realistic should reasonably account for actual behaviour; on this point the theory of game does not score very well. This is linked with the array of 'unrealistic' postulates of the theory:

(1) It requires that the individuals playing the game have a consistent pattern of preferences and strive to maximize their utility. From this consistency postulate, a new 'modern' type of utility is devised. Therefore, game theory becomes a method of solving a problem of maximizing expected 'utility' (128, pp. 4-5).

(2) Secondly, game theory assumes perfect knowledge, (not of future events), but of the preference patterns (utility functions) of his opponents (128, p. 5).

Game theory is then a model of rational behaviour (in the case of uncertainty) based on a priori postulates. The behaviour postulates are not derived from empirical data because any knowledge available is not sufficiently precise to be incorporated in a mathematical model, (128, p. 5). Game theory "is not descriptive in the sense that it describes how people do behave" (161, p. 186), but may become prescriptive (normative) under the condition that the players are all 'rational' equally well informed and sophisticated. Any direct application of the theory is then not in sight. After such

a balance sheet it is not surprising to find someone saying that "it is not clear how a theory which does not or could not explain actual behaviour can be relevant to potential behaviour" (161, p. 186).

As it stands now, its only important role in the theory of the firm will be in the field of conceptualization. The game theory might eventually furnish new models in terms of market structure and behaviour, models that were beyond the imagination of the previous contributors in this domain.[38] It will also give imagination to the business managers in conceptualizing oligopolistic relations. Before game theory, the managers could proceed inductively; having analyzed the former behaviour of competitor(s), they tried to guess the future behaviour of their opponents. With game theory, "it is possible to approach the analysis by a more deductive route. Instead of asking inductively, what we can infer from the competitor's past behaviour one seeks to determine a rival's most profitable counterstrategy to one's own 'best' moves and to formulate the appropriate defensive measures" (18, p. 348).

A subsidiary role of game theory will be to "formulate what is the nature and role of information in economic situations"

---

[38]"It cannot be reiterated sufficiently that the value of the theory of games to economics...lies with its ability to formalize vital and subtle concepts" (190, pp. 28-29).

(190, p. 23). Besides that, its future relations with the traditional theory of the firm are not quite defined mainly because they do not use the same time framework. The traditional theory is timeless, the alternatives are all available at the same time, and the problem is to choose among them, abstracting from time. While in game theory decisions and events are sequential and consequently must be timed (128, p. 8).

(3) Decision Theory:

Highly related to game theory is decision theory which can be treated under the tools of sub-optimization. In its present form it is often the method used in some low centers of the firm to determine courses of action that will optimize their situations.

The difference between game theory and decision theory is only a matter of degree. Both can be used for decision-making. However decision-making can be affected by three types of conditions: certainty, risk, and uncertainty (128, p. 13). Game theory can be used in the first two cases, while decision-theory proper is suited for the third case.

Game theory can be used under conditions of certainty because many of the situations can be decided in a linear programming framework. By relating the linear programming problem to a zero-sum game, game theory can be used. Similarly, in the case of risk (in the sense of Knight) decisions can

be taken by maximizing expected utility (computed out of probabilities given by the measurable risks). Here, also, as seen in the preceding sub-section, game theory can be utilized.

Decision-theory proper is reserved to the situation where no probability figures are available to make computations, or the probabilities are not meaningful (128, p. 13). In other words when decisions have to be taken under ignorance, some decision rules or criteria have to be devised. The mathematics of decision-theory are similar to those of game theory and are characterized by a normal-form game matrix, and it uses the same Bernouilli utility function (219, p. 385). This can even be full optimization.

Ignorance can also be present at lower decision centers. Various specialized statistical techniques have been devised to operate at these levels. They are usually encountered in the managerial literature under the name of statistical-decision theory. One example is quality-control where the statistical technique helps to determine how big a sample should be to help maintain a uniform quality in production (18, pp. 384-385). In this form it is a managerial tool, and is perfectly integrable into the conventional theory of of the firm. However if it is used to operate at the overall level of decision-making, it is difficult to reconcile it with traditional theory; the problems of reconciliation are then similar to those encountered with game theory.

(4) Simulation:

This is another routine which could be used either as a sub-optimizing tool or as a full optimizing tool. Simulation involves experimenting with an operational model. This is the next thing to real life experimentation for the economist. In the theory of the firm, it amounts to constructing a model whose parts and components account for the essential characteristics of the situation that we want to depict or the problem of the firm that we want to solve. Once the model is built (either in a mathematical form or otherwise) initial data are fed in and the model is then run. The difference between mathematical models and simulation is that mathematical models are constructed to be solved; they yield equilibrium values, while simulation involves operating a model (mathematical or not) to picture the dynamic behaviour of its components.

Simulation can be a method of finding out what will happen to the final solution of a model when some of its initial condition(s) is (are) changed. As such it is not inimical to the traditional theory of the firm which itself is a mathematical model, as with Cournot. Simulation can be used at various levels, department level or at the over-all level. As of now, I have not heard of successful simulation models at the over-all level, although models for an entire industry can be built (48a, p. 538). Simulation can even be an instrument of research to discover empirically behaviour

postulates about components of the model.[39]

Simulation[40] models of the firm are now of two types: computer type and open-ended type (48a, p. 535; 28, p. 344). The computer type functions automatically once the model has been fed into the computer. Sometimes, as in "job shop scheduling" it uses a form of heuristic programming (28, p. 345). In open-ended type of simulation each "department" or "cost center" of a firm is assigned to a person which plays the game according to its ability. The complete answer is reached partly through adding the behaviour of the decision makers to the original model. Business games of this type are open-ended simulation. This is why simulation is sometimes called "gaming" (192, p. 910).

Furthermore, simulation model can work under certainty where the components at any time or position in the model react in a definite pattern. Simulation can also work under uncertainty. This means that at some point in the process of tracing the effects in the model, the mode of reaction of the components is in the form of a probability. To deal with this added difficulty we need to apply the Monte-Carlo statistical technique.[41]

---

[39]Some business games are not designed to teach but to discover how people react in face of uncertainty. See Hoggatt, (105, pp. 192-204).

[40]For the particular methodology of building simulation models see Orcutt, (153, pp. 893-908).

[41]"The Monte Carlo technique is a process whereby data are generated by use of some random number generator." (28, p. 345).

Simulation however will proably be capable in the future of doing other things than sub-optimizing jobs. Already Schubik (192, pp. 915 ff.) assures us that although not published, there is current work in simulation dealing with the firm and its relationship to its environment and simulation dealing with aggregations of firms.

In general, simulation promises to predict, furnish norms about the behaviour of firms and be an instrument of empirical research, or an instrument of 'theoretical' experimentation (when a model is tested for self-consistency under dynamic conditions).

As long as simulation remains a sub-optimizing tool, it might be reconcilable with the traditional theory of the firm, at least in some cases. However, under uncertainty it is methodologically different from the conventional model (where the alternatives are known beforehand). For instance when we need to use the Monte-Carlo method, the choice is probabilistic; in some other cases, for example, job shop scheduling, only gross attempts at profit maximization are possible under heuristic programming. Furthermore, and now we are moving toward full optimization, it can be expected that as simulation models get more comprehensive (as promised by Shubik) and begin to be fed in terms of econometric data, the mechanistic profit maximization hypothesis of the traditional model will be discarded to make room for something resembling a behavioural theory of the firm (59, pp. 81-95).

(5) The Theory of Inventory Management:

As noted previously for other sub-optimizing methods, this is not an independent discipline. This body of knowledge called inventory management, utilizes other sub-optimizing routines like linear-programming, statistical-decision theory, simulation, calculus of variations, servomechanism, and game theory (191, p. 215; 223, p. 231). These tools are used to solve inventory problems. This involves extending the production model of the firm where output is immediately disposed of through sales to one where both production costs and inventory carrying charges are taken into account. The usefulness of making these computations and managing inventories accordingly is based on the hypothesis that it might be cheaper to produce discontinuously but by great quantities and carry an inventory to provide for a continuous flow of output (if required by sales orders). Or it might be also cheaper to carry on production continuously even if sales are discontinuous (seasonal). In both cases inventories are necessary. Which is the best method, or more explicitly how should the production flow and inventory be managed to minimize costs is the subject of a sub-optimizing technique called inventory management.[42]

---

[42]The classic in the domain is Within (223). On the literature on the economic-lot-size, Bowman and Fetter (28) can be consulted.

Shubik (191, p. 218) has summarized the theory as follows: (see also Whitin 223, pp. 231-32).

"Let Y represent the expected number of yearly sales, let Q be the size of the lot purchased, and C be the unit cost of an item purchase; S is the procurement expense incurred in placing the order; I is the carrying cost, expressed as a percentage of procurement cost. ...Then the total variable costs for the period under consideration are:

$$\frac{QC}{2} \cdot I + \frac{Y}{Q} \cdot S$$

The first term in the formula gives the average carrying charges for the period, based on the assumption that inventories are allowed to fluctuate from zero to Q. The second term gives the procurement expenses Y/Q, the number of times during the period the orders are placed; and this multiplied by S gives the total procurement expenses. Upon differentiating the above expression with respect to Q and setting the resultant expression equal to zero, we obtain the optimum lot formula:

$$Q = \sqrt{\frac{2YS}{I \cdot C}}$$

More complicated models, involving uncertainty both in terms of procurement and in sales are devised in Whitin and others (28, pp. 286 ff.). In this case some of the variables take the form of frequency distributions, and "safety allowances are necessary to allow for unpredictable variations in demand or time lag" (223, p. 20).

As such these models are oriented toward cost minimization rather than profit maximization (224, p. 552). Yet it is possible to relate this sub-optimizing routine to the traditional theory. This is done through including in the cost curves both the carrying costs and ordering costs of raw materials, goods in process and finished goods, and this for any level of output. The derivative of this curve gives the marginal cost that corresponds to the marginal cost of the conventional model (223, pp. 83 ff.).

(6) Location Theory:

When some of the previous efficiency methods permitted the firm to optimize under conditions of both of certainty and uncertainty, they were to a certain extent integrating time into the theory. On the other hand, location theory attempts to integrate the space factor into the theory of the firm.

The traditional theory of the firm is spaceless (ponctiform) in the sense that it presumes production and sales as occurring on the same spot. In the traditional theory distance from the market is taken care of by a proper correction (in terms of transport costs) in the demand curve of the firm. Although it is possible for the firm to reach equilibrium with a corrected demand curve (produced by a given location), in the planning stage the firm may have a choice to make among alternative locations. This problem of choosing among different locations is akin to the one of

choosing among different products under product differentiation. This choice is properly made under a subsidiary theory called location theory.[43]

The traditional theory of the firm, (as in the case of the production function), presupposes this problem already settled. The firm is supposed to have chosen the best location, and the theory starts from there, for example, given demand curves. This is why location theory developed itself outside the main body of economic analysis (168, p. 3).

As of now the location theory has many uses:

(1) As a sub-optimizing method, it can determine the best location of a warehouse, a plant.

(2) As a full-optimizing method it can determine the pattern of locations (in the case of a multi-plant firm) that will maximize overall profits.

(3) It can be used also as a simulator to represent the relationships of the firm with its environment; for example Enke represented spatial competition with an electric analogue (75, pp. 240-47).

(4) There has been a feedback effect on economic theory. For instance it was found that spatial

---

[43]The literature is very extensive. See Claude Ponsard (168). The literature begins in 1826 with Thunen, passes through Weber and Palander in the 1920-1930, to be restated by Hoover and Losch since then.

model was an ideal one to study duopoly and oligopoly and other market structures. The feedback consists in the independent discovery of theorems in spatial models that can be applied back into traditional theory.[44]

(5) It is now used as an illustration of models of imperfect competition or of product differentiation. In this case the literal space of the previous models is replaced by "economic space".

(A) As an optimizing method (sub or full) it amounts usually to finding the "best" location for a plant, store, warehouse. The solution of these problems can be attempted through two types of tools:

(1) Conventional location theory.

(2) Operations Research methods.

(1) The Conventional Theory of Location

It is presently available in the works of Greenhut and Isard.[45] The theory has set itself the goals of either finding the least cost location or of finding the maximum profit location. To achieve that it has isolated the following location factors:

(a) Demand factors influenced by the location of

---

[44]See Hotelling, Lerner and Singer, Smithies, Chamberlin, etc..

[45]Greenhut (92, chapters 11, 12) and Isard (112, chapters 5, 6, 7).

of markets and the location of competitors.

(b) Cost factors: which includes: location and costs of land or other facilities necessary for production; the location and cost of raw materials, labor, and other items such as insurance and power.

(c) Transport costs: of the various items involved like raw materials, finished products. Also the subjective inconveniences of being away from the markets.

Since these main factors are present almost everywhere but in different degrees or at different costs, and since for production and sales purposes they must be brought together in a unique location, the choice of a particular production location will determine the costs of operations through the addition of transport costs to the regular production costs. That particular location will also determine the net revenue from sales, (gross sales minus transport costs if the firm absorbs transport costs) or if transport are not absorbed, the location will determine the number of customers that could eventually be served at various prices. Finally a given location (and given transport costs) will determine which competitors the firm will have, and to what extent its market will be restricted.[46]

---

[46]Fetter (82). See also Hyson and Hyson (111, pp. 219-227) where they show that the boundaries of the market are hypercircles. Losch (127) will reduce the market area to a hexagone.

The relations between these various factors and their importance have been represented by various spatial models,[47] where the 'dimensional' equilibrium of the firm is determined.[48] Furthermore, the equilibrium of the firm has been determined not only under regular pricing practices but also under various refinements of standard economic theory, such as spatial price discrimination (109, pp. 182-191), discontinuities (3, pp. 212-230).

The trouble with this otherwise beautiful theory is that one important factor in the location of firms has not as yet been successfully integrated in the theory (in its predictive form): this class of factors is called: purely personal considerations (92, chap. 10). Empirical researches yielded the following result in this respect: "A study of the locations of eight small firms in Alabama reveals personal considerations as the primary force in five of the cases and a secondary specific factor in a sixth case" (92, p. 277). This reduces quite drastically the usefulness of such apparatus.[49] It seems that the higher the level of optimization the harder it is to build a comprehensive location

---

[47]See an original model in Hoover (108, chap. 3). The models before 1955 are summarized by Ponsard (168, pp. 353-404).

[48]See Greenhut (92) where a general theory of plant location is attempted in chapter 13.

[49]This made a reviewer of Greenhut remark rather caustically: "était-il nécessaire d'écrire tant de pages pour arriver à ce simple résultat" (Guitton, 93, Sept. 1958).

theory.

(2) When problems involving optimization are of a lower level, for example, the location of a single warehouse, operations research, when using a locational framework is now quite successful, in a normative sense, in determining the best location. This success is partly due to "crucial" simplifications that are made (18, p. 410) to reduce the number of alternative locations to a manageable proportion. When reduced to the essentials, the "warehouse location model" resembles to the standard transportation problem (solved by linear programming) or the inventory problem. When formulated in a linear program the problem is to minimize the cost function of operating a warehouse, keeping in mind transport costs from the factory to the warehouse and from the warehouse to the customers or points of distribution (18, p. 411).

(B) The feedback effects and the use of location theory as simulator for standard economic theory began seriously in 1929 when Hotelling studied duopoly. This was followed by similar studies by Lerner and Singer, and Smithies. The use of spatial models show that if space is taken into account and that consumers and firms are not located on the same spot, the conventional model of pure competition becomes almost an impossibility.[50] This is the first feedback effect of a

---

[50]"it appears probable, that spatial elements alone are sufficient in all, save a few special cases, to prevent pure competition in the demand and supply schedules of goods usually considered homogeneous" (74, p. 627).

sub-theory to the main theory. Chamberlin in 1951 (going back to his Ph.D. thesis) welcomed the suggestion of translating his monopolistic competition into a spatial model when he said: "we may speak of economic space instead of literal space, and conceive buyers distributed symbolically throughout a multi-dimensional area containing all the various aspects of the products in addition to convenience."[51]

Therefore, location theory can become a proving ground to "test" the logical consistency of standard economic theory.

(7) Marketing Theory:

In some quarters (18, p. 394; 7, chap. 1) marketing theory is hailed as a full fledged theory because "marketing has its special problems and may therefore...find it useful to develop further its own body of theory" (18, p. 394). On the other hand, there are people[52] who explicitly reject this kind of claim, and see marketing simply as an applied version of the "received" theory of the firm.[53]

---

[51]Chamberlin (37, p. 347). For the model see my M.A. thesis pp. 129-136. For a drawing of the demand curve of the monppolistic competition of Chamberlin see Losch (127, pp. 105-110).

[52]"It will be held that there need be no special economic theory of marketing" (Mehren 142, p. 125).

[53](142, p. 125): "that the received theory of the firm, applied to the marketing phases of production and providing for simultaneous adjustment of many profit determinants in each of many factors and product markets provides fairly adequate guides to what the firm can do and should

Without agreeing with the tenets of the existence of a theory of marketing, one might say that attempting to make of it an applied version of the traditional theory of the firm is a way to become guilty of misuse of an instrument which was never made for that. At most, the traditional theory can give some proficiency in model building.

In business administration, marketing covers all those problems or functions relating to the distribution of products from the producers to the consumers.[54] In detail they include various occupations like merchandising, selling, storing, shipping, displaying, delivering, advertising and packaging. Viewed from this angle, "marketing techniques pertain to economic theory, institutional economics, applied psychology, sociology, anthropology" (7, p. 8). Will it ever constitute a theory of its own? This depends how successful it will be in developing a specific method of connecting all these various elements. If it remains exclusively

---

do in order to maximize pfofits in various structures of competition; that such theory will indicate both the data and methods of analysis necessary for the solution of a wide variety of marketing problems...." Also (142, p. 126): "Marketing Theory is defined here as that part of the economic theory of the firm relevant to the marketing phases of the production process."

[54]"Marketing behaviour embraces a wide range of organized activities in the distribution and exchanges of goods and services. In the American Economy, distribution provides the link between mass production and individualized consumption" (6, p. 65).

an applied field of either economic theory, or psychology, there will never be such a thing as marketing theory.

Yet even if its cohesiveness is incomplete at the moment, it has until now defined certain problems of the firm, for example, location, or storing. The traditional theory of the firm used to look at these problems as sub-optimizing ones, in the sense that it supposed them already solved. If an optimum sum of $1,000. has been decided by economic theory to be spent on advertising, traditional theory supposed that an optimum allocation of this sum among newspapers, radios, televisions has been worked out. Similarly for the location of a warehouse or type of packaging used, or pricing technique, since the product can be sold to different customers over the country, the "proper" price discrimination should have been worked out. Seen in this perspective, marketing theory by solving these sub-problems, it then is a sub-optimizing technique. For instance, pricing in a retail store is not as simple as traditional theory pictures it. Detailed pricing is an art, for example, charm prices (62, p. 490, also chap. 8 and 9). Economic analysis may demand that a price be enforced. But to obtain this average price for a number of commodities, this may require a high price y to start with (at the beginning of the season) and eventually a low price β to get rid of the excess stock remaining on the shelves at the end of the season. In other problems, such as delivery, or canvassing of customers,

various operation research techniques may be used. One celebrated achievement is the solution to the problem of the travelling salesman.[55] Other marketing problems require the use of queuing theory, statistical-decision theory, linear programming.

SECTION IV: EXTENSIONS OF THE THEORY OF THE FIRM THROUGH UNORTHODOX CHANNELS

(A) Introduction:

We have seen in section II of this chapter that extensions of the theory of the firm along traditional lines was quite difficult. In many places the walls are already cracking in the sense that the main postulates of the theory must yield if the theory is to grow. Another point to note is that, dissatisfied with the traditional models, many economists want a more complete extension even if it means doing away with the main postulates of the traditional Cournotian model: profit maximization, marginal analysis, and the automatic entrepreneur. Their ambition is to build a 'more realistic' model of the firm; a model which would resemble something like a "behavioural or cybernetic" theory of the firm (26). Yet their purpose is simply to increase the predictive power of the theory within the general theory of price.

---

[55]This is a theorical solution. I do not pretend that many salesmen solve this problem every Monday morning.

Others want to extend it primarily for managerial purposes. They are appalled by the impossibility (through the traditional theory of the firm) of being able to predict in the particular case of an individual firm what will happen in given circumstances.[56] As long as this 'extension' is only for use in the managerial sphere, economists per se should not object because, manifestly, it is simply an attempt to build a managerial theory of the firm. Yet this move will become an issue if there is an attempt to reintroduce this managerial model by the back door into a standard economic model. Many people, such as Boulding, would favor such an extension of the theory for its feedback effects which are supposed to increase the worthiness of the economic theory in general.

In a way, the issue between the tradionalists and the "revisbnists" (as the above people will be called) is methodological. The questions to ask are: Do more 'realistic' models of the firm increase the predictive value of price theory? And does the traditional economic theory absolutely need their more 'realistic' models to perform its job? Finally, this presupposes that the "new" elements can be incorporated into the theory without undue damage to the former framework. Already, Machlup (130) has said no to

---

[56]"The current theory provides an inadequate conceptualization for those who wish to study the behaviour of firms per se" (133, p. 187).

Boulding on this matter.

As of now the issue cannot be solved empirically because although the extension of the theory along unorthodox lines is proceeding at a fast pace, it has not yet reached the stage of completion where it could be tested. Testing not being available, we must limit our exposition to a short account of the ways unorthodox economic theory is being developed.[57] This description may then serve to evaluate as far as possible, the chances of the "new theory" of being successful in eventually getting better results than the traditional one.

Being an extension, the "new Theory" does not dispense entirely with the whole of the traditional theory; on the contrary, the framework is similar to these distinguishing properties:

(1) It takes account of a greater number of variables.

(2) The goals of the firm might now be multiple. It does not postulate profit maximization in all cases.

(3) Marginal analysis is no more a <u>sine qua non</u> requirement.

(4) The entrepreneur becomes a complex phenomenon. The entrepreneur is no more automatic.

---

[57]Note that we will not describe the managerial theory of the firm; this is a field by itself and is not covered by the present essay.

(5) It embodies many of the efficiency tools mentioned before as an integral portion of its own theoretical framework. In this case the efficiency methods do not all assume the same form (in terms of use) they had in the traditional framework. For instance game theory may become a direct and primary method of determining the mode of behaviour and price policy.

Before discussing its methodological worthiness, let us note its differential feature. There are hopes to develop the new theory to the point where it will be a general theory of the firm "that takes the firm as its unit of study and the prediction of firm behaviour with respect to such decisions as price, output, resources allocation, as its objective."[58] The new theory is then:

(1) an individualized model of the firm,

(2) its model is oriented towards predicting the behaviour of firms in typical circumstances.

As such it must include all the significant variables that influence the behaviour of real firms. The new theory presumes that the following factors or processes are important:

(1) the organizational structure,

(2) the entrepreneurial factor,

---

[58]Cyert and March (55) as cited by Cohen (48a, p. 536).

(3) the financial structure,

(4) the methods of decision-making,

(5) the processes of establishing goals for the firm.

These factors are not independent of each other but on the contrary are interdependent in the sense of mutual causation. For example, the organizational structure will influence the locus of the entrepreneurial center. These two shape the decision-making process which in turn (with the two preceding ones) determine the goal(s) of the firm and how it will attain this goal(s). Moreover, the method(s) used to attain the goal may change the goal itself. For instance the way it deals with uncertainty and ignorance about the relevant data will determine which goal(s) it will consider attainable.

(B) The Organization Factor in the Theory of the Firm:

Organization is one feature (and an important one) of the new theory of the firm. The proponents of the new theory require the integration of the organizational factor in the theory of the firm because they regard the firm as an adaptative organism. Therefore "to predict the short-run behaviour ...or its behaviour in a complex and rapidly changing environment, it is not enough to know its goals. We must know also a great deal about its internal structure and particularly its mechanism of adaptation" (196, p. 255). Why is it so?

In other words is it true that the organization factor makes a lot of difference?

The influence of the organization factor has been established in other fields, such as public administration. Here two institutions, identical in every respect except the organizational structure, will react differently under the same external stimuli and will set different goals (195). Extrapolating, some economists inferred that two firms identical in everything that is costs, would set different prices and output (55, p. 130), if they have a different organizational structure. Furthermore, they infer that the mere presence of an organization inside a firm will automatically produce goals other than the one of profit maximization. Why?

First, an organizational structure means that decisions are taken by groups instead of by individuals and in such a case group norms prevail over individual norms (55, pp. 132-133). Presumably these group norms are "watered" down individual norms because of the implicit compromises that come when groups take decisions. Second, organization begets the problem of communication and information. This has the effect of yielding biased pieces of information as they arrive at the desks of policy makers, and of yielding biased orders as they reach the rank and file executors of orders.[59] This

---

[59]This will be developed extensively later when we will give an exposition of the influence of organization theory.

because of the "noise" which is inherent to any communication system (44, pp. 211-212). Consequently this type of theory yields a non-profit maximization hypothesis because organizations yield imperfect knowledge.

The implications of such a situation (if it exists) are illustrated by the duopoly model of Cyert and March (55, pp. 135 ff.). Where duopolist No. 1, because his data are channelled through a cost center, persistently underestimates demand, expects conjectural variation, while duopolist No. 2, because his data are channelled through a sales center, has a bias toward consistently overestimating demand, and underestimating conjectural variations. The duopoly model integrating the organizational factor yields a different solution from the standard Counot duopoly solution.

One postulate must now be justified: is it true that since information must go through some channels, they are always biased to a degree important enough to change the course of the policies? Furthermore, can this go on indefinitely, and why is it not corrected? Would not the manager realize at some time that his market research department persists in producing "inflated" sales forecasts? This presupposes that the management is so far remote from the rank and file that it cannot know the truth.

We have now to account for another product: the organizational factor changes the goals themselves. Organizations are made of components which try to maximize their individual

utility functions. Presumably, except in the case of deterministic behaviour, without authoritative coordination there would be no way of having the organization to function coherently if all its elements were permitted to achieve their own goals. The function of the peak coordinator is then to devise a system of roles which when filled will achieve the goal(s) of the organization (156, pp. 187-191). This requires organizational techniques such as planning and control. Now, it is these techniques that will introduce fissures into the single original goal which may have been profit maximization. Planning and control of roles taking can be achieved mainly by having the organization comply with a budget and a set of stand-by rules of procedures.[60]

We must now present the arguments (of the revisionists) which illustrate how budgeting (which is the chief instrument of planning and control) can lead to the non-maximization of profits.

(1) Since the budget is made much before the actual situations occur, this will prevent firms from maximizing profits in the short-run. Budgets are for one year or more (in the case of investment budgets) and as such prevent the firm from adjusting its policies fast enough to cope with the ever changing reality.

(2) The budget is a political compromise (12, p. 98)

---

[60]For the budget see (56, p. 49 ff.), for the stand-by orders see (156, p. 191).

between various departments: production, sales and advertising, and as such might suffer from "watering" down of the ideal profit maximization. The controller of the firm might favor survival policies, the salesmanager might be sales-volume minded.

What precedes has been sufficient to build a folklore which qualifies the big firms as sluggish, rigid, etc.

(3) Especially in big enterprises, there must be at least two levels of management:

(a) top management preoccupied with problems of equity, control retention, and general policies.

(b) Middle and lower management preoccupied with implementing the general directives of top management (52, p. 88).

The budget cannot be the exclusive work of top management. Since the enterprise might be dealing in thousands of different products or sizes, varieties of items, top management must limit itself to determine the main lines of the budget. That is, top management will be supplemented by middle management which will do some detailed work (in pricing, for example) but will also have to delegate power to lower management by issuing stand-by orders. For instance in the case of a department store, the budget will not determine the price of a lot of items. These items will be priced by a rule of thumb method (markup percentage) by lower management (12, p. 101).

(4) The budget can be an instrument of control only if it sets definite bench-marks or targets to evaluate the performance of various cost and revenue centers in the organization. The usual procedure is to set up the well known accounting standard costs (52, sec. 2). Now these standards could either be ideal or reasonable. Presumably the ideal would be akin to maximum performance (profit maximization). Yet, it is unlikely that these ideal performances be a part of the budget for reasons inherent to its own making. Remembering that lower management has an influence in the type of information they send to the top, in the setting of some detailed norms in the budget, it is unlikely that they will act in such a way as to facilitate the inclusion of targets which they might have difficulty in attaining. To account for uncertainty (at its own level) the lower management usually tries to obtain targets that will leave them some slack. This was later called organizational slack (55, pp. 53 ff.). The final effect is that the observed cost curves of firms usually include slack and as such are not the lowest possible cost curves (52, sec. 5). The standards included in the budget are then reasonable or satisfactory targets. This seems the main clue to infer that planners strive for satisfactory profits instead of maximum profits.

(5) Budgets may strive for satisfactory profits for two other reasons connected with uncertainty.

(a) Budgets imply prediction about future events and therefore

must account for the general uncertainty that surrounds any economic prediction (56, p. 49). One way to minimize risks in this situation might be to strive for a lower but more easily and more certainly attainable goal. This is one way to bring in satisfactory profits.

(b) There is another type of uncertainty connected with the organizational structure itself.[61] The employees and lower and middle management are not "will-less" agents. Therefore there is an element of uncertainty in predicting the extent to which they will carry on the prescriptions of the budget. There must be an allowance made for their interpretative modifications of the stand-by orders. All this amounts to modifying the ideal goal of profit maximization.

(c) There is, further, the uncertainty connected with the imperfect knowledge fed to the top management. These people do not fail to suspect this, and they must account for a further margin of uncertainty.

There are then three types of uncertainty which will be integrated in the budget:

(i) uncertainty about economic events outside the firm

---

[61](52, sec. 2). See also Granick (91, pp. 84-85) where he says that even in Russia, in practice norm revision was not carried on by lower management when it should have done so, that is when new more productive machines were introduced. The consequence it gradually accumulated slack which was obliterated only in the next periodic campaign of norm raising by Stakhanovite workers (shock workers). "Small wonder that norm making...was at a low level and was considered one of the weakest elements of the plants' work" (91, p. 85).

such as future prices of raw material, the regularity of supply, sales,

(ii) the deviation that the employees will produce in carrying out orders. The degree of this deviation is uncertain,

(iii) the deviation (from the truth) produced by the communication system. This degree of deviation is also uncertain.

(6) One must also include as a source of modifications of the budget some external influences which might reinforce the goal toward satisfactory profits and other goals. Some of these are organizational:

(a) Since in modern companies, there is a divorce between ownership and management, these two groups may be influential enough to introduce their own goals. For instance, under uncertainty, a goal of satisfactory profits may be adequate to keep the shareholders happy, mainly because of its survival component. While a goal of satisfactory profits may (presumably because of a lower price, or a greater advertising campaign) satisfy some of the goals peculiar to management: such as increasing market share, greater size of the firm.

(b) Both groups may favor satisfactory profits because maximizing profits may increase the danger of socialization, anti-trust suits. Here the personal goals of these groups might prevail over those of others. Yet this is not always so; in some cases middle management may try to maximize profits

even if higher managements (for the reasons mentioned) would prefer satisfactory profits; a recent example can be found in the law suits, against electrical manufacturers in the United States, which showed that middle management maximized profits (entered into agreements) while infringing the "official" policies of the firms.

The budget is then the means by which the organizational factor is proved to be important in modifying the goals of the firm.

This point of view has been publicized enough to invite testing. Up to now, statistical testimonies have appeared on two sides: those who accept this point of view and those who reject it.

Statistical support has been provided by Alt, Cyert and March and by Lanzillotti, to name a few. Let us note a few characteristics of these testimonies.

Alt's sample is taken from a large department store and presumably the above type of theory of the firm would best explain the pricing policies of large retail stores. Yet it might not apply at all to small stores. If a big store is defined as something over $100,000 (yearly sales) manipulating his own statistics (12, p. 95) shows that this theory does not apply to 96.5 per cent of the whole population of retail stores of the United States which do 57.7 per cent of all the business. Besides, that, a few remarks can be made about various statements of Alt. This in no way is meant

to be a refutation, but they might help build a perspective. The remarks regard specifically the operations of a department store.

(1) Alt is scandalized by "the almost casual manner in which the markup percentage is reached" (12, pp. 101 and 107). For him this is a sign of non or imperfect marginalism. The first thing to note is that the markup is a tool of expediency. Consequently, it might be more economical to have only a certain range of markups. However, one thing which was not mentioned by Alt is that in some stores there might be a great variety of markups. This is surely an indication of some attempt at maximizing returns on certain classes of items. Even within each department, for example, appliances, dress, toys, the markups differ whether the items are classified as fast moving or slow moving, or are standard items. In the department of electrical appliances, stoves never carry the same markup as automatic washing machines. Similarly in the service department radio tubes do not have the same markup as ordinary television tubes and these do not have the same markup as picture tubes (which have a lower markup). Within the same service department markups range from 25 per cent to 300 per cent. (The 300 per cent is on small condensers).

Secondly, the apparent unconcern with figuring a markup comes from the fact that a great proportion of items dealt with in a department store are branded and carry a "suggested

price" which in some countries cannot be infringed because of a resale price maintenance law or when the law has been repealed because changing the price might bring complications with the suppliers.

Thirdly, markups are flexible over time or over types of distribution set up, for instance the presence of discount houses.

Curiously, the markup of the retailer is not always a rule of thumb devised in a non-marginalist way. The mark-up may be set up in a discrete form, but marginalism is present when the decision is taken to classify "a particular item", its classification determining the markup it will carry, and by re-classifying we simply change the markup. The loss-leader is an indirect confirmation that retailers know about the elasticity of demand.

(b) Alt does not like the controller interferring with price setting by lower management. This is not necessarily a sign of non overall profit maximization because, as remarked by Alt himself the controller is varying the price while simultaneously he is varying selling expenditures. Consequently this shows only that the controller may be using marginal analysis at a higher plane. He may be trying to satisfy the equilibrium between price, quality variation and selling costs à la Dorfman and Steiner.

In 1956 Cyert and March (56) produced another sample of business behaviour, taken this time from the manufacturing sector (farm implements). Their conclusions did not stem

directly from the sample itself but from statistical manipulations which render the conclusions much less interesting. A more complete sample was provided in 1958 by Lanzillotti (120). It includes 20 companies taken from the biggest non banking United States firms. The management of these companies is supposed to have answered candidly the interviews or questionnaires by saying that they do not maximize profits in their pricing policies but that they mainly:

(1) try to achieve a target return on their investment,

(2) try to achieve a target share of the market,

(3) try to keep peace with rivals, or meet competition!

We may remark: (1) that many behaviours described by Lanzillotti are not anti-marginalist. For instance (120, p. 931) he remarked that the targets 'varied' with the kind of product, for example, higher returns on new products than on old ones. (2) The range of actual returns differed markedly from the proposed goals. For instance United Steel is supposed to achieve a goal of 8 per cent return while in fact the range was from 7.6 per cent to 14.8 per cent. Obviously 8 per cent is only a floor and the company tries to obtain as much as it can with the constraints imposed by labor unions and government tolerance. (3) The sample obviously refers to an oligopolistic situation where interdependence is recognized.[62]

---

[62]This is realized by referring to the wider study: Pricing in Big Business, by Kaplan, Dirlam, and Lanzillotti, (114). "Big business does not as a rule engage in price cutting at least of the day to day variety.... But size does not give

In such a case the demand curve is probably kinked at a critical price, and therefore the price variable is not much used as competitive weapon.[63] As may be expected in such a case pricing is only one of the competitive parameters and is used in conjunction with quality, service, and style competition, some companies relying more on these last weapons to achieve their goal(s) than on price (114, pp. 260 ff.). The main defect of this otherwise magnificent inquiry is to have tried to explain pricing policy in isolation. In fact the firm's model may very well be one where marginal adjustments are made in price, quality, and selling expenditures. In such a model price can vary (or not vary) without this being directly connected with any tangible goal. (4) No wonder the conclusions of such a study are vague.[64] (5) Finally this work illustrates the drawbacks of making inquiries through questionnaires or interviews. The authors noted that "strategic memoranda summarizing the considerations at an

---

immunity from price wars.... Hence price wars are the complement of rigidity; but they are indicative of a lack of power of the big firms to maintain consistent pricing..." (114, p. 284).

[63]For example "On the whole, the executives tended to minimize price policy as such, or to discuss it in such broad terms that interviewers initially found it difficult to apply the explanation to particular situations" (114, pp. 3-4).

[64]Talking about the big companies they had analyzed, the authors (114) conclude: "general uniformities in their pricing are wanting" (114, p. 290). This is based on rather negative findings about pricing. For instance they also admit that "it is not always evident that the price maker in the various big companies know there is a policy, even if an inconsistent one" (114, p. 276).

important conference leading to price discussion are rarely found" (114, p. 5). They note the tendency of price makers to rationalize when interviewed, or to withhold details of price making because the executives are reluctant to reveal actual costs and revenue figures (114, p. 6). This shows that the interview or questionnaire method is a very unreliable method for obtaining information of this type from businessmen. Businessmen are not politicians and their natural method of tackling problems is not to talk about them but to take decisions when obliged to do so. Therefore, when they are questioned they have a tendency to make logical their behaviour. I think it is Machlup who once suggested that the only way to "know" about business behaviour would be to "live" in a business for some years,[65] go through their files, and be as inconspicuous as possible.

No large scale inquiry has been made along these lines. However the opponents[66] of the conclusions of 'organizational theory' devised a questionnaire which is advertised as containing no leading questions.[67] The inquiry was made in a

---

[65]This has been done recently by Cyert, Simon and Trow (58) who actually planted an observer in a going concern contemplating the purchase of a computer.

[66]Early (70) and Firestone (83). Both studies are based on the same research project.

[67]Yet, it has been judged tendencious by H. A. Simon (196, p. 264).

sample of firms which resembles those of Cyert and March and Lanzillotti, but seems to account for relatively smaller firms. It seems also that being not the giants of those of Lanzillotti, the oligopolistic behaviour is less evident. As such it is not a cross section of American firms, but is limited to 110 replies from a group of 'excellently managed' firms. The position of the opponents has been stated by Early (71) as an hypothesis of modified profit maximization which they oppose to the hypothesis of satisfactory profits of Cyert and March and the satisficing hypothesis of Simon. In 1957 the hypothesis is that the tendency toward profit maximization appears along several dimensions (71, p. 234). Yet in 1960 the original position begins to be watered down. The creed is now: "We postulate rational choice among measured alternatives, and a rigorous drive for profit and growth. Our firms are not conceived as attempting to "maximize" profits over any time period, actual or conceptual, but as moving step-by-step towards reaching goals of growth, rising profit and strengthened competitive position" (72, p. ix).

If now (in 1960) the firm of the opponents of organization theory does not anymore maximize anything it is curious to note that a work of a follower of Early is devoted to prove "that a broadly marginalist model is appropriate to the economic analysis of the firm" (83, p. 1). The purpose of the inquiry seems to demolish some basic postulates of the organization theory of the firm as expounded by the

people of Carnegie Tech. For instance, contrary to Cyert and March, the inquiry shows: (1) that committee work does not inhibit marginalism, (2) that budgets are not rigid; planning is short term, (3) the selling center is highly represented in decision-making; and all these are signs of active marginalism. On this topic of marginalism, Firestone's method is to infer it in an indirect manner. The following characteristics are rated as evidence of marginalism:

> "(1) accounting techniques which segment costs and revenues of the firm's various operations, determine variable costs, and distinguish between fixed costs assignable to particular segments
> (2) organizational structure facilitating this segmentation and differentiation of costs
> (3) the use of marginal accounting...
> (4) maintenance of a 'short-dated term horizon'..." (83, p. 1).

However curiously it is <u>not</u> the frequency of these characteristics that is tested. What seems to be tested is the marginalist attitudes of the firms in market behaviour, internal group decision-making, budgeting planning, and relation of company size to marginalism. Such a haphazard experiment interests us solely because while making these explorations the inquiry produces some empirical refutations at some of the tenets of organization theory. For instance:

(1) The firms do not rely much on cost-plus pricing, especially in new products. Elasticity of demand is extensively used through market analysis by the management of the firm.

(2) "There is no suggestion in the data that joint decision-making is associated with attitudes or behaviour that would run counter to a basic profit maximization hypothesis" (83, p. 35).

(3) Through the simultaneous use of different types of budgets, (short-dated, periodic, continuous) which are modified through feedbacks, the 'rigidity' of budgets is no more an impediment to marginalist behaviour.

(4) In the case of big firms (the realm of Lanzillotti and Co.) the findings are that "the larger firms appear to be more marginalist rather than less" and "there is no evidence that large firms become careless, sluggish, or non-marginalist" (83, pp. 62-63).

These conclusions seem to prove too much, especially their tone is much too optimistic as compared with what we were promised in the foreword; usually it is the other way around. Undoubtedly, their worthiness depends upon the statistical techniques employed and the reliability of the questionnaires and interviews employed, all things which we evaluated as quite unreliable. When an outside observer reads these con-

flicting testimonies, all based on statistical inquiries, he begins to wonder about the soundness of this method of proof (this is why in our section on methodology we required much more than a simple empirical verification of the hypothesis).

Before closing this section on organization theory, it might be worthwhile to have a quick look at a third type of testimony.[68] This testimony is an a priori model simultaneously tested by an empirical situation.

The purpose of the model is to show that the organization factor has a definite and noticeable influence on the decision-making of the firm.

(A) The a priori model:

Suppose a firm which has the same production function, face the same markets (selling and buying) as another firm, but the original firm is owned and operated by its employees. This type of organizational setup is called "market syndicalism" (221, p. 567). Its employees will try to maximize their individual revenues from whatever source. Consequently the Illyrian firm (as it is called) will have to operate in such a way that each employee's wages plus his share in the profits of the firm are maximized, while the capitalist's

---

[68]This testimony is by Ward (221). Note that the model presented by Ward was probably not made with the intention on our present use.

firm would simply maximize profits.

We must note here that the organization basically changes the goal of the firm. From there, it can be shown that (in a certain competitive situation) the Illyrian firm will have negatively sloped supply curve, and will produce a different output than the capitalist firm. Interesting implications are isolated by Ward. For example:

(1) "change in the fixed costs of the competitive Illyrian firm leads to a change in output in the same direction"(221, 573).

(2) "A change in price...leads to a change in output in the opposite direction" (221, p. 575).

Therefore, not only the goal but the behaviour of the firm is changed in a noticeable way, to the extent that the traditional tools of theory of the firm are incapable in predicting events in the country of Illyria.

(B) Empirical basis:

This "curious" organizational setup is not derived either from dubious statistical investigations or from thin air, but approximates the real life situation in present day Yugoslavia.

(C) <u>Unorthodox Extensions of the Theory of the Firm Through the Personal Goals of the Entrepreneur. The Problem of Multiple Goals</u>.

The classical and the neo-classical entrepreneur and its firm (except for Schumpeter) had been simply

relays in the economic system--transmitting and reacting in a neutral manner to external stimuli. The new theory re-establishes the entrepreneur's motivation as the central figure of the firm. As such, like the organizational factor, its introduction will not be neutral as far as the process of determining and attaining goals of the firm is concerned. The traditional model assumed profit maximization on the assumption that the entrepreneur is rational. The new theory says that the rationality of the entrepreneur "is consistent with maximization of other things as well as profits" (156, p. 206). Rationality in the new theory of the firm means maximizing a given utility-index furnished by the entrepreneur. This is not necessarily identical with the empty tautology that what the entrepreneurs do is what they maximize because they behave in this way.

Up to now this is not very damaging for the traditional theory. But this picture will change when it is shown how the entrepreneur can maximize his utility function, especially in the case of uncertainty. In the traditional model, time was taken care of by maximizing the possible present capital value of a stream of expected revenues extending in the future as far as the planning horizon permitted. But when the entrepreneur is no more a simple relay, uncertainty will become a subjective component of the entrepreneur's motivation. Any objective criterion of profit maximization is now inadequate (155).

Tintner (213), Alchian (4), and Papandreou (156, p. 208) have brought to our attention the fact that when we couple entrepreneur's motivation with uncertainty "expectations are not single-valued" and this "forces us to substitute preference-function maximization for profit-maximization analysis" (156, p. 208). Actually Alchian stated that it is impossible to maximize profits because there is no such thing as a maximizing distribution (4, p. 212). Entrepreneur's motivation will determine how we will deal with uncertainty. Since many methods are valid, it is possible that he will pursue another goal than maximum profits, or even a collection of more or less consistent goals. For instance, he may put survival as the main goal and thus may prefer not to maximize profits if such a maximization increases his risks of failure. He may favor a certain assets distribution. Here are a few unorthodox theories that might illustrate decision process of the entrepreneur under uncertainty:

(1) Game theory:

The minimax solution which is obtained by the theory of games is a decision method to maximize the utility of the entrepreneur by <u>minimizing</u> his maximum possible loss (144, p. 13). Here the entrepreneur is risk-minded, and this method is especially useful when uncertainty is identified with a situation of oligopoly. This tendency will be found in the entrepreneur who prefers to play safe. He might have

a tendency to under-expend, not raise output enough to maximize profit.

(2) On the other hand, there may be the entrepreneur who is sales-minded. His method of dealing with uncertainty (especially in an oligopoly situation) might be to increase sales (read: revenues). He links his security (survival) with a certain share of the market. But since his utility function may not only include survival but also power, salary (if he is a salaried entrepreneur), growth of the firm, large sales might also be a means to satisfy his utility function.

Now, this hypothesis is no different from traditional theory if large sales are compatible with profit maximization. However the new theory holds that in most cases maximizing sales is usually incompatible with maximizing profits. The new theory is therefore different. This is not a new hypothesis,[69] however it has recently been put back into fashion by Baumol (16, and 17). The modern formulation goes like this: The "hypothesis is that oligopolists typically seek to maximize their sales subject to a minimum profit constraint" (16, p. 187). The businessmen are then assumed (1) to determine independently a non-maximum profit goal,

---

[69]The non-maximization hypothesis has a long history beginning in Schumpeter (183, pp. 92-93).

and (2) to maximize sales, (this implying that they are good marginalists) at the expense of maximum profits.

(1) The determination of a non-maximum profit target is probably the loose end of this theory. It needs to be justified. The theory hypothesises a minimum profit goal which is not the maximum profit. Here are a few reasons:

(a) Previously, we have shown that internally organizations work through budgets when the profit goal can rarely be the maximum.

(b) Externally, (as far as shareholders are concerned) the profit target does not need to be maximum to induce new capital to flow in. As long as the dividend rate and appreciation of stock (through reinvested profits) is enough to enable and satisfy present and prospective shareholders the firm may have no problem to finance capital expansion (17, pp. 50-54).

(2) On the other hand maximizing sales (after the profit constraint is satisfied) will help:

(a) To obtain funds from prospective shareholders who do not rate profits (income) as the deciding factor in choosing an investment outlet but also security; similarly for bankers who lend short-term capital and count sales as a security.

(b) The good-will and long-run considerations. These items are well served by maintaining a given

share of the market.

(c) Growing firms with respect to their problems of personnel and marketing outlets. It is easier to attract and retain personnel and marketing outlets with large and growing sales.

(d) To minimize uncertainty in oligopolistic situations.

(e) To achieve the personnel goals of the hired management or entrepreneurs. Large and growing sales increase morale, salaries, power, of these people.

The importance of this hypothesis on the model of the firm consists in yielding a solution different from the one obtained in the traditional model. For instance (17, pp. 55, 74, 75):

(1) It is possible for the output of the firm (in equilibrium) to be greater than the one under straight profit maximization.

(2) A change in overhead costs (fixed costs) might change the output of the firm.

(3) There is a bias towards non-price competition.

The Brookings inquiry (114) (besides the impressionistic experience of Baumol as a consultant) provides a partial empirical verification of the hypothesis that at least in the case of big firms, market share and target profits are goals of firms.

(D) Unorthodox Extensions of the Theory of the Firm by Integrating the Financial Factor.

The traditional theory determines price and output of the firm simply by comparing revenues and costs functions. Hence the blueprint of the operations of the firm is reduced to a profit and loss statement. However because the incurring of costs and the collection of revenues and the distribution of profits do not instantaneously offset each other, the firm achieves permanence by maintaining outstanding assets and liabilities. In other words, the firm not only has a profit and loss statement but also a balance sheet to take care of.

There are many ways to deal with this complication. One way would be to build a theory which would register all the transactions in the manner they affect the balance sheet items, and would postulate that a given structure of these items tends to persist because of a process of homeostasis (or a thermostat principle) (25, chap. 2). We cannot discuss it further at this point because any probing in this approach requires bringing in biological analogies which are better dealt with in a behavioural context.[70]

Another way is to utilize the old framework (profit and loss) with a financial constraint. This constraint

---

[70]See below in sub-section F for the approach by extending the theories of the firm through behavioural theories.

accounts for the requirement in terms of working capital which is necessary to carry on production. Up to now, the neo-classical theory of the firm was completely unable to integrate this factor in its production function.[71] Yet working capital is necessary since the two flows of money revenues and expenditures do not exactly offset each other. Hence the integration of money capital into the model is necessary. One method of doing so(the one of Gabor and Pearce) is quite complicated and amounts to a reconstruction of previous capital theories.[72] An alternative method to deal with this requirement of money capital and which has the advantage of not disturbing old controversies, is to introduce the notions of liquidity and control, that is if liquidity is too low (insufficient working capital), the present owners lose the control of the firm.

The first notable study of control retention through the financial factor was made by Reder (169, pp. 455 ff.) who hinted that entrepreneurs may not maximize profits (grow at the optimum rate) if their capital (including further reinvestments of profits) is insufficient to enable them to retain control over the firm. Actually, in a firm there is

---

[71]"There is no place anywhere in the neoclassical theory of the firm for working capital. It cannot logically be included in the production function" (90, p. 539).

[72]Of the Austrians, Wicksell and of the Neoclassicals.

a relation between liquidity, profits and control (51, p. 1205). Control may be lost through two main channels: (51, p. 1205), insufficient liquidity and insufficient profits. Control maintenance amounts to expanding the output of the firm while reconciling the exigencies of liquidity and profit making.[73] A trivial example is encountered when to restablish the cash position a firm might have to sell some of its goods (or other assets) below its target price. The contention of Cooper is that the firm is often forced to make violent readjustments which throw it off in an oscillatory manner along an equilibrium line. This line is something like an $\pi = M$ curve depicting the various combinations of π (profits) and M (cash position) necessary at various levels of output. Alternatively retention of control can be studied through a box diagram (51, pp. 1219-1221) where one set of indifference curves show the various profit and cash positions which are compatible with a given output, and where the other set of indifference curves in which each curve depicts the relative preference of the entrepreneur between losing control through illiquidity or through lack of profits, and this also at each level of output.[74] The set of tangency points becomes the locus

---

[73]The firm constantly exchanges cash for profit position via oscillation in output, until equilibrium is restored" (51, p. 1212).

[74]If the entrepreneur sees the risk of losing control through lack of M as a perfect substitute to losing control through lack of profits, this set of indifference curves will be straight lines.

of his expansion path (the method he will choose to expand output).

We must finally bring the money capital into the picture as a fixed amount as opposed to a variable amount. The traditional theory supposed the supply of capital either as a perfectly elastic curve at the going interest rate or as an upward sloping curve, if imperfections in the market were introduced. This does not correspond to today's conditions. For instance the allocation of working capital by banks is not made solely through a price mechanism (banks cannot lend at a rate higher than 6 per cent) but through a system of rationing. Funds being not unlimited, firms must have recourse to a form of capital-budgeting.

There is a vast literature covering various aspects of what is called the theory of investment. Problems of replacement decisions, lease or buy, when to refund a debt, the size of the plant, all can be studied by this method. However, at that level capital budgeting remains a sub-optimizing technique. A more interesting illustration at the over-all level of the firm is encountered when decisions to expand, contract or even move out of an industry, are considered. An interesting model based on a somewhat flimsy basis has been built by Baumol (14, pp. 119-134; 17, chap. 5). If we accept that the average rate of profit per dollar invested increases with the size of the firm,[75] capital

---

[75]Note here that the size does not mean size of output as in the traditional theory, but the "amount of owned and borrowed capital, invested in either liquid or illiquid form" (17, p. 34).

budgeting will induce firms to "grow" and to move out of previous industries as both (or either) their supply of funds can be increased and the rate at which they can be borrowed is lowered. The difference with traditional analysis is that a reduction of the rate of interest will not induce firms to increase their present output but to move out to another more profitable industry!

A by-product of this whole analysis is to induce the manufacturing of sub-goals or goals that rival straight profit maximization. The entrepreneur may now strive for liquidity and control which will become goals per se, at the expense of profit maximization. An empirical sign of the importance of these new goals is found in the tendency, especially since the last War, for corporations to practice internal financing.[76] This phenomenon is quite foreign to traditional theory where distribution of profits was automatic or at least where the firm would theoretically send its profits funds to the financial market and reinvest them according to the prevailing market rate of interest. Rationality required that if the internal rate was lower than the market rate the firm should distribute all its profits. However this does not seem the practice today. Firms do not

---

[76] "The evidence indicates that American industrial corporations are relying to an increasing extent on internal sources for both long-term and short-term capital.... Increasing liquidity has become imperative to additional growth..." (210, p. 63).

distribute the greater portion of their profits because of liquidity and control reasons especially when they want to grow. If, in addition, Baumol's hypothesis that as they will grow their average rate of profit per dollar invested will increase is accepted, this becomes another incentive to retain profits.[77]

(E) Unorthodox Extensions of the Theory of the Firm in the Fields of Size and Growth

The traditional theory being a single product approach made the size of the firm almost synonymous with the output of the product in question. The output axis becomes the size axis.[78] When this theory is extended to a multiple product case, it becomes a bit ambiguous to talk about size. In this approach, one may talk about equilibrium output for a given product or a given group of products (160, p. 11). Here, the process of growth of a firm is the method or reconciling the differing optima, where the optima are: technical, managerial, financial, marketing, and a proper balance of risks and fluctuations (176, chap. 7). Therefore the traditional theory explains the growth of firms "in terms of the

---

[77]By the same token this produces a tendency to distribute only "conventionally adequate dividends" (see 210, p. 56) and this reinforces our former postulates of non-interference of stockholders in profit goal setting.

[78]Robinson prefers to measure size "by physical unit of capacity" (176, p. 10), for example the size of a spinning firm is measured by the number of spindles. See also the same measure in Simon and Bonini (197, p. 613).

balance of advantages and disadvantages of being a particular size" (159, p. 531).

The new theory has a tendency to measure the size of firms either in monetary units, for example, the size of its assets (17, p. 34) or in physical capacity (176, p. 10; 197, p. 613) or "by some measure of the productive resources it employs" (160, p. 24), this last measure being comprehensive enough to take into account the "talents" of its managers. Properly speaking, the new theory emphasizes the "size" of the organization rather than the size of the output. Therefore when the organization is assimilated to an organism, growth becomes "a continuous on-going or 'unfolding' process" (159, p. 531). A by-product of this biological process of growth is "that there is now no optimum or even most profitable size of firm" (160, p. 2). The new theory approach to growth postulates that there is something similar to a biological "predisposition to grow which is inherent in the very nature of firms" (159, p. 531). But, in contrast with a literal biological context, the new theory "makes growth depend on human motivation--in the usual case on the businessman's search for profits."[79]

Once the notion of a single equilibrium size is out

---

[79](159, p. 531). It might be à propos to comment on Mrs. Penrose's evolution. She started (158) in 1952 with a violent sortie against biological theories of the firm which are the basis of the new theory of the firm. This produced a controversy (in The American Economic Review 1953) and it seems that she still bears the scars she collected from her

of the way, and is replaced by the notion that firms can grow indefinitely, the most darying hypothesis becomes possible. We will limit ourselves to two.

The first and more sedate one is the one of Mrs. Penrose: Firms grow because they have what can be called an "internal incentive..." expansion itself tends to create opportunities for further expansion--opportunities that did not exist before the expansion did not exist" (159, p. 532). There are two main reasons for that:

(1) During the original expansion firms get hold of factors some of which cannot be utilized completely. This creates some 'local' excess capacity in these few factors. Some readjustment, for example, increasing some others, is then necessary.

(2) Upon completion of the expansion, the planning power of the entrepreneur is released and becomes also in excess capacity--ready to plan a new expansion.

---

encounter with Alchian. This can be appreciated in her 1955 article (159) when she watered down her previous traditional position to one which is almost indistinguishable from one of a member of the "new" theory. In her book, which is simply an expansion (in words) of her 1955 article, there are numerous vehement professions of faith in her master Machlup; but still the reader is left unconvinced. This puts her in a more delicate position. For instance in the quotation above, she does not dare stipulate "maximum profits" while in her book (160) only once (p. 29) does she talk about maximizing profits; in other places (pp. 183-186) she settles for the new theory approach which implies simply more profits without maximizing them.

Why is it that the firm does not grow in one step? The firm at any moment of time has only a limited supply of managerial resources with adequate experience in a particular firm. Mrs. Penrose then postulates that experience on the part of executives is necessary.[80] Granted that, the horizon of the firm is therefore limited at any moment of time. Yet, since the management is susceptible of learning, this horizon will recede with time. This is why firms grow over time.

The second approach (197, pp. 610-611) is much more revolutionary. The basic assumption is that the Gibrat's law of proportionate effect holds.[81]

More explicitly:

(1) firms grow proportionately whatever their size,

(2) new firms are being 'born' in the smallest size class at a relatively constant rate.

This is enough for Simon and Bonini to build a stochastic model which reveals itself as a very good model for a skewed distribution, as it is found by empirical researches. This type of theory of growth is completely a priori and satisfies

---

[80] This is doubtful. Experience can be bought, for instance the popularity of consultant organizations. Yet she bases her contention on Heller's well known work: "limited managerial services are often stated by management itself to be the important restriction on expansion" (159, p. 537).

[81] (197, pp. 608-609). The assumption is that "the distribution of percentage changes in size over a year, of the firms in a given size class is the same for all size classes. That is to say, we assume that a firm randomly selected from those with a billion dollars of assets has the same probability of growing, say, 20 per cent, as a firm randomly selected from those with a million dollars in assets."

the methodological criterion of Friedman: it can predict, while the one of Mrs. Penrose cannot.

(F) Extensions of the Theory of the Firm Through Behavioural Theories[82]

In a sense, most of the theories outlined up to now in this section IV are behavioural hypotheses. What is especially called behavioural theories attempt to predict the values of the action-parameters of the firm (output, price, size) by replacing the entrepreneur as an economic man by a more complex entrepreneur whose behaviour cannot be predicted by the former rules of the game.

Behaviourism is not a newcomer to economics. It reached a peak with the Institutionalists of the period 1920-1935. They were responsible for the introduction of group action in the analysis of economic behaviour. Its recent resurgence is identifiable with three main currents of behaviourism, probabilistic, biological, and psychological.

(1) Probabilistic Behaviourism:

A few models of deducing a priori the probable behaviour of entrepreneurs have been built recently. The most simplified

---

[82]It might be premature to attempt any systematization here because there are at least two important works that are scheduled to appear soon: (1) Cyert and March: Behavioural Theory of the Firm (Prentice-Hall, Englewood, 1962), and (2) Meyer, J. R.,: Business Motivation and the Theory of the Firm.

and radical has put forward by Achian (4, p. 214 ff.). In the case of uncertainty a stochastic model can be built which, whether or not the entrepreneurs consciously devise courses of action to make profit, will conform to the characteristics of a chance dominated model. Marshall's representative firm could therefore be built through such a method, and would thus be reduced to "a set of statistics summarizing the various 'modal' characteristics of the population" (4, p. 217).

Today (196, pp. 260, 264) Simon mentions that the behaviour of consumers can be tested by 'binary choice experiment'. However as of now this had not been done in the case of firms.

(2) Biological Behaviourism:

This is still a crude form of behaviourism popularized in the 1950s by Boulding (25, chap. 1, 21) and anathematized by Mrs. Penrose (158, p. 816 ff.).

This approach is also sociological in the sense that the theory intends to explain and predict the behaviour of human groups. As such it differs fundamentally from the traditional economic theory which is based on a more or less modified version of the economic man. Drawing on "the biological study known as 'ecology'" (25, p. 5) sociologists visualize the society as an ecological system. From these, economists subdivide the aggregate society system into various groups (all related through human ecology) such as consumers, firms. Each group forms an ecological system.

In terms of model building an economic ecological system is different from the traditional ones which were either atomistic, i.e. pure competition, or mechanical, i.e. the traditional firm which could be reduced to an independent set of cost and demand functions, while in "the ecological system the relation of the components to the system (firm) as a whole stands somewhere between these two extremes. Components are not associated in a merely random fashion as in the gas chamber, nor are they rigidly connected with each other as in a machine. Instead, there is a tendency for each part to occupy a certain position or status in relation to the others: An ecological system operates something like a machine with the unique capacity for repairing or remodeling itself" (7, pp. 29-30). This leads to functionalism as the special method for the study of the firm which is now an ecological system made of a set of roles. The function of each component of a system determines both the structure of the group behaviour (7, p. 17) and the status of each role. The main product of this analysis is to show that:

(1) A business firm, when it is a group (organization) "does not have a set of desires and goals like the individual but is at bottom only a means of expressing and realizing the goals of the individuals who belong to the system" (7, p. 53). It is only in the case of the "individual bachelor, owner, entrepreneurs running small single-product firms as unincorporated enterprises" (191, p. 219) that the goals of the

components is automatically the goal of the system.

(2) Nevertheless, systems operate in some respect as if they had goals of their own. Therefore, firms will tend to achieve the goals produced by ecological systems. These goals are survival and growth (7, pp. 53-60).

This approach has many weaknesses, but it has the merit of being based on an operationally 'approved' scientific knowledge: ecology. This is to be contrasted with traditional economics which bases its behaviour postulates on a false and outmoded psychology.

Those who dismiss summarily biological analogies will have to get rid of another opponent: ecology, which curiously is not at all foreign to economics. One author pushed the paradox as to define ecology as "the economics of plant and animal societies" (7, p. 64).

(3) Psychological Behaviourism:

Its main contemporary prophet is H. A. Simon (196). This approach intends, through the use of the psychological notion of aspirational levels to present a "fuller theory of the firm" that would explain "the actual process of making business decisions" (196, p. 279). Such a detailed theory is necessary because the firm is an adaptative organism which as such cannot adapt itself completely and instantaneously to changes in its environment. Therefore in a dynamic world, to predict the short run behaviour of a firm "we must know a great deal about its internal structure and particularly its

mechanism of adaptation" (196, p. 255). Furthermore organization theory has shown that it is likely to have a multiplicity of goals (conflicting or not). This is another reason why internal processes of determining and reconciling these goals must be integrated into the analysis.

Behaviourism is visible at two levels of the theory of the firm: At the internal equilibrium level the first thing that must go is not rationality but profit maximization which is too complicated to achieve in real life (196, p. 260). Maximizing is replaced by 'satisficing'. At the external equilibrium level (the one where the firm relates itself to its environment) the behavioural hypothesis of Galbraith's countervailing power is called in for duty. The whole thing is permeated by learning theories and expectation-forming processes.

This forces the theory to be quite elaborate: it must picture the psychological laws and data of behaviour that preside at:

(a) the determination and reconciliation of goals within the firm,

(b) the use of its power and bargaining ability in dealing with other components of the economic system,

(c) the formation of expectations and the perception of alternatives.

The resulting model of the firm is much different from the

traditional one. While the traditional model is mechanical and made of a "list of prescribed or proscribed behaviours" (196, p. 274), the behavioural model "would consist of a large number of criteria...of routines for generating possible courses of action, of computational procedures for assessing the state of the environment and its implications for action and the like" (196, p. 274). The theory requires, then, a vast amount of information in terms of behaviour of the various components of each system, and this information is far from being available, as of now. However, at the moment, much work is done to fill these gaps. There seems to be two sources of behaviour information: (1) simulation, (2) econometric measurement of business motivation.

Both types of simulation (computer and open-ended) are used to supply behavioural information. A recent example[83] utilizes computer simulation to predict an initial market situation and a specific organization set up, the output, price, and profit of the participants as time goes by. The model describes the decision process, and it shows how the goals given initially are eventually changed as environmental conditions change. The model of, Cyert, Feigenbaum and March utilizes behaviour components, for example, how organizations predict the behaviour of competitors, and the steps involved in the decision process. Granted these, the

---

[83]Pending the appearance of their book we rely on (59).

duopoly model which they build, made of an ex-monopolist and of a splinter(they behave differently because of their different origins), is a remarquable fit for a real life case history: American Can Company and Continental Can Company. That does not mean that this type of analysis can a priori yield all the behaviour hypothesis needed, but it is interesting to note that with only nine decision rules plus a set of 'proper' initial data, the authors can obtain a successful prediction model.

A different but also interesting approach to devise behaviour hypothesis consists in having real people, instead of a computer, playing the game (105). In the previous model of Cyert, Feigenbaum and March, the computer is told "how" to react and "how" to revise and make his own estimates out of initial data. However once he has been taught how to behave rationally, determinism sets in. The present method uses humar reactions instead of computer reactions to obtain initial modes of behaviour and also predictions of where a system will stand after a certain period of time. As before, a certain market situation, cost and demand, are given; certain rules (of the game) are supplied, but not the detailed procedural methods (each player is free to decide his own course of action as he sees best). It is the purpose of the experiment to yield two things:

(a) we might obtain information as to how people behave in conflict situations,

(b) eventually when enough empirical research will have been done in this way, the data can be summarized and used in an analytical model that might predict the outcomes in real life situations, at least on a probability basis.

There is, finally, the econometric approach. Some people take a different course. According to Professor J. R. Meyer,[84] determining the proper postulates for a behavioural theory of the firm is basically an empirical question. This approach belongs mainly to statistical methodology. The technique is to examine the financial reports of the firms and try to isolate the relevant hypothesis by multivariate analysis, either through the techniques of principal component, regression, factor analysis, correlation, and the like. Contrary to his former study (144) the statistical technique emphasized will be principal component.

(G) Methodological Considerations of the New Theories of the Firm

(1) Characteristics of the New Theories:

Contrary to the traditional theory, which is primarily static, the new theories are dynamic in the sense that they show how the variables are affected by the passage of time, for example, sequential decision making. Yet this dynamism

---

[84] See his forthcoming book.

is not of the kind usually encountered when the static theory of value is 'dynamized'. One peculiarity is that this dynamic system does not tend towards an equilibrium of any kind, not even a dynamic equilibrium. (There is no such thing as an equilibrium size for a firm in the new theories). In the realm of the firm, as anywhere in economics, there are indisputable advantages in terms of showing the path of expansion in dynamizing the theory.[85]

The theories are simultaneously, and with a few modifications, both normative and descriptive. The normative aspect of these theories usually carry "such labels as management science, engineering economics, and operations research" (196, p. 254). The economist can now tell the manager "how" and "what" to do. The time is near when a welfare analysis based on these theories will be devised so that in the end we will have a yardstick that will enable us to determine what to do with the "good Monopolist" or how to classify mergers in "good" and "bad" ones. Let us note here that the traditional theory was never able to devise an undisputed measure of "workable competition". As to the descriptive property, being theories, the new theories will not be photographic films of what is happening in a particular corporation, but the empirical basis of their postulates surely will

---

[85]Yet this entails the cost of increasing the complexity of the theory. However this item is dealt with below.

give an "average" picture more typical of what is going on. Their behavioural content is especially designed to equip them for explaining the process of goal setting and interactions within firms. On this subject of goals, the economic man who presided at the destinies of the traditional firm is now replaced by a complex entrepreneur which is rational but where rationality does not necessarily require profit maximization.

Furthermore, as mentioned previously, the new theories integrate a lot of what is known as managerialism, (yet not the whole of managerialism). We have seen that many efficiency tools are compatible with both types of theories (traditional or revised); however in the case of the revised theories, the efficiency methods are sometimes not only compatible, but even form a portion of the new theories. This is because sometimes the implementation of some efficiency methods entails (a) structural modifications in the organization and (b) a consequent revision of the goals. Finally, managerialism is not foreign to economics. When economics is defined à la Robbins (173, p. 16) as the study of the allocation of scarce means among competing ends, managerialism becomes economics within the firm. Nobody will deny that the type of planning of the managers is exclusively concerned with the allocation of factors services in such a way as to produce as great an output as possible (where output is a form of utility). Critics objecting to the inclusion of managerialism

in the theories had to change their definition of economics. With them the function of economics is only "to explain how our economics system works" (147a, p. 157), and of course not to advise entrepreneurs how they should behave. It is not necessary then to make our economic models realistic to the point of managerialism. But even there, the revisionists will challenge the traditionalists (77, pp. 48-49).

However, no indiscriminate increase in realism should be welcomed. Increase in realism carries with itself a cost, and any extension of the theory must take into account both the advantages and the costs. On this ground, one cannot go along with Boulding who in the early days of the revision (26) was advocating realism for the sake of realism. Machlup was justified, in his comments, in temporizing by advocating the use of various models varying in various degrees of realism according to the problem at hand.

(2) Critics of the Theories:

Generally speaking, the new theories have been criticized as too complex, no better for making predictions, too comprehensive, irrelevant, not necessary, and inappropriate for achieving the purposes of the science of economics.[86]

As the theory becomes oriented towards the individual firm it is reducing the level of aggregation at which it is

---

[86]See Penrose (160), Mason (139), Machlup (130), Bodendhorn (22), Cohen, (48a), Farrell (78).

working. Now "the lower the level of aggregation at which the class of phenomena in which we are interested occurs, the greater the amount of complexity and detail concerning the behaviour of firms which should be incorporated into our microeconomic models. Increasing the amount of detail embodied in the model entails of course, an increase in the difficulty of both formulating and manipulating the model" (48a, pp. 534-535). This is not denied by the revisionists, especially since they have now the means, through computer models, to deal with this difficulty.[87]

The revisionists hold that simple models are so general and the economic world so complex as to prevent any direct applicability of the traditional model of the firm.[88]

In many cases, the argument between the revisionists and the traditionalists really boils down to the degree of comprehensiveness (amounts of detail) that should be included in the models. Nobody favors including everything in drawing up the list of important items; the difference is that the traditionalists stop before the revisionists do.

---

[87] "Computer models however can be made as complex and realistic as our theories permit..." (48a, p. 535).

[88] Already Marshall had anathemized the simplified approach: "in this world...every plain and simple doctrine as to the relations between cost of production demand and value is necessarily false: and the greater the appearance of lucidity which is given to it by skillful exposition, the more mischevious it is. A man is likely to be a better economist if he trusts to his common sense, and practical instincts, that if he professes to study the theory of value and is resolved to find it easy" (136, p. 306).

More serious is the contention that it is no better in terms of predictions than the traditional theory, and the traditionalists usually cite the purely competitive case where the paraphernalia of organization would not give different results. However once we leave this market structure, both approaches seem to become silent. The truth is that as of now we have no reliable theory of any kind, to make predictions in individual cases. But it seems that it might be worthwhile to try the new approach. Even a "traditionalist" mellowed down recently (78, p. 563) to the point of inviting his colleagues to "abandon (the) beautiful simple profit maximization theories and develop an organizational theory of the firm."

Others (160, p. 14) have questioned the possibility of translating some components (for example, the organizational factor) into the conventional costs and revenue curves. However simulation techniques seem to integrate into a single model both costs, revenue, and organizational data (59).

(3) Confrontation with Methodological Requirements:

To determine a priori the chances of the new theories of proving their usefulness and theoretical worthiness we might examine them briefly with regard to the methodological criteria established previously in section I of this chapter.

The theories usually score quite well with respect to requirements Nos. 2, 4, 5. As far as requirement No. 1 is

concerned (predictions), not much is available at the moment, at least not enough to be decisive.[89] It seems also that there is no available study made to test the overall consistency (requirement No. 3), neither to test the compatibility of its subsidiary theories (requirement No. 6).

Its main methodological characteristic is that its psychology is empirically grounded. The fact that it can deal with theoretical difficulties such as ignorance, is surely reassuring. Similarly it does not require a logician turned entrepreneur to run the firm. The degree of rationality required is compatible with what can be expected in real situations.

These methodological differences seem to be the real source of the controversies in the field of the theory of the firm. It is therefore a bit misleading to (77, p. 45; 160, p. 9) characterize the situation as a "joust" between proponents of marginal analysis and empiricists. In fact, it is a bit more complicated than that. It is a methodological contest between traditionalists and revisionists; and although the empirical matter is part of the question, other

---

[89] "First it must be shown that the theory (the behavioural theory) is at least as good as other existing theories in its ability to predict firm behaviour.... Convincing demonstrations on either side of this point are not available. In our judgement, a major portion of the effort in the next decade of research on pricing should be directed to answering this question (or to make it irrelevant)" (59, p. 81).

things are involved, such as, what elements should be included in the theory of the firm?

(4) Future:

Although professions of faith are not lacking and new converts join the ranks, the ordeal of the new theories is not over as yet. The reasons are varied:

(a) Many of its theoretical components are still incomplete. To begin with, "the theory of organization is more a set of words than a reality" (193, p. 110). Besides, we do not know much about the significance for economic theory of many of the tools such as linear programming. As of now we think that many of them have no economic content. We still do not have a complete model of the entrepreneur (196, p. 274). On the other hand game theory seems to have been deceptive.

(b) There are no solid realization imputable to the new theories (48a, p. 538). Yet some hopes are found in the undisputable fact that the revision will surely give a big boost to what is called managerialism. It is likely that accounting is and will continue to be reformed. The use of economic consultants from outside or from the staff of the firm itself will in general contribute to make firms behave more rationally and in general will furnish the basis for scientific managerialism. Economists as model builders are well suited for this type of job (18a, p. 143). In turn, managerialism may become a provider of behavioural inputs for

the economic theory of the firm; (104, p. 154) and with this feedback, the circle is completed in the sense that we may be able to complete our economic theories.

SECTION V: CONCLUSIONS

The upshot, if economic theory follows this path of evolution and extension, will be to render it completely non-political. Economics, at least that portion which is called economics of the firm, may grow to the point of being able to feed itself with all the inputs necessary, including the pervasive behaviour assumptions, without any connection with a particular political context being postulated. In other words, if it is not based on strict profit maximization, but on behavioural assumptions internally determined by the organizational set up of the firm, this theory will be equally utilizable in a capitalist context such as the United States as it would be in a socialist context. This is what we mean by the future non-political character of tomorrow's economics.

We hope in this thesis to have traced the evolution of the theory of the firm from how it appeared in a political framework like the one of the 19th century to its transformations when 20th century came with its problems of concentration of economic power, and to its eventual revisions as the capitalist system moves toward a mixed system.

APPENDIX A

THE DIVISION OF LABOR

The division of labor can take three forms:

(1) Spontaneous Case

Strictly speaking, the laborers of a society can specialize without the aid of a blueprint or scheme designed by or under the authority of an entrepreneur as we see it under present capitalistic methods, for example, the factory system.

Spontaneous cooperation and specialization existed in primitive societies under impetus other than the entrepreneur or market forces. Malinowski showed that in some societies the "production of goods and services was embedded in political, religious, social and kinship institutions. The economic activities of individuals were motivated by social and political obligations, by kinship, by friendship and by magical rituals and beliefs. Exchange was carried on by gift giving, and ceremonial distribution of goods; barter existed only with persons outside the tribe; buying and selling were non existent. The Trobrianders were a typical nonmarket people..." (167, p. 345). The tribes were not firms[1] but sociological groups which more or less

---

[1]See appendix AA.

correspond to our concept of household. Briefly, specialization can be brought about exclusively through taboos, customs, without the underlying economic motives being present.[2] Now let's not introduce the economist's "hidden" market mechanism that is supposed to govern the making of these taboos because the anthropologists are categorical: "outside of that slim strip of a few centuries in which the price-making markets and consequently money as a means of exchange have become general. Prehistory, early history, and indeed as Kal Bucker was the first to proclaim, the whole of history apart from those last centuries, had economies the organization of which differed from anything assumed by the economist. And the difference, we now begin to infer, can be reduced to one single point--they possessed no system of price-making markets" (167, p. 241).

(2) Market Mechanism Case:

Gradually some societies evolved, and eventually production and distribution were carried on with the help of markets. In the beginning this might have come about through barter, but gradually money was used to carry on exchange, savings, and transactions. Simultaneously, firms (with their entrepreneurs) entered the picture; yet they did not immediately monopolize all production. The manorial system excluding

---

[2]See also the anthropological works of R. Firth, (84, chap. IV, and 85, chap. VI).

the artisans) was a stage where the individual farmer (serf), after having satisfied his obligations toward the lord could look for a fair or other similar market device in order to exchange the surplus in his crop if any. To that degree, the market could govern the "specialization" of these households. However the manor was not a firm because the relations between the lord and the serf were not based on a system of prices for the inputs, arrived at through the working of economic forces. The lord did not run his estate the way an entrepreneur does, hiring inputs at certain prices and trying to maximize output. In the beginning he was in the role of the government of today which depends on taxes. In other words, his activity was focused on distribution; this is why he favored big outputs. However, as emancipation changed the relations between ruler and ruled, some manors may have evolved toward the firm system. Let us note that at that time the artisan was already a firm. The apprentices specialized under the blueprint and authority of a master who paid for their services and whose product he sold.

In England, much later (around 1860), many commodities involving much specialization in terms of processes, were not produced by a single economic organization, but by variety of different independent outfits (artisans), each specializing in a particular process (205, pp. 192, 193). The coordinating function was carried on exclusively through the market mechanism. Each process was the subject of a bargain in terms

of prices between the outfits. This is still the pattern today in cities like New York in the manufacturing of some women's dresses, hats, and in some branches of textiles. Empirically, it is easy, then, to show that the market mechanism can bring about specialization without requiring the participants to be parts of a single firm.

To have a case, where a commodity both in its intermediate and final stages is produced entirely by the coordination of the market mechanism, requires that more than one household be involved without any of them becoming a firm. Their relations with one another must remain those of exchange, not of production. An economic example may be found in the situation where a group of households sporadically enter into a multilateral agreement with one another to undertake a specific project, for example a bee to build a communal road, or a well. The difference between this case and the case of spontaneous cooperation is that here there is bargaining both in terms of the amount of services furnished and in the amount of product received as a reward, each household making the usual computations (with market prices) as to the value of services furnished and the product received. Yet no particular household "sells" the "product" to another.

(3) <u>The Case of the Firm</u>:

Finally, specialization can be undertaken within a coordinating unit called a firm. This is the "pin factory"

approach. It was described in detail in Adam Smith's Wealth of Nations Book I, Chapter I. Here the market is no longer used to coordinate the factors of production but instead the authority and blueprint of an entrepreneur. (see appendix AA).

The scheme of production is, then, as follows:

(1) Non specialization, (self sufficiency).

(2) Specialization. This can take place under three forms:

(a) spontaneous

(b) market mechanism

(c) within a firm.

It is now time to discuss various possibilities. The first one was: Can a firm exist without division of labor? A firm, by definition is engaged solely in activities of production. Such a firm cannot be present in the self-sufficiency case because one of the characteristics of a firm is to buy the services of the households and sell the resulting product. Obviously in a world of self-sufficiency there will be no firms, merely households. As a consequence, the firm can then exist only in the case of specialization (division of labor) and we have thus proved our point.[3] Regarding the

---

[3]The division of labour in this sentence refers to the one that is brought about by the market or the society. In other words this does not require that the division of labour is carried on necessarily within the firm. For instance, above, we referred to an example where the production of a commodity was not carried on within a firm but through the coordination of the market mechanism. The artisans involved were firms which might have been of the one-man type. This does not involve division of labour within a firm. It is in this sense that we can talk of doctors, in private practice, as firms.

second possibility (e.g. division of labor outside the firm), the foregoing schema shows that it is possible. This is seen in cases 2a and 2b. In this context, the division of labor will absolutely entail the existence of a firm only if the costs of using the market mechanism are greater than the costs of using an entrepreneur, that is the costs of setting up a firm.[4] In the carrying on of production they often are, and therefore most production is carried on by firms.

---

[4]The costs of using the price mechanism are thus:
(1) Discovering what the relevant prices are,
(2) The cost of negotiating and concluding a separate contract for each transaction of exchange between the factors of production.
(3) The desirability of making long-term contracts.
(4) The difficulty of forecasting.
etc.....
From Coase: (47, pp. 331-352).

APPENDIX AA

THE CONCEPTS OF THE FIRM AND OF THE HOUSEHOLD

(a) FIRM:

A firm is an economic organization which buys factors of production, or their services, combines them, and sells the final product. Production by a firm requires three functions:

(1) buying of inputs,

(2) their coordination by the entrepreneur which consists in deciding how much and what to buy, how to use these inputs, and in what kind of production,

(3) selling output.

Formally two persons (they may be fictitious) are involved: a seller of inputs, and a buyer of inputs (which also coordinates and resells the final product). This is not to mention the buyer of output.

Examples from our business world are easy to find, yet a far fetched one might show the fine points: A magician is hired by a promoter on the condition that he will work anywhere for a stipulated price. This promoter immediately sells this contract for a roadshow. According to our definition this promoter is a firm. From the point of view of the

economy he "produced" the magician. The division of labor (or specialization of tasks) was carried as follows. One element specialized in magic tricks, while the other (the promoter) specialized in discovering and "evaluating" magicians. Not only did the promoter "produce" but he also carried exchange transactions: he sold a contract and bought labor services. He is then a firm.

Embedded in the notion of the firm are such facts as:

(1) the outright selling of services by the household so that it renounces to any further title to the product,

(2) the risk elements in producing through time (e.g. the possibility that the selling price of the product eventually varies) are not shared by the households,

(3) coordination at the interior of the firm replaces the other methods of inducing cooperation between factors, e.g. taboos, customs, market relationships, etc.,

(4) A firm is a separate entity (a fictitious person) having a life independent from the one of its members,

(5) the firm maximizes something: profits, sales, etc..

(b) HOUSEHOLD:

Among other things, it is the economic unit which furnishes the factors' services and sell them exclusively to a firm (never to the final customers).

When production is carried by firms it engages only in exchange transactions. In other social contexts, e.g.

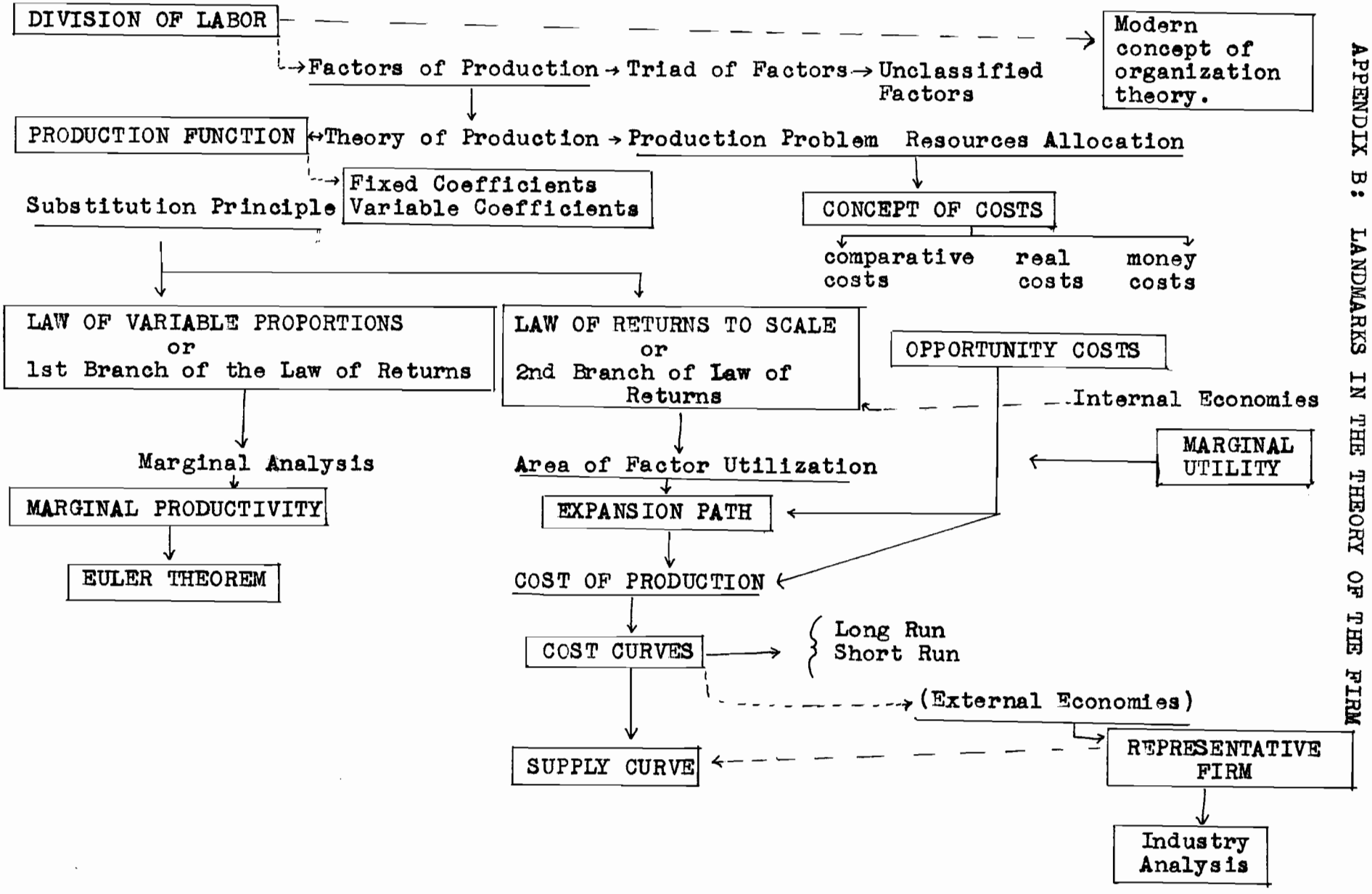
primitive societies, the serf of the middle ages, were households as long as they did not produce for a market and did not hire outside help, or hired its own members. When distribution inside the group is made on a membership basis exclusively, and not for services rendered, that is a case of a household. When distribution of product is done on the basis of the work done, the family farm becomes a firm.

Yet in the case of the early self-sufficient farm of North America it is a bit more complex. Strictly speaking they were households since the head of the family did not hire any outside help, but he coordinated the efforts, maximized output, and distributed the product on a household basis; yet this does not mean that part of the theory of the firm did not apply to him. He might be rational in the use of his resources, e.g. he might be aware of the laws of returns; but he will never utilize his resources according to the law of marginal productivity, e.g. the brightest son might have been sent away to get an education. He might not exploit or have children for economic motives.

Economic rationality does not necessarily apply to the household because as such it maximizes utility, happiness, etc., if it maximizes anything. In such a case the farm is both a household and a firm (to the extent that it uses the apparatus of the firm). Similarly the present professional man is both a household and a firm.

These hair splitting distinctions have their utility.

They help to keep the categories mutually exclusive. They put light on some knotty problems like the one of the lazy entrepreneur which some people brandish as a case where the firm does not maximize profits. Actually this case can be handled safely with the above dichotomy. In this example the traditional hypothesis of the firm maximizing profits is kept intact, and we simply postulate that the household sells too few entrepreneurial services; yet the firm maximizes whatever amount of these services it gets hold of. The utility function of the household may be peculiar but this has nothing to do with our model of the firm. The difficulty is sent back to the psychologist, permitting the economist to stand aloof of these treacherous waters.


DIVISION OF LABOR
Modern concept of organization theory.
Factors of Production
Triad of Factors
Unclassified Factors
PRODUCTION FUNCTION
Theory of Production
Production Problem Resources Allocation
Fixed Coefficients
Variable Coefficients
Substitution Principle
CONCEPT OF COSTS
comparative costs
real costs
money costs
LAW OF VARIABLE PROPORTIONS or 1st Branch of the Law of Returns
LAW OF RETURNS TO SCALE or 2nd Branch of Law of Returns
OPPORTUNITY COSTS
Internal Economies
MARGINAL UTILITY
Marginal Analysis
Area of Factor Utilization
MARGINAL PRODUCTIVITY
EXPANSION PATH
EULER THEOREM
COST OF PRODUCTION
COST CURVES
Long Run
Short Run
(External Economies)
SUPPLY CURVE
REPRESENTATIVE FIRM
Industry Analysis

APPENDIX C

BRANCHES OF THE LAWS OF RETURNS

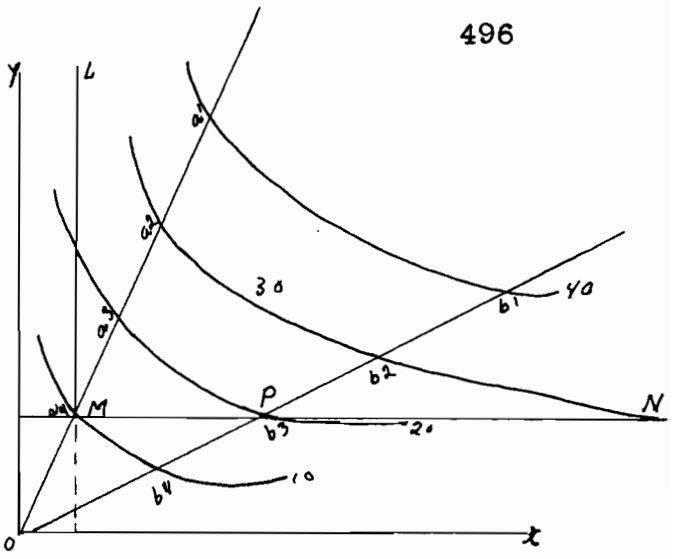
The two branches of the Law can be contrasted as follows:

Let us suppose that the firm is confronted with a certain set of relations between factors used and product produced. This relationship is called the production function. It can be pictured by an equal product contour map; that is a production diagram using isoquants. The gist of the derivation of this contour map is as follows: In the case of a firm using two factors and producing one output, the diagram needs to be in three dimensions, one axis for each of the factors (if we use only two factors) and the third one (vertical) measuring the product. As we increase the quantities of any of the factors used, the output produced increases for a certain range (as read on the third dimension). To cover all the possibilities opened by the production function, we begin by keeping one factor (x) constant at one unit and vary the other factor (y) from 0 to n units, and we note in the third dimension the locus of these points; this locus is a product line. Next we do the same thing for the ensuing value of x, for example, 2, etc. and note the locus described. If this process is carried to its limit, (covering all the values of x and y including the fractional values if it is possible), the collection of all these lines forms a surface

called a production surface. This surface may look like a mountain, that is a plane that slopes upward, increasing as any of the factors increases, until a summit is reached. It is customary to use a production function having the characteristic that the quantity of product increases at least for a certain range.

On this surface taking the third dimension as a reference, we connect all the points corresponding to the same "height" (same production) of the mountain. Allen, (10, p. 273) calls that "taking horizontal sections". This will yield a "product contour" on the mountain. If we do that to all the "heights" of the mountain, the sloping plane is now a collection of contour lines, each representing a constant amount of product.

Furthermore, since the production surface is sloping upward (in some cases until a summit and we usually do not utilize the other portion), none of these contour lines will stand vertically exactly above another one. This property is important because it permits us to project all these contour lines on the horizontal plane, reducing our diagram to a figure in two dimensions, such that moving in the northeast direction in the plane means meeting successively isoquants representing higher and higher levels of output. Furtheremore a surface regularly sloping upward means nonintersecting



isoquants. This figure will consist of two axes, x and y, for the two factors and of a series of projected contour lines (iso-products or iso-quants), all in the first quadrant. Selecting only a few of these isoquants, for example, only those labeled at a certain multiple of an original one, we obtain a diagram looking like the one above. We have now all what is necessary to separate diagrammatically the two branches of the law of returns.

In the diagram, moving in a parallel direction to any of the two axes, depicts the working of the first branch. Parallel to one axis meaning that one of the factors, (the one of the other axis) is kept constant. The production is measured at the meeting points of this parallel and the iso-quants. Eventually along MN or ML (being parallels to x and to y) we must sooner or later experience diminishing returns, for example, MP $<$ PN $<$ etc.... otherwise this law which is based on an independent empirical reasoning would be violated. The second branch of the law is illustrated by any northeast movement in the first quadrant. That is, by simultaneously increasing both factors of production. We do not need to increase the factors in the same proportion, because as long as we move in the northeast direction, the size of the firm

is changed.

There is a variety of production functions, that is of shapes and distances between isoquants (labeled as in figure above). However, only one production function at a time can be pictured on a production diagram.

To contrast the two branches of the law, let us use a homogeneous production function of degree one. In such a case on any diagonal from the origin, for example, Oa, Ob, the distances between each successive isoquant (Nos. 10, 20, 30, 40, etc.) are equal; that is

$$a^4a^3 = a^3a^2 = a^2a^1 = \text{ etc.....}$$

$$b^4b^3 = b^3b^2 = b^2b^1 = \text{ etc.....}$$

Such a case is called constant returns to scale. Increasing factors in a certain proportion will increase production in the same proportion.

Therefore, this diagram shows that simultaneously (and with the same production function), one branch, (the first), experiences diminishing returns while the other (the second branch) can experience constant returns.

APPENDIX D

EQUATIONS OF THE MODELS FROM COURNOT

(AD,1) The computations are as follows:

No cost case: total revenue is maximum when $a - 2bx = 0$.

e.g. when $x = \frac{a}{2b}$

The optimum price is: $p = a - bx$

$$= a - b\left(\frac{a}{2b}\right)$$

$$= a - \frac{a}{2} = \frac{a}{2}$$

With Costs: The optimum is when $ax - bx^2 - cx$ is maximum. This is so when its derivative is zero, or when:

$a - 2bx - c = 0$ or when $x = \frac{a - c}{2b}$

Consequently the optimum price is: $p = a - bx$ or

$$p = a - b\left(\frac{a-c}{2b}\right) \quad ; \quad p = \frac{a + c}{2}$$

(AD,2) Using Joan Robinson's formula we write $MR = p - \frac{p}{e}$, or $e = \frac{+p}{-MR + p}$. Replacing p by its value: $a - bx$, we have:

$$e = \frac{+(a - bx)}{-(a-2bx) - (a - bx)} = \frac{a - bx}{bx}$$

(AD,3) The derivation is as follows: Optimum output of firm No. 1 is: $x_1 = \frac{k_{11}}{2b}$ (because $MR = 0$). Replacing

this value in demand curve of firm No. 1 ($p_1 = k_{11} - bx_1$) we obtain $p_1 = k_{11} - b\left(\frac{k_{11}}{2b}\right) = \frac{k_{11}}{2} = \frac{a - x_{21}}{2}$

(AD,4) For a derivation see below (AD,7).

(AD,5) If x_2 increases by one unit; in other words if firm No. 2 changes his output from x_{21} to x_{22} such that $x_{22} - x_{21}$ = one unit, then x_{11} changes to x_{12} the change in output by firm No. 1 is $x_{12} - x_{11}$.

Now, if $x_{11} = \frac{a - bx_{21}}{2b}$ and $x_{12} = \frac{a - bx_{22}}{2b}$

$$x_{12} - x_{11} = \frac{a - bx_{22} - (a - bx_{21})}{2b} = \frac{bx_{22} + bx_{21}}{2b}$$

$$= -\frac{1}{2}(x_{22} - x_{21}) \text{ or } = -\frac{1}{2} \text{ (change in output of firms No. 2).}$$

If the change in output of x_2 is one unit; then the change in output in x_1 is one half of a unit. Consequently each time the competitor of firm No. 1 will increase its output by one unit, firm No. 1 will decrease his output by one half unit. The total increase on the whole market is consequently (in terms of output), one half unit.

The corresponding change in price is: (where x is total output).

If $p' = a - bx'$ and $p'' = a - bx''$

(where $x'' - x' = .5$ units)

The change in price is:

$$p' - p'' = a - bx' - (a-bx'')$$

$$" = -bx' + bx''$$

$$" = b(x'' - x') = b(\tfrac{1}{2})$$

Then the change in price = .5b

(AD,6) We have to show that if $x_2 = \frac{a-c}{b}$, then $x_1 = 0$. Here is how: to maximize profits, firm No. 1 must maximize its total revenue function which can be written as:

$$ax_1 - bx_2x_1 - bx_1^2 - cx = TR$$

The derivative with respect to x_1 or $\frac{dR}{dx_1}$ is:

$$a - bx_2 - 2bx_1 - c = 0 \quad \text{or}$$

$$x_1 = \frac{-c + a - bx_2}{2b} \quad \text{: replacing } x_2 \text{ by } \frac{a-c}{b}$$

gives $x_1 = 0$

In such a case the demand curve of firm No. 1 is equal to

$$p_1 = a - b\left(\frac{a-c}{b}\right) - bx_1$$

$$p_1 = a - a + c - bx_1$$

$$p_1 = c - bx_1$$

(AD,7) The derivation of these values is as follows: Note that the reasoning is inspired from Fellner (80, p. 61) although our method and notation are different.

First let us establish the formula of the reaction curve of each duopolist:

$$p = a - bx, \text{ then, } p_1 = a - bx_2 - bx_1$$

$$TR_1 = ax_1 - bx_2x_1 - bx_1^2 \quad ; \quad MR_1 = a - bx_2 - 2bx_1$$

Firm No. 1 maximizes profits when $MR_1 = 0$ in the case of no cost, or when $MR_1 = C$ in the case of constant marginal costs.

Consequently in the case of no cost:

$$a - bx_2 - 2bx_1 = 0$$

where $x_1 = \frac{a - bx_2}{2b}$ and similarly for firm No. 2 we would find that $x_2 = \frac{a - bx_1}{2b}$ are the equations of the reaction curves.

Since $x = x_1 + x_2$

then $x = \frac{a - bx_2}{2b} + \frac{a - bx_1}{2b} = \frac{2a - b(x_1 + x_2)}{2b}$

or $x = \frac{2a - bx}{2b}$ or $2bx = 2a - bx$; $3bx = 2a$

Total equilibrium output is then $x = 2a/3b$

Since $x_1 = x_2$; thus $x_1 = x_2 = \frac{a}{3b}$

The ruling equilibrium price is obtained as follows

$p = a - bx$, then $p = a - b\left(\frac{2a}{3b}\right) = a/3$.

(AD,8) In the case of costs the reaction curves become:

$$x_1 = \frac{(a-c) - bx_2}{2b} \quad \text{and} \quad x_2 = \frac{(a-c) - bx_1}{2b}$$

Equilibrium aggregate output x is then:

$x = x_1 + x_2 = \frac{2(a-c) - bx}{2b}$. Solving, we obtain

$x = \frac{2(a-c)}{3b}$.

each duopolist will then supply $x_1 = x_2 = \frac{a-c}{3b}$

The ruling price, obtained by substitution is $p = \frac{a - c}{3}$

The individual demand curve for each duopolist will, when equilibrium is reached in the market, look like this:
(in the case of no cost)

$$p_1 = a - b\left(\frac{a}{3b}\right) - bx_1 = \frac{2a}{3} - bx_1$$

in the case of costs:

$$p_1 = \frac{2a + c}{3} - bx_1$$

We should note here that Fellner (80, p. 61) says that where there are costs, the letter a can be replaced by a - c; as seen in the last equation this is not always true, because the letter a is now replaced by a + c. However Fellner is right for the limited use he makes of this.

(AD, 9) These elasticity formulas are established as previously:

if $e_1 = \frac{p}{-MR + p} = \frac{2/3a - bx_1}{-(2/3a - 2bx_1) + (2/3a - bx_1)}$

$e_1 = \frac{2/3a - bx_1}{bx_1}$; similarly $e_2 = \frac{2/3a - bx_2}{bx_2}$

(AD,10) We know that equilibrium price is $p = a/3$

.... output for each duopolist is $x_1 = a/3b$. Replacing the value of x_1 in the elasticity formula we obtain $e_1 = \frac{2/3a - b\left(\frac{a}{3b}\right)}{b\ (a/3b)} = 1$.

(AD, 11) In the case of costs, equilibrium output is

$x = \frac{a-c}{3b}$, replacing, we obtain

$$e_1 = \frac{2/3a - b\left(\frac{a-c}{3b}\right)}{b\left(\frac{a-c}{3b}\right)} = \frac{a + c}{a - c}$$

(AD, 12) We know from (AD, 2) that $e = \frac{a - bx}{bx}$; since equilibrium value for $x = 2a/3b$ then $e = \frac{a-b\left(\frac{2a}{3b}\right)}{b\left(\frac{2a}{3b}\right)} = \frac{1}{2}$

This is the value of elasticity on the market demand curve when the system is in equilibrium.

(AD, 13) Aggregate output can be obtained as follows: The demand curve of firm No. 1 can be written as:

$$AR_1 = a - b(x-x_1) - bx_1$$

where $x = x_1 + x_2 + x_3 + \ldots\ldots x_m$ so that output put on the market (except for the one of firm No. 1) is $\underline{x - x_1}$. In terms of (AD,7), x_2 of (AD,7) is $x - x_1$ of the present footnote. Consequently, utilizing the material of (AD,7) we can write in the case of no cost the formula of reaction curve of firm No. 1 as $x_1 = \frac{a-b(x-x_1)}{2b}$ and similarly for firm No. 2, $x_2 = \frac{a-b(x-x_2)}{2b}$ etc. for firm No. n, whose $x_m = a - \frac{b(x-x_m)}{2b}$

Since $x = x_1 + x_2 + \ldots\ldots x_m$, then

$$x = \frac{na - b(xn - x)}{2b} = \frac{na}{(n + 1)b}$$

(AD, 14) To obtain equilibrium price we replace the value of x in the market demand curve:

$$p = a - bx$$

$$p = a - b\left[\frac{n.a}{(n+1)b}\right] - a - \frac{n.a}{n+1} = \frac{a}{n+1}$$

(AD, 15) The equilibrium output of each firm is obtained by dividing the aggregate output by the number of firms n (there is no reason why the firms should be of different size):

then, $\frac{n.a}{(n+1)b} \cdot \frac{1}{n} = \frac{a}{(n+1)b}$

the aggregate output of all the other firms $(x - x_1)$ is equal to:

$$x - x_1 = \frac{n.a}{(n+1)b} - \frac{a}{(n+1)b} = \frac{a(n-1)}{b(n+1)}$$

(AD,16) The equilibrium demand curve of firm No. 1 is:

$$p_1 = a - b(x - x_1) - bx_1$$

$$p_1 = a - b\left[\frac{a(n-1)}{a(n+1)}\right] - bx_1 = a - a\left[\frac{n-1}{n+1}\right] - bx_1$$

$$p_1 = \frac{2a}{n+1} - bx_1$$

(AD. 17) The MR_1 (equilibrium) is the derivative of TR_1 (equilibrium).

Since $TR_1 = x_1 p_1$ (where p_1 is the equilibrum one) consequently, $TR_1 = \frac{2ax_1}{n+1} - bx_2$; by derivation we obtain $MR_1 = \frac{2a}{n+1} - 2bx_1$.

(AD, 18) The elasticity of the equilibrium demand curve for each firm is e_1 is found as follows: (see AD,9 for the method):

$$e_1 = \frac{\frac{2a}{n+1} - bx_1}{-\left(\frac{2a}{n+1} - 2bx_1\right) + \left(\frac{2a}{n+1} - bx_1\right)} = \frac{2a - bx_1(n+1)}{bx_1(n+1)}$$

Knowing that the equilibrium output of firm No. 1 is $\frac{a}{(n+1)b}$ replacing this value in e_1, will yield the numerical value at which each firm will operate.

Here

$$e_1 = \frac{2a - b\left(\frac{a}{(n+1)b}\right)(n+1)}{b\left[\left(\frac{a}{(n+1)b}\right)(n+1)\right]} = 1$$

The elasticity at which the system will operate on the market demand curve is: (replacing x by its value $\frac{n \cdot a}{(n+1)b}$

$$e = \frac{a - b\left[\frac{na}{(n+1)b}\right]}{b\left(\frac{n \cdot a}{(n+1)b}\right)} = \frac{1}{n}$$

(AD. 19) In the case of costs, the aggregate equilibrium output is obtained as follows: (see AD, 13) while here the only difference is that $MR_1 = c$, (c = costs). In other words:

$a - b(x - x_1) - 2bx_1 = c$; then $x_1 = \frac{(a-c) - b(x - x_1)}{2b}$,

Similarly, $x_2 = \frac{(a-c) - b(x - x_2)}{2b}$, etc.....

Consequently $x_1 + x_2 + \ldots\ldots x_n = x = \frac{na - nc - b(nx - x)}{2b}$

$$x = \frac{n(a-c)}{(n+1)b}$$

(AD, 20) The equilibrium price becomes:

$$p = a - b\left[\frac{n(a-c)}{(n+1)b}\right] = \frac{a + nc}{n+1}$$

Note here that the method of Fellner which consists of replacing a by a - c does not work.

(AD, 21) The equilibrium output of each firm is:

$$\frac{x}{n} = \frac{n(a-c)}{(n+1)b} \cdot \frac{1}{n} = \frac{a-c}{(n+1)b}$$

Consequently the output of the rest of the firms $(x - x_1)$ is:

$$(x - x_1) = \frac{n(a-c)}{(n+1)b} - \frac{(a-c)}{(n+1)b} = \frac{(n-1)(a-c)}{(n+1)b}$$

(AD, 22) The equilibrium demand curve of each firm is equal to:

$$p_1 = a - b(x - x_1) - bx_1$$

$$= a - b\left[\frac{(n-1)(a-c)}{(n+1)b}\right] - bx_1, \text{ simplifying,}$$

$$p_1 = \frac{2a + c(n-1)}{(n+1)} - bx_1$$

(AD, 23) The marginal revenue curve is obtained by setting TR_1 as $= \frac{[2a + c(n-1)]}{(n+1)} x_1 - bx_1^2$, and deriving gives, $MR_1 = \frac{2a + c(n-1)}{(n+1)} - 2bx_1$

(AD, 24) The elasticity formula of the equilibrium demand curve of each firm is:

$$e_1 = \frac{\frac{2a + c(n-1)}{(n+1)} - bx_1}{-\left[\frac{2a + c(n-1)}{(n+1)} - 2bx_1\right] + \left[\frac{2a + c(n-1)}{(n+1)} - bx_1\right]}$$

$$= \frac{2a + c(n-1)}{bx_1(n+1)} - 1$$

in e_1, if n tends to infinity, then we know that x_1 tends toward 0; at the limit, n is infinity and e_1 tends toward infinity also.

As far as the elasticity of e of the market demand curve is concerned it can be obtained as follows:

We know that $e = \frac{a - bx}{bx}$ or $\frac{p}{bx}$. Knowing that as n increases toward infinity p tends toward c (see AD, 20). Consequently as n tends toward infinity e tends toward $\frac{c}{bx}$, which is neither 0 nor infinity but a stable positive value.

APPENDIX E

HOMOGENEOUS FUNCTIONS

by Dr. Carman Miller of the Department of Mathematics,
University of Saskatchewan.

(A) Derivation of Function No. 1. Homogeneous of degree one.

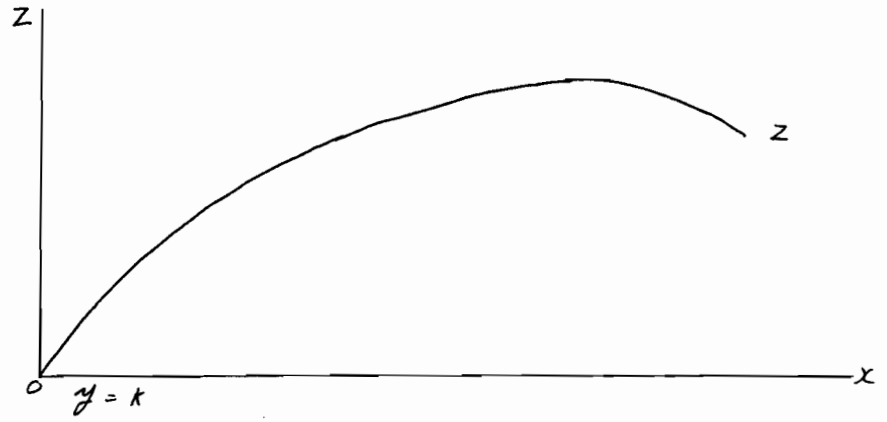
(B) Derivation of Function No. 2. Homogeneous of degree two.

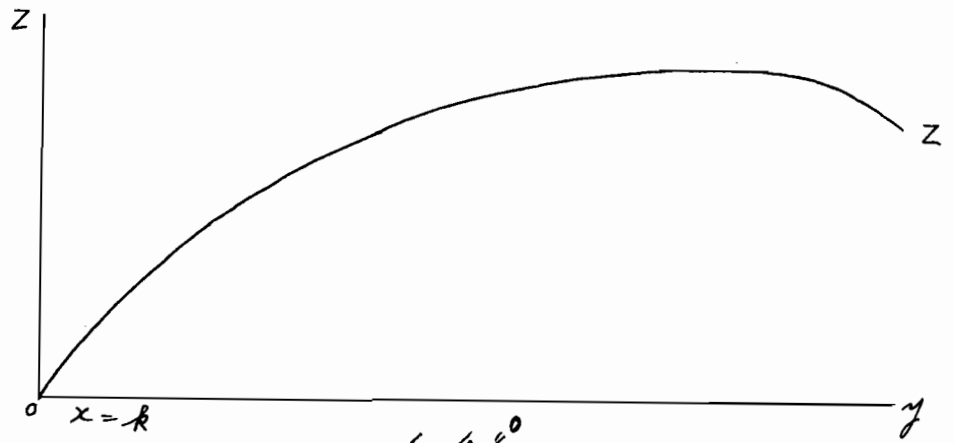
Function No. 1

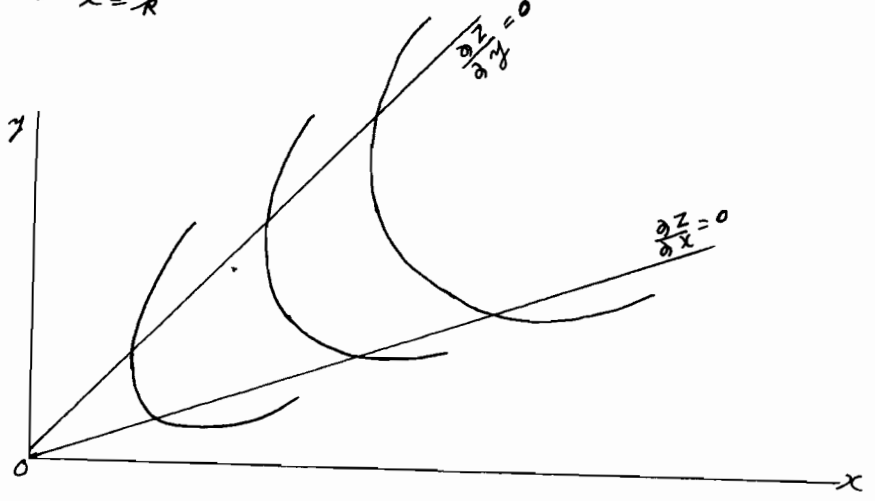
We seek a function $Z = Z(x,y)$ with the following properties:

1. $Z(x,y)$ is homogeneous of degree one so that $Z(tx,ty) = tZ$, and $x\frac{\partial Z}{\partial n} + y\frac{\partial Z}{\partial y} = Z$
2. $Z(x,y) = 0$ on the x - axis and on the y - axis.
3. $Z(x,y) > 0$ in the first quadrant.
4. The function $Z(x,k)$, k a positive constant, has exactly one maximum for $x > 0$.
5. The function $Z(k,y)$, k a positive constant, has exactly one maximum for $y > 0$.
6. The abcissas at which the family of functions $Z(x,k)$ have maxima all lie on a straight line through the origin in the xy - plane.
7. The abcissas at which the family of functions $Z(k,y)$ have maxima all lie on a straight line through the origin in the xy - plane.

8. The level curves Z (x,y) a constant cut these two lines (6 and 7) at points where the curve is parallel to the axes. These points on the level curves are where y has a minimum value with respect to x and where x has a minimum value with respect to y.



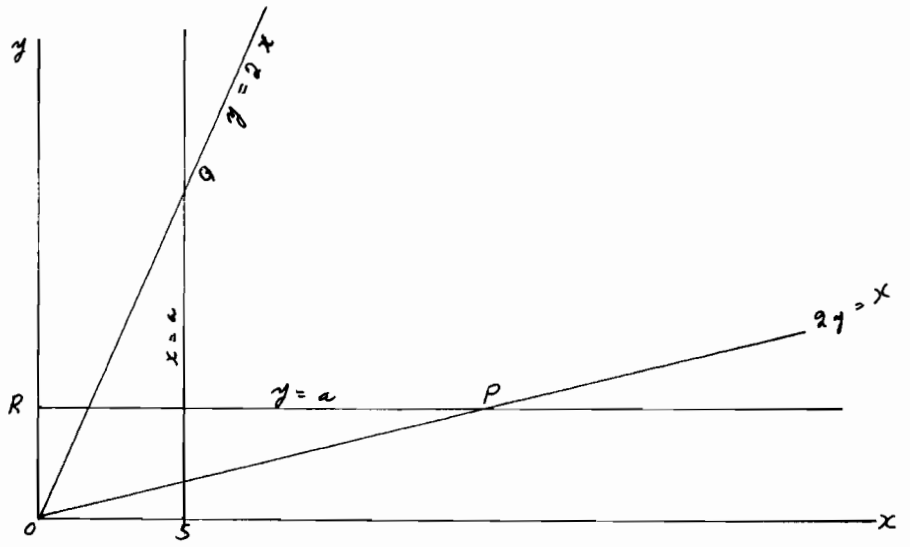




Various functions homogeneous of degree one where investigated before the following one with the desired properties was found, namely,

$$Z = \frac{x^2 y^2}{(x+y)^3}$$

For this function $\frac{\partial Z}{\partial x} = 0$ for $x = 2y$. The maximum value of Z on the line $y =$ a constant is $\frac{4}{27} y$ at the point $(2y, y)$. $\frac{\partial Z}{\partial y} = 0$ for $y = 2x$. The maximum value of Z on the line $x =$ a constant is $\frac{4}{27} x$ at the point $(x, 2x)$.



The maximum value of Z on the line RP is taken at the point P and is $\frac{4}{27} y$.

The maximum value of Z on the line SQ is taken at the point Q and is $\frac{4}{27} x$.

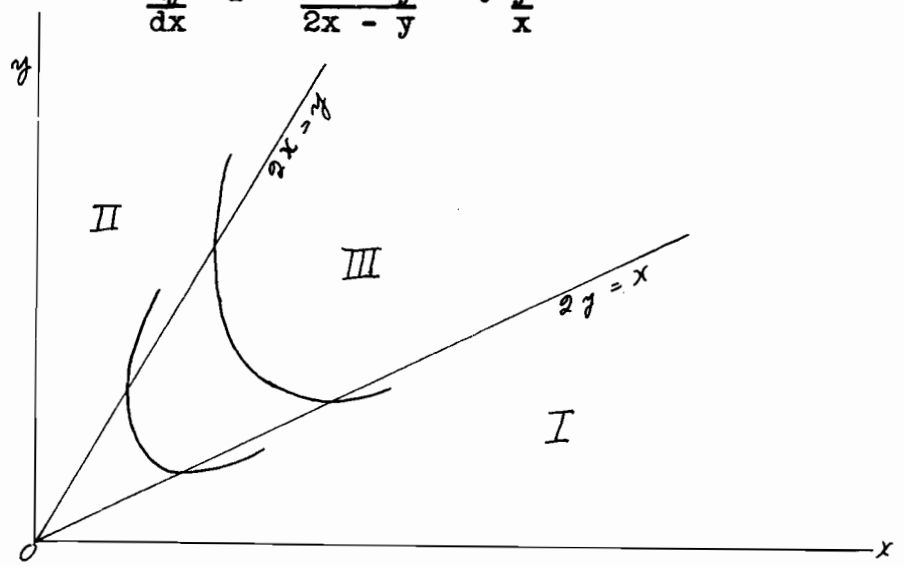
To find the nature of the level curves

$$\frac{x^2 y^2}{(x+y)^3} = \text{a constant}$$

we shall find the differential equation of this family of curves as follows. We have $2 \log x + 2 \log y - 3 \log (x + y)$ = a constant. Hence $\left(\frac{2}{x} - \frac{3}{x+y}\right) dx + \left(\frac{2}{y} - \frac{3}{x+y}\right) dy = 0$

Hence, on the family of level curves,

$$\frac{dy}{dx} = \frac{x - 2y}{2x - y} \cdot \frac{y}{x}$$



Now $x - 2y$ is positive in the region I and negative in the regions II and III, while $2x - y$ is positive in the regions I and III and negative in the region II. Thus, the curves are parallel to the x - axis on the line $2y = x$, parallel to the y - axis on the line $2x = y$, of positive slope in the regions I and II, and of negative slope in the region III. A few of the curves are sketched.

We note that this particular function

$$Z = \frac{x^2 y^2}{(x + y)^3}$$

has all the desired properties. Having demonstrated all

these properties we note something more important, namely, that we have found a method to construct the function.

Had we arbitrarily selected the two straight lines $2y = x$ and $2x = y$ we could have solved the differential equation

$$\frac{dy}{dx} = \frac{x - 2y}{2x - y} \cdot \frac{y}{x}$$

to get the family of curves

$$\frac{x^2y^2}{(x + y)^3} = c, \text{ a constant}$$

Then, the desired function would have been

$$Z = \frac{x^2y^2}{(x + y)^3}$$

This suggests that if we were to arbitrarily select the two straight lines $x = \alpha y$ and $Bx = y$ with $\alpha > 1$ and $\beta > 1$ then the solution of the differential equation

$$\frac{dy}{dx} = \frac{x - \alpha y}{\beta x - y} \cdot \frac{y}{x}, \quad \alpha > 1, \quad \beta > 0$$

would yield a whole family of functions with the desired properties.

The differential equation is homogeneous of degree one. The method of solution is standard. One makes a change of variable by the substitution

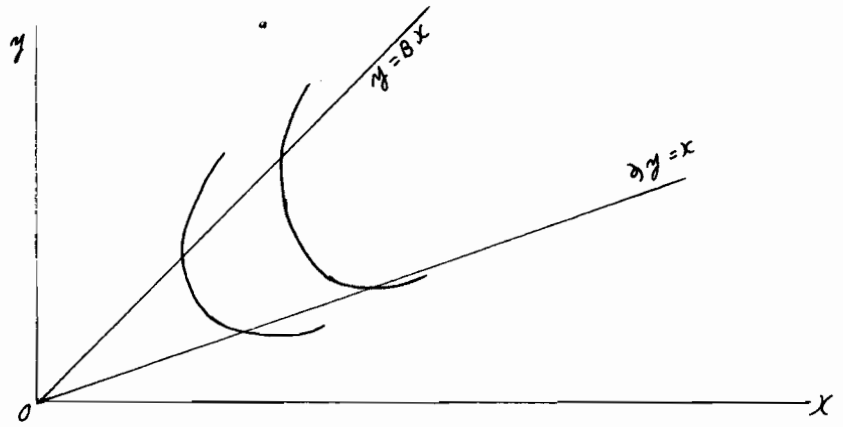
$$y = vx, \quad \frac{dy}{dx} = v + x\frac{dv}{dx}$$

which yields a first order differential equation with variables

x and v separable. The solution involves an arbitrary constant c. When one replaces c by Z there results the family of functions

$$Z = \frac{x^{\frac{\alpha}{\alpha-1}} \cdot y^{\frac{\beta}{\beta-1}}}{\left[(\beta-1)x + (\alpha-1)y\right]^{\frac{\alpha(\beta-1)}{(\alpha-1)(\beta-1)}}}$$

which are homogeneous of degree one and have all the other desired properties.



Function No. 2

We seek a function Z (x.y) with the following properties:

1. Z (x,y) is homogeneous of degree n so that

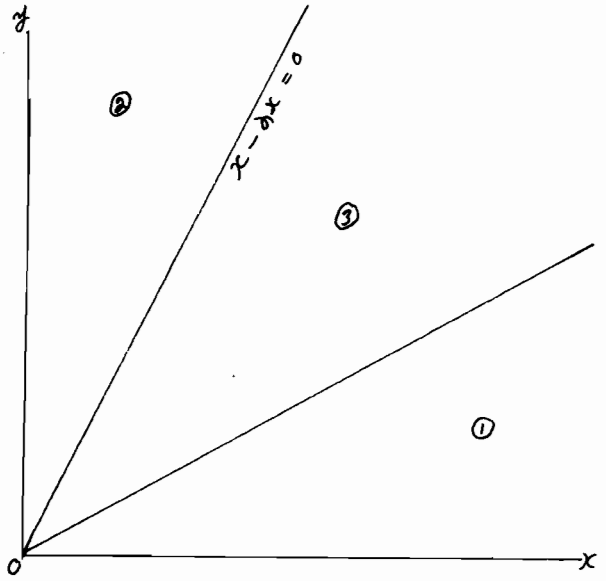
$$Z(tx, ty) = t^n Z(x,y)$$

and $$x\frac{\partial Z}{\partial x} + y\frac{\partial Z}{\partial y} = nZ$$

2. Z(x,y) = 0 on two straight lines

$$y - \alpha x = 0 \quad , \quad 0 < \alpha < 1$$
$$x - \beta y = 0 \quad , \quad 0 < \beta < 1$$

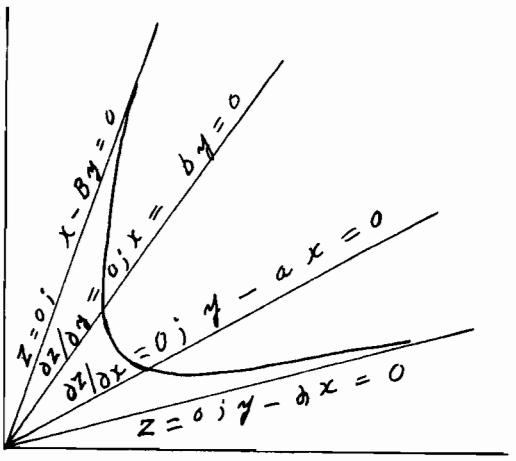
through the origin in the first quadrant.



3. $Z(x,y) > 0$ in the region (3) between these two lines; $Z(x,y) < 0$ in the regions (1) and (2) outside these lines.

4. The function Z(x,k), k a positive constant, has exactly one maximum which is in the region (3), and if x is the value of x which produces this maximum, the locus of the points (xm, k) is a straight line y - ax = 0, $a < a < 1$, through the origin in the xy-plane.

5. The function Z (k, y), k a positive constant, has exactly one maximum which is in the region (3), and if ym is the value of y which produces this maximum, the locus of the points (k, ym) is a straight line x - by = 0, $\beta < b < 1$, through the origin in the xy - plane.



6. Each level curved Z(x,y) = a constant cuts the lines y - ax = 0 and x - by = 0 at points where the curve is parallel to the axes; between these points the slope of the level curve is negative but otherwise the slope of the level curve is positive and

the level curve approaches the two lines

$y - \delta x = 0$ and $x - \beta y = 0$ asymptotically.

One way to arrive at such a function is to solve the differential equation,

$$\frac{dy}{dx} = -\frac{y - ax}{x - by}, \qquad \begin{matrix} 0 < a < 1 \\ 0 < b < 1 \end{matrix}$$

Another way is to start with the function

$$z = z(x, y) = \gamma (x - \beta y)(y - \delta x),$$

(γ a positive constant)

which is homogeneous of degree n = 2.

Both methods lead to the same result if

$$\delta = \frac{1 - \sqrt{1 - ab}}{b}, \qquad \beta = \frac{1 - \sqrt{1 - ab}}{a},$$

$$a = \frac{2\delta}{1 + \delta\beta}, \qquad b = \frac{2\beta}{1 + \delta\beta},$$

so that $a/b = \beta/\delta$.

$$\frac{\partial z}{\partial x} = [-2\delta x + (1 + \delta\beta) y]\gamma$$

$$\frac{\partial z}{\partial y} = [(1 + \delta\beta) x - 2\beta y]\gamma$$

Thus, $\frac{\partial z}{\partial x} = 0$ on the straight line y - ax = 0

and $\frac{\partial z}{\partial y} = 0$ on the straight line x - by = 0

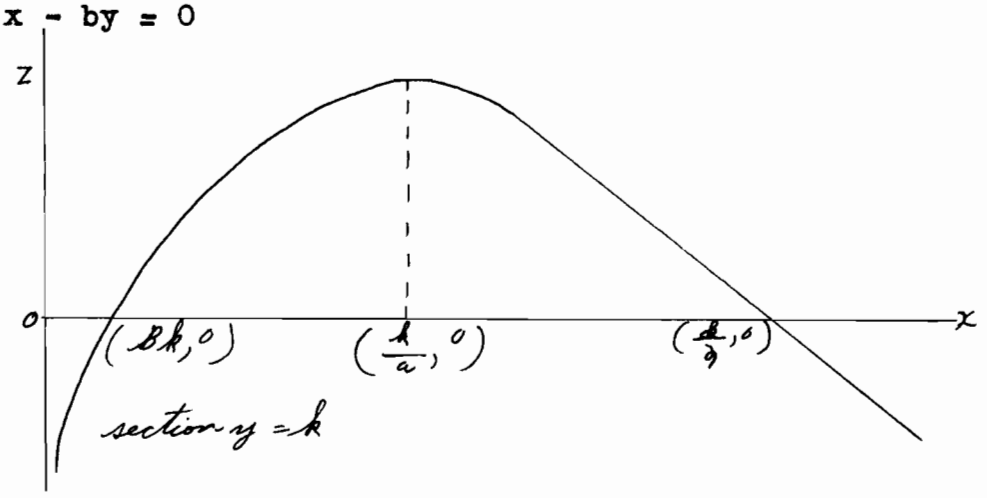
$$Z(x, k) = -\delta\gamma x^2 + (1 + \delta\beta)\gamma k x - \beta\gamma k^2$$

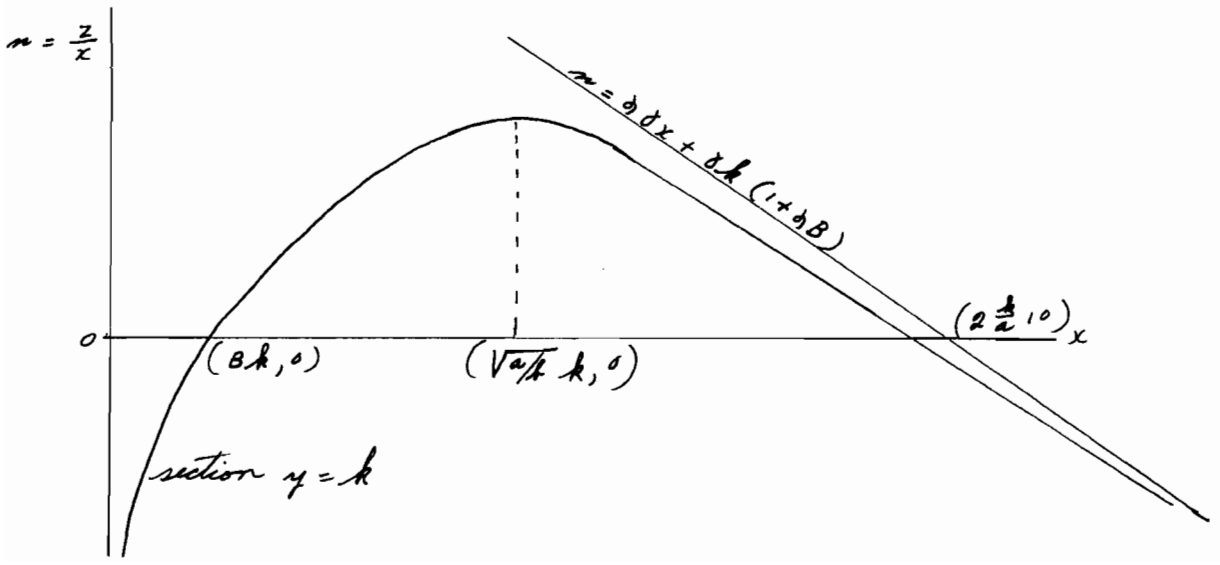
Thus, $\frac{\partial z(x, k)}{\partial x} = -2\delta\gamma x + (1 - \delta\beta)\gamma k,$

and $\dfrac{\partial^2 z(x,k)}{\partial x^2} = -2\alpha\gamma$

Thus, the sections of the surface Z = Z (x,y) by the planes y = k are curves which are concave downward with exactly one maximum at the point $x = \frac{k}{a}$ directly above the intersection of the plane y = k with the line y - ax = 0. Similarly, the sections of the surface Z = Z(x,y) by the planes x = k are curves which are concave downward with exactly one maximum at the point $y = \frac{k}{b}$ directly above the intersection of the plane x = k with the line

x - by = 0





Let us also study the functions

$$m(x,k) = \frac{Z(x,k)}{x} = -\alpha\gamma x + \gamma k(1+\alpha\beta) - \frac{\beta\gamma k^2}{x}$$

Thus, $\frac{\partial m(x,k)}{\partial x} = -\alpha\gamma + \frac{\beta\gamma k^2}{x^2}$,

and $\frac{\partial^2 m(x,k)}{\partial x^2} = -\frac{2\beta\gamma k^2}{x^3}$

The graphs of these functions are concave downward with exactly one maximum for $x > 0$ at the point where

$$x = \sqrt{\beta/\alpha}\cdot k = \sqrt{a/b}\cdot k$$

$m(x,k) = 0$ for $x = \beta k$ and for $x = k/\alpha$,

the same values for which $Z(x,k) = 0$. For small positive values of x the graph of the function approaches the negative m-axis asymptotically. For large values of x the graph of the function approaches the straight line $m(x,k) = -\alpha\gamma x + \gamma k(1+\alpha\beta)$ asymptotically.

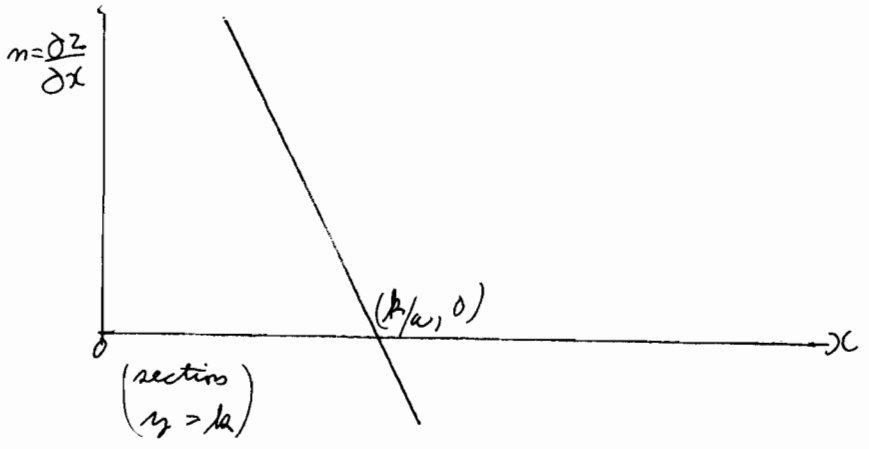
Let us also study the functions

$$N(x,k) = \frac{\partial^2 Z(x,k)}{\partial x} = -2\alpha\gamma x + \gamma k(1+\alpha\beta)$$

The graphs of these functions are straight lines of slope

x - intercepts $\frac{k}{a}$,

n - intercepts $\gamma k(1+\alpha\beta)$



APPENDIX F

CONDITIONS OF EQUILIBRIUM OF THE FIRM, THE SINGLE PRODUCT CASE

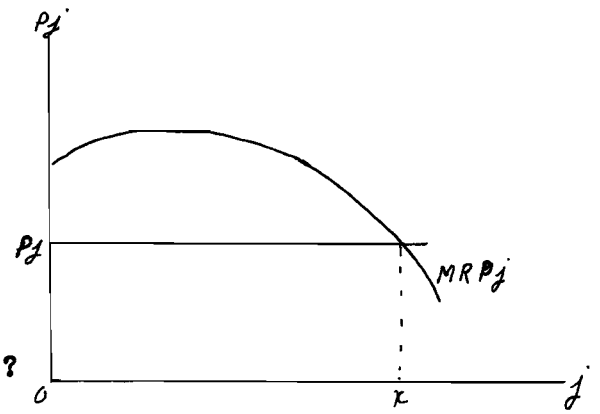
(i) Least cost condition : $P_j = MRP_j$

(ii) Maximization of Profits: MC = MR

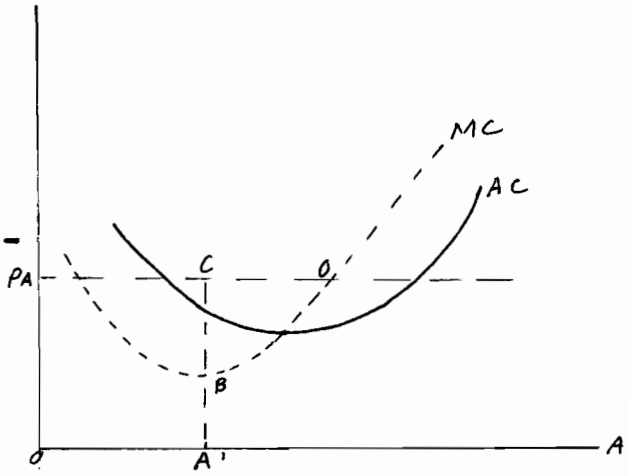
If $p_j < MRP_j$, then the entrepreneur will tend to hire more units of factor j; thus increasing output. However if input has to be increased because there is a possibility of gaining more profits by factor hiring, that means that the system was not previously at a point of maximum profits. Consequently if MC = MR, that means that $p_j = MRP_j$ hold. We can show now that if $p_j = MRP_j$ holds, then, MC must equal MR. This is done as follows.

The reasoning is a reductic ad absurdum:

Suppose we would write $p_j = MRP_j$ and that MC < MR. Under profit maximization hypothesis this is not a stable position because the manager (because of the second item) will try to increase production. Yet he can do that only by increasing the employment of some factor j. However this leads us to an absurdity because in terms of the diagram at right if $p_j = MRP_j$ holds, increasing employment of j will lead to losses. How can a profit be engendered by losses?



Since this cannot be so, our initial set of propositions cannot hold. Similarly we can prove that the set of propositions: $p_j = MRP_j$ and MC $>$ MR would lead to a similar absurdity. It remains to be shown that the least cost conditions (i) and profit maximization (ii) are two different conditions because one can be satisfied without the other being so. For instance condition (i) can be fulfilled without condition (ii) being so. Suppose that the firm stops producing at output A', where MR $>$ MC. Yet this is not an impediment preventing the firm from operating at the least cost point (the cheapest way to produce output A') because condition (i) to be satisfied requires only that we are on the expansion path, (that is of all the possible average cost curves, we are on the lowest one). Consequently being on the expansion path means that output A' is produced by the combination of x and y such that $p_x = MC_{A'} \cdot MP_x^{A'}$ and $p_y = MC_{A'} \cdot MP_y^{A'}$.



However condition (ii) cannot be satisfied except if condition (i) is also satisfied. See:
if condition (ii) is satisfied, MC = MR, and that we are in equilibrium in factor hiring: $p_x = MRP_x = MP_x \, MR$

$p_y = MRP_y = MP_y \, MR$

since condition (ii) postulates that MC = MR we can replace MR by MC, this yields:

$$p_x = MP_x MC = MC/MI_x \quad \text{or} \quad p_x MI_x = MC$$

$$p_y = MP_y MC = MC/MI_y \quad \text{or} \quad p_y MI_y = MC$$

or that $p_x MI_x = p_y MI_y$ (since they are both equal to the same thing MC) or that $p_x = MP_x$, which is the condition (i).

$$p_y = MP_y$$

If condition (i) would not be satisfied this could be only if a wrong MC is used to start with.

APPENDIX G

CONDITIONS OF EQUILIBRIUM OF THE FIRM; THE MULTIPRODUCT CASE

The general conditions of equilibrium (in terms of output and products produced and factors hired) of a general multiproduct long run model can be outlined as follows:

(1) The first condition is that any quantity of product (a, b, c...n) must be produced with the combinations of factors (x, y, ... m) such these combinations be the least cost ones.

This requires:

$$\frac{MP_x^a}{MP_y^a} = \frac{MRC_x}{MRC_y} \quad ; \quad \ldots\ldots \quad \frac{MP_y^a}{MP_m^a} = \frac{MRC_y}{MRC_m}$$

Similarly

$$\frac{MP_x^b}{MP_y^b} = \frac{MRC_x}{MRC_y} \quad ; \quad \ldots\ldots \quad \frac{MP_y^b}{MP_m^b} = \frac{MRC_y}{MRC_m}$$

....

etc.

$$\frac{MP_x^n}{MP_y^n} = \frac{MRC_x}{MRC_y} \quad ; \quad \ldots\ldots \quad \frac{MP_y^n}{MP_m^n} = \frac{MRC_y}{MRC_m}$$

From the above set of equations we can write:

$$\frac{MP_x^a}{MP_y^a} = \frac{MP_x^b}{MP_y^b} = \ldots\ldots\ldots \frac{MP_x^n}{MP_y^n} = \frac{MRC_x}{MRC_y}$$

$$\ldots = \ldots = \ldots\ldots\ldots = \ldots$$

$$\frac{MP_y^a}{MP_m^a} = \frac{MP_y^b}{MP_m^b} = \ldots\ldots\ldots \frac{MP_y^n}{MP_m^n} = \frac{MRC_y}{MRC_m}$$

The above set of equations can be rewritten as:

$$\frac{MI_z^a}{MI_x^y} = \frac{MI_z^b}{MI_x^b} = \ldots\ldots\ldots\ldots \frac{MI_z^n}{MI_x^n} = \frac{MRC_x}{MRC_z}$$

etc. ..... for any pairs of factors.

Now being at the least cost combinations (tangency of isoquant and isocost, the MC of any product will either be: MR $MRC_y . MI_y^a$ or MRC_x, MI_x^a . (for the reasoning see appendix F).

We can then write:

$$MRC_x . MI_x^a = MRC_z . MI_z^a = \ldots\ldots MRC_m . MI_m^a = MC^a$$

$$\ldots\ldots = \ldots\ldots = \ldots\ldots = \ldots\ldots$$

$$MRC_x . MI_x^n = MRC_z . MI_z^n = \ldots\ldots MRC_m . MI_m^n = MC^n$$

which can be rewritten as:

$$MRC_x = MC^a . MP_x^a = \ldots\ldots\ldots MC^n . MP_x^n$$

$$\ldots\ldots = \ldots\ldots = \ldots\ldots\ldots\ldots$$

$$MRC_m = MC^a . MP_m^a = \ldots\ldots\ldots MC^n . MP_m^n$$

or more shortly: (if i = any product a, b, c, ... n and if j = and factor x, y, ... m) the <u>least cost</u> conditions can be written as:

$$MRC_j = MC^i . MP_j^i$$

(2) The profit maximization condition now requires that an optimum output for each product is produced. This requires:

$$MC^a = MR^a \; ; \; MC^b = MR^b \; ; \ldots\ldots\ldots MC^n = MR^n$$

or shortly: $MC^i = MR^i$.

This implies that the rule of hiring factors is satisfied. (See previous reasoning appendix F). That is for each factor, the following conditions must hold, otherwise output will change:

$$MRC_x = MRP_x^a = \ldots\ldots\ldots\ldots MRP_x^n$$
$$\ldots\ldots \quad \ldots\ldots \qquad\qquad\qquad\qquad \ldots\ldots$$
$$MRC_m = MRP_m^a = \ldots\ldots\ldots\ldots MRP_m^n$$

or more shortly: $MRC_j = MRP_j^i$

A similar reasoning as in appendix F would show that equilibrium in factor hiring cannot be satisfied except if profits are maximized and vice versa.

Furthermore condition No. 2 implies condition No. 1:

$MRC_j = MRP_j^j = MP_j^j . MR^i$. Now since in the situation where profits are maximized $MR_i = MC_i$, we can substitute one for the other this gives: $MRC_j = MC^i . MP_j^i$, which is the least cost condition. However condition (1) can be satisfied without condition (2) being so because in this case condition (1) yields at a certain output $\underline{ii}$, $MRC_j = MC^{ii} . MP_j^{ii}$. Yet at that output it may happen (if profit maximization conditions is not satisfied) that $MC^{ii} = MR^{ii}$. Consequently being different we cannot substitute MR^{ii} for MC^{ii}, in the above equation to obtain optimum hiring of factors which is a formulation of condition (2).

BIBLIOGRAPHY

1. Abbott, L., "Vertical Equilibrium Under Pure Quality Competition," American Economic Review. Vol. 43, 1953, pp. 826.

2. _____, Quality and Competition. Columbia, N. Y. 1955.

3. Ackley, G., "Spatial Competition in A Discontinuous Market," Quarterly Journal of Economics, Vol. 1942. pp. 212-230.

4. Alchian, A. A., "Uncertainty, Evolution and Economic Theory," Journal of Political Economy. Vol. 58, 1950. pp. 211-222.

5. Archibald, G. C., "'Large' and 'Small' Numbers in the Theory of the Firm," Manchester School. Vol. 1959. pp. 104-109.

6. Alderson, W., "Survival and Adjustment in Organized Behavior System" in (8, 65-88).

7. _____, Marketing Behavior and Executive Action. Irwin, Homewood, 1957.

8. _____, and Cox, R., (eds) Theory in Marketing. Irwin, Homewood, 1950.

9. Allen, G. C., The Industrial Development of Birmingham and Black Country 1860-1927. London 1929. Cited by Stigler (205, pp. 192-193).

10. Allen, R. G. D., Mathematical Analysis for Economists. Macmillan, London, 1956.

11. _____, Mathematical Economics. London, Macmillan, 1956.

12. Alt, R. M., "The Internal Organization of the Firm and Price Formation: An Illustrative Case," Quarterly Journal of Economics. Vol. 63, 1949, pp. 92-110.

13. Bain, J. S., Barriers to New Competition. Harvard University Press. 1956.

14. Baumol, W. J., "Firms with Limited Money Capital," Kyklos, Vol. 6, 1953. p. 119.

15. Baumol, W. J., "Activity Analysis in One Lesson," American Economic Review, Vol. 48, 1958. p. 837.

16. _____, "On the Theory of Oligopoly," Economica, N.S., Vol. 25, 1958. pp. 187-199.

17. _____, Business Behavior, Value and Growth. MacMillan N. Y. 1959.

18. _____, Economic Theory and Operations Analysis. Prentice-Hall, Englewood, 1961.

18a. _____, "What Can Economic Theory Contribute to Managerial Economics," American Economic Review, May 1961. pp. 142-147.

19. Beach, E. F., Economic Models, Wiley, N.Y. 1957.

20. Bennion, E. G., Linear Programming and Game Theory. Bureau of Business and Economic Research. Michigan State University. East Lansing. 1960.

21. Biet, B., Théories Contemporaines du profit. Editions Genin. Librairie de Medicis. Paris 1956.

22. Bodenhorn, D., "A Note on the Theory of the Firm," Journal of Business (University of Chicago), Vol. 32, 1959. p. 164.

23. Boulding, K., Economic Analysis. Harper, N. Y., 1941.

24. _____, Economic Analysis Rev. Ed. Harper, N.Y. 1948.

25. _____, A Reconstruction of Economics. Wiley, N. Y. 1950.

26. _____, "Implications for General Economics of More Realistic Theories of the Firm," American Economic Review, May 1952, Vol. 42, p. 35.

27. _____, and Spivey W. A., Linear Programming and the Theory of the Firm. MacMillan & Co., N. Y., 1960.

28. Bowman, H. B., and Fetter, R. B., Analysis for Production Management. Rev. Ed. Irwin, Homewood, 1961.

29. Brems, Hans, "The Interdependence of Quality Variations, Selling Efforts and Price," Quarterly Journal of Economics, Vol. 62, 1948. pp. 418-440.

30. _____, Product Equilibrium Under Monopolistic Competition Harvard, 1951.

31. Brems, Hans, "Input-Output Coefficients as Measures of Product Quality," American Economic Review, Vol. 47 1957. p. 105.

32. Brown, H. G., "Opportunity Cost: Marshall's Criticisms of Jevons," American Economic Review, Vol. 21, 1931. pp. 498-500.

33. Cannan, E., Theories of Production and Distribution from 1776 to 1848. 2nd Ed., King and Son, London, 1903.

34. Carlson, Sune, A Study on the Pure Theory of Production. King and Son Ltd., London, 1939.

35. Cassel, J. M., "On the Law of Variable Proportions," in (79, pp. 103-119).

36. Chamberlin, E. H., The Theory of Monopolistic Competition. 6th edition. Harvard, 1950.

37. _____, "Monopolistic Competition Revisited," Economica N.S., Vol. 18, 1951. p. 343.

38. _____, "Full Cost and Monopolistic Competition," E.J., Vol. 62, 1952. p. 318.

39. _____, "The Product as an Economic Variable," Quarterly Journal of Economics, Vol. 67, 1953. p. 1.

40. _____, Towards a More General Theory of Value. Oxford University Press. N. Y. 1957.

41. _____, "On the Theory of Oligopoly," in (40, pp. 31-43).

42. Champerrowne,P.G., "Review of Brem's Product Equilibrium Under Monopolistic Competition," Economic Journal, Vol. 62, 1952. p. 353.

43. Clark, J. B., Distribution of Wealth. MacMillan, N.Y. 1899.

44. Cleland, S., "A Short Essay on a Managerial Theory of the Firm," in (27, pp. 202-217).

45. Clemence, R. V., Essays (Schumpeter). Ed. by R. V. Clemence. Addison-Wesley, N.Y. 1951.

46. _____, Ed. Readings in Economic Analysis. Vol. 2. Addison-Wesley Press Inc., Cambridge. 1950..

47. Coase, R. H., "The Nature of the Firm," in (207, pp. 331-352).

48. _____, "Monopoly Pricing and Interrelated Costs and Demands," Economica, N.S., Vol. 13, 1946. pp. 278-94.

48a. Cohen, K. J.: "Simulation of the Firms," American Economic Review, Proceedings, Vol. 50, 1960. p. 534.

49. Cole, A. H., "Entrepreneurship and Entrepreneurship History," in Change and the Entrepreneur. Harvard University Press, 1949. pp. 85-108.

50. _____, Business Enterprise in Its Social Setting. Cambridge, Harvard, University Press, 1959.

51. Cooper, W. W., "The Theory of the Firm, Some Suggestions for Revision," American Economic Review, Vol. 39, 1949. p. 1204.

52. _____, "A Proposal for Extending the Theory of the Firm," Quarterly Journal of Economics, Vol. 65, 1951. p. 87.

53. Coppock, J. D., Economics of the Business Firm. MacGraw-Hill Book Company, New York. 1959.

54. Cournot, A., Researches into the Mathematical Principles of the Theory of Wealth. MacMillan Company, N.Y., 1929.

55. Cyert, R. M., March, J. G., "Organizational Structure and Pricing Behavior in Oligopolistic Market," American Economic Review, Vol. 45, 1955. p. 129.

56. _____, _____, "Organizational Factors in the Theory of Oligopoly," Quarterly Journal of Economics, Vol. 70, 1956. p. 44.

57. _____, _____, "Introduction to Research on a Behavioral Theory of the Firm," Behavior Theory of the Firm Project, Working Paper No. 1. Grad. Sch. of Ind. Admini. Carnegie Inst. of Tech., Oct. 17, 1957.

58. Cyert, R. M., Simon, H. A., Trow, D.B., "Observation of Business Decision," Journal of Business, (University of Chicago), Vol. 29, 1956. p. 237.

59. Cyert, R. M., Feigenbaum, E. A., March, J. G., "Models in a Behavioral Theory of the Firm," Behavioural Science, Vol. 4, 1959. p. 81.

60. Danhof, C. H., "Observations on Entrepreneurship in Agriculture," in Change and the Entrepreneur, Harvard University, 1949.

61. Davenport, H. J., The Economics of Alfred Marshall. Oxford University Press, London. 1935.

62. Dean, Joel, Managerial Economics. Prentice Hall, N. Y. 1951.

63. Dehem, R., Traité d'Analyse Economique. Dunod, Paris. 1958.

64. Dorfman, R., Application of Linear Programming to the Theory of the Firm. B.B.R.R., University of California. 1951.

65. _____, "Mathematical or Linear Programming a Nonmathematical Exposition," American Economic Review, Vol. 43, 1953. pp. 797-825.

66. _____, and Steiner, "Optimal Advertising and Optimal Quality," American Economic Review, Vol. 44, 1954. p. 826.

67. _____, Samuelson, P. A., Solow, R. M., Linear Programming and Economic Analysis. McGraw-Hill, N.Y. 1958.

68. Downie, Jack, The Competitive Process. Duckworth & Co., London. 1958.

69. Due, J. F., and Clower, R. W., Intermediate Economic Analysis, 4th ed. Irwin, Homewood, 1961.

70. Early, J. S., "Marginal Policies of 'Excellently' Managed Companies," American Economic Review, Vol. 46, 1956, pp. 44-70.

71. _____, "Discussion of a Paper by H. A. Simon: Economics Organization Theory and Decision Making," American Economic Review, Proceedings May 1957, Vol. 47. pp. 330-335.

72. _____, "Foreword," to Firestone, F. S.: Marginal Aspects of Management Practices.

73. Edgeworth, F. Y., Papers Relating to Political Economy, Vol. 1. MacMillan and Co., London. 1925.

74. Enke, S., "Space and Value," Quarterly Journal of Economics, Vol. 56, 1942. pp. 627-37.

75. _____, "Equilibrium Among Spatially Separated Markets: A Solution by Electric Analogue," Econometrica, Vol. 19, 1951. p. 40.

76. Evans, G. H., "A Century of Entrepreneurship in the United States," in The Entrepreneur, Research Center in Entrepreneurial History, Harvard. Cambridge 1957.

77. Farrell, M. J., "Deductive Systems and Empirical Generalizations in the Theory of the Firm," Oxford Economic Papers, Vol. 4, 1952. p. 45.

78. _____, "Discussion," American Economic Review Proceedings. May 1960, Vol. 50. pp. 561-564.

79. Fellner, W., and Haley, B. F., (eds.), Reading in the Theory of Income Distribution. Blakiston. Philadelphia, 1946.

80. _____, Competition Among the Few. Knoff, N.Y. 1949.

81. _____, Modern Economic Analysis. McGraw-Hill, N. Y. 1960.

82. Fetter, F., "The Economic Law of Market Areas," Quarterly Journal of Economics, Vol. 38, May 1924. pp. 520-9.

83. Firestone, F. N., Marginal Aspects of Management Practices. M.S.S. Business Studies. 1960.

84. Firth, R., Primitive Polynesian Economy. Routledge & Sons, London 1939.

85. _____, Primite Economics of the New Zealand Maori. Dutton & Co., N.Y. 1929.

86. Friedman, M., Essays in Positive Economics. Chicago Press. 1953.

87. Frisch, R., "Monopoly-Polypoly- The Concept of Force in the Economy," in International Economic Papers No. 1. Macmillan. London, 1951.

88. _____, "Marshall's Theory of Value," Quarterly Journal of Economics. 1950. Vol. 64, p. 495.

89. Gabor, A., and Pearce, I. F., "A New Approach to the Theory of the Firm," Oxford Economic Papers Vol. 4, 1952. p. 252.

90. _____, _____, "The Place of Money Capital on the Theory of Production," Quarterly Journal of Economics, Vol. 72, 1958. p. 537.

90a. Gide, C. et Rist, C., Histoire des Doctrines Economiques. 7th Ed. Paris. 1947.

91. Granick, D. , Management of the Industrial Firm in the U.S.S.R. Columbia, N. Y. 1954.

92. Greenhut, M. L., Plant Location in Theory and in Practice. Chapel Hill University of North Carolina Press. 1956.

93. Guitton, Henri, "Revue de Plant Location Theory and Practice, (Greehut). Revue Economique. Sept. 1958.

94. Hague, D. C., "Alfred Marshall and the Competitive Firm," Economic Journal. Vol. 68, 1958. p. 673.

95. Hahn, F. H., "The Theory of Selling Costs," Economic Journal, Vol. 69, 1959. p. 293.

96. Haley, B., ed., A Survey of Contemporary Economics. Vol. 11, Irwin Chicago. 1952.

97. Henderson, J. M., and Quandt, R. E., Microeconomic Theory. McGraw-Hill. N. Y. 1958.

98. Haney, L. H., History of Economic Thought. MacMillan, N. Y. 1949.

99. Hicks, J. R., The Theory of Wages. Macmillan, London. 1932.

100. _____, "Marginal Productivity and the Principle of Variation," Economica O.S., Vol. 12, 1932. pp. 79-88.

101. _____, "Linear Theory," Economic Journal, Vol. 70, 1960. p. 671.

102. Hieser, R., "A Kinked Demand Curve for Monopolistic Competition," Economic Record, Vol. 39, 1953. p. 19.

103. Hitch, C., and McKean, R., "Suboptimization in Operations Problems," in Operations Research for Management. Ed. McCloskey and Trefethen. John Hopkins Press Baltimore. 1954.

104. _____, _____, "What Can Managerial Economics Contribute to Economic Theory," American Economic Review, Proceedings. Vol. 51, May 1961. p. 147.

105. Hoggatt, A. C., "An Experimental Business Game," Behavioral Science, Vol. 4, 1959. pp. 192-204.

106. Hood, W. C., "Linear Programming and the Firms," Canadian Journal of Economics and Political Science. Vol. 18, 1952. p. 208.

107. Hotelling, H., "Stability and Competition" in (207, pp. 467-485).

108. Hoover, E. M., Location Theory and the Shoe and Leather Industry. Harvard, Cambridge Mass. 1937.

109. _____, "Spatial Price Discrimination," Review of Economic Studies. Vol. 4, June 1937. p. 182.

110. Hutchison, T. W., A Review of Economic Doctrines, 1870-1929. Oxford Clarendon Press. 1953.

111. Hyson, C. D., and Hyson, W. P., "The Economic Law of Market Areas," Quarterly Journal of Economics. Vol. 64, 1950. pp. 350-357.

112. Isard, W., Location and Space-Economy, Wiley, N.Y. 1956.

113. Jevons, S., The Theory of Political Economy. Macmillan, London. 1911.

114. Kaplan, A., Dirlam, J., Lanzillotti, R. F., Pricing in Big Business. The Brookings Institution, Washington, D.C. 1958.

115. Kaysen, C., "Review of Strategy and Market Structures," American Economic Review, Vol. 50, 1960. p. 1036.

116. Keller, I. W., Management Accounting for Profit Control. McGraw-Hill, 1957.

116a. Klein, L. R., Economic Fluctuations in the United States 1921-1941. Wiley, N. Y., 1950.

117. Knight, F., Risk, Uncertainty and Profit. Houghton Mifflin. N.Y. 1921.

118. _____, On the History and Method of Economics. University of Chicago Press. Chicago. 1956.

119. Koopmans, T. C., Three Essays on the State of Economic Science. McGraw-Hill, N.Y. 1957.

120. Lanzillotti, R. F., "Pricing Objectives of Large Companies," American Economic Review, Vol. 48, 1958. pp. 921-941.

121. Lee, W. W., Economic Fluctuations: Growth and Stability. Irwin, Homewood, 1959.

122. Leftwitch, R. H., The Price System and Resource Allocation. Rev. Ed. Holt, Rinehart and Winston, N.Y. 1960.

122a. Leibenstein: Economic Theory and Organizational Analysis. Harper, N.Y. 1960.

123. Lekatchman, R. , A History of Economic Ideas. Harper. N.Y. 1959.

124. Lerner, A. P., Economics of Control. MacMillan, N.Y. 1944.

125. Lester, R. H., "Shortcomings of Marginal Analysis for Wage-Employment Problems," American Economic Review, Vol. 36, 1946. pp. 63-82.

126. Liebhafsky,H.H., "A Curious Case of Neglect. Marshall's Industry and Trade," Canadian Journal of Economics and Political Science. Vol. 21, 1955. p. 339.

127. Losch, A., Economics of Location, Yale New Haven, 1954.

128. Luce, R. R., and Raiffa, H., Games and Decisions. Wiley, N.Y., 1957.

129. Machlup, F., "On the Meaning of Marginal Product," in (79, pp. 158-175).

130. _____, "Comments on a Boulding's Paper: Implications for General Economics of More Realistic Theory of the Firm," American Economic Review, Vol. 42, May 1952. p. 69.

131. _____, The Economics of Sellers' Competition. John Hopkins, Baltimore, 1952.

132. Makower, H., Activity Analysis. Macmillan, London, 1957.

133. Margolis, J., "The Analysis of the Firm: Rationalism, Conventionalism, and Behaviourism," Journal of Business. (University of Chicago) Vol. 31, 1958. p. 187.

134. _____, "Traditional and Revisionist Theories of the Firm: A Comment," Journal of Business (University of Chicago), Vol. 32, 1959. p. 178.

135. Marsh, D. B., World Trade and Investment. Harcourt, Brace and Co. N.Y. 1951.

136. Marshall, A., Principles of Economics. 8th Ed. Macmillan, London, 1952.

137. _____, Industry and Trade. Macmillan, London. 1920.

138. Mason, E. S., "Price and Production Policies of Large-Scale Enterprise," American Economic Review. Supplement Vol. 29, 1939. p. 61.

139. _____, "Comment" in A Survey of Contemporary Economics (Haley). p. 221.

140. Maxwell, J. A., "Some Marshallian Concepts, Especially the Representative Firm," Economic Journal, Vol. 68, 1958. p. 691.

141. McCloskey and Trefethen , Operations Research for Management. John Hopkins Press, Baltimore, 1954.

142. Mehren, G. L., "The Theory of the Firm in Marketing," in (8).

143. Menger, K., "The Laws of Return," in Economic Activity Analysis. O. Morgenstern (ed.) N.Y. Wiley, 1954. pp. 419-582.

144. Meyer, J. R., and Kuh, E., The Investment Decision. Harvard, 1957.

145. Mill, J. S., Principles of Political Economy. Ed. W. F. Ashley. Longmans, Green.

146. Mishan, E. J., "Theories of Consumer's Behaviour: A Cynical View," Economica, N.S., Vol. 28 No. 109, 1961. pp. 1-12.

147. Modigliani, F., "New Developments on the Oligopoly Front," Journal of Political Economy, Vol. 66, 1958. p. 215.

147a. _____, "Comments," American Economic Review, Vol. 51, May 1961. p. 157.

148. Neuner, J., Cost Accounting. Rev. Ed. Irwin, Chicago. 1942.

149. Newman, J. W., Motivation Research and Marketing Management. Cambridge, Harvard University Press, 1957.

150. Neumann, Von, and Morgenstern, O., Theory of Games and Economic Behavior, Princeton, N. J. 1944.

151. Nogaro, B., Dévelopement de la Pensée Economique.

152. Operations Research Seminar Montreal 1958. Proceedings Operations Research Seminar, Ed. by Leetham, Simpson Ltd., Montreal.

153. Orcutt, G. H., "Simulation of Economic Systems," American Economic Review, Vol. 50, 1960. pp. 893-908.

154. Ozga, S. A., "Imperfect Markets Through Lack of Knowledge," Quarterly Journal of Economics, Vol. 74, 1960. p. 29.

155. Paakkanen, J., Hintateorian Realismi. (In Finnish). English Summary: "The Realism of Price Theory and the Price Decisions of the Firm."

156. Papandreou, A. G., "Some Basic Problems in the Theory of the Firm," in (96).

157. Pareto, V., Manuel d'Economie Politique. 2nd Ed. Giord et Frères, Paris, 1927.

158. Penrose, E., "Biological Analogies in the Theory of the Firm," American Economic Review, Vol. 42, 1952. p. 804.

159. _____, "Limits to the Growth and Size of Firms," American Economic Review, Vol. 45, May 1955. p. 531.

160. _____, The Theory of Growth of the Firm. Wiley, N.Y., 1959.

161. Peston, M., "A Review of Games and Decisions," (Luce and Raiffa), Economica N.S., Vol. 27, 1960. p. 185.

162. _____, "Returns to Scale," Oxford Economic Papers, Vol. 12, 1960. p. 133.

163. Pigou, A. C., Economics of Welfare, 4th Ed. Macmillan, London. 1932.

164. _____, Memorials of Alfred Marshall. Macmillan. London, 1925.

165. _____, "Empty Economic Boxes," in (207, pp. 131-143).

166. Pirou, G., Les Théories de L'Equilibre Economique. Domat-Montchrestien, Paris, 1938.

167. Polanyi, K., Arensberg, C., Pearson, H., Trade and Markets in the Early Empires. The Free Press, Glencoe, Illinois. 1957.

168. Ponsard, C., Economie et Espace. Sedes. Paris, 1955.

169. Reder, M., "A Reconsideration of the Marginal Productivity Theory," Journal of Political Economy, Oct. 1947, pp. 450-458.

170. Robertson, D. H., "Wages Grumbles," in (79, pp. 227-237).

171. _____, Economic Commentaries. Staple Press London, 1956.

172. Robbins, Lionel, "Remarks Upon Certain Aspects of the Theory of Costs," Economic Journal, Vol. 44, 1934. pp. 1-19.

173. Robbins, Lionel, An Essay on the Nature and Significance of Economic Science. 2nd Ed. London, Macmillan. 1935.

174. _____, Robert Torrens and the Evolution of Classical Economics. N. Y. St. Martin's Press. 1958.

175. _____, "The Representative Firm," Economic Journal, Vol. 38, 1928. p. 387.

176. Robinson, E. A. G., The Structure of Competitive Industry Nisbet, Cambridge. 1958.

177. Robinson, Joan, "Euler's Theory and the Problem of Distribution," Economic Journal, Vol. 44, 1934, p. 398.

178. _____, Economics of Imperfect Competition. Macmillan, London. 1948.

179. Rolph, E., "The Discounted Marginal Productivity Doctrine," in (79, pp. 278-294).

180. Rothschild, K. W., "Price Theory and Oligopoly," in (207, pp. 440-467).

181. Ryan, W., Price Theory. Macmillan, London. 1958.

182. Samuelson, P. A., Economics. 4th ed. McGraw-Hill N.Y.

183. Schumpeter, J. A., The Theory of Economic Development. Cambridge Mass. Harvard University Press. 1934.

184. _____, "Economic Theory and Entrepreneurial History," in Clemence (45).

185. _____, History of Economic Analysis. Oxford University Press, N.Y. 1954.

186. Scitovsky, T., "Some Consequences of the Habit of Judging Quality by Price," Review of Economic Studies, Vol. 12, 1944-45. p. 100.

187. _____, Welfare and Competition. Irwin, Chicago, 1951.

188. Schneider, E., Pricing and Equilibrium. London, W. Hodge. 1952.

189. Schoeffler, S., The Failures of Economics. Cambridge, Harvard University Press, 1955.

190. Shubik, M., "The Role of Game Theory in Economics," Kyklos, Vol. 6, 1953. p. 21.

191. _____, "Economics, Management Science, and Operations Research," Review of Economics and Statistics, Vol. 40, 1958. pp. 214-220.

192. _____, "Simulation of the Industry and the Firm," American Economic Review, Vol. 50, 1960. p. 908.

193. _____, "Approaches to the Study of Decision-Making Relevant to the Firm," Journal of Business (University of Chicago), Vol. 34, 1961. p. 101.

194. _____, Strategy and Market Structure. Wiley, N. Y. 1959.

195. Simon, H. A., "Recent Advances in Organizational Theory," in Research Frontiers in Politics and Government by S. H. Bailey and others, Brookings Lectures, 1955.

196. _____, "Theories of Decision-Making in Economics and Behavioral Science," American Economic Review, Vol. 49, 1959. p. 253.

197. _____, and Bonini, C. P., "The Size Distribution of Business Firms," American Economic Review, Vol. 48, 1958. p. 607.

198. Smith, A., Wealth of Nations. Methuen & Co., London. 1920.

199. Soper, C. S., "Review of Brems' Product Equilibrium under Monopolistic Competition," Economic Record, Vol. 28, 1952, pp. 105-108.

200. Spencer, M., and Siegelman, L., Managerial Economics. Irwin, Homewood, 1959.

201. Stäkelberg, H., The Theory of the Market Economy. Hodge & Co., London. 1952.

202. Stigler, G. J., Production and Distribution Theories. MacMillan N.Y. 1941.

203. _____, The Theory of Price. MacMillan, N. Y. 1946.

204. _____, The Theory of Price. Rev. Ed. MacMillan, N. Y. 1952.

205. _____, "The Division of Labor is Limited by the Extent of the Market," Journal of Political Economy, Vol. 1951.

206. _____, "The Kinky Oligopoly Demand Curve and Rigid Prices," in Stigler (203) pp. 410-440.

207. _____, and Boulding, K. E., (eds.) Readings in Price Theory. Allen and Unwin, London. 1956.

208. Stonier, A. and Hague, D., A Textbook of Economic Theory. Longmans Green & Co. N.Y.

209. Strauss, I. H., "The Entrepreneur: The Firm," Journal of Political Economy, Vol. 52, 1944. p. 112.

210. Suajanen, W. W., "Enterprise Theory and Corporate Balance Sheets," Accounting Review, Vol. 32, 1958. p. 56.

211. Swartzman, D., "The Methodology of the Theory of Returns to Scale," Oxford Economic Papers, Vol. 10, 1958. p. 98.

212. Thirly, G. H., "The Subjective Theory of Value and Accounting Cost," Economica, N.S. Vol. 13, 1946. pp. 32-49.

213. Tintner, G., "The Theory of Choice under Subjective Risk and Uncertainty," Economica, IX 1941. pp. 298-304.

214. Triffin, R., Monopolistic Competition and General Equilibrium Theory. Cambridge, Harvard University Press. 1940.

215. Van Sickle, C. L., Cost Accounting 2nd Ed. Harper and Brothers, 1947.

216. Viner, J., "Costs," in Encyclopedia of Social Sciences, MacMillan, N.Y. 1935. pp. 466-475.

217. _____, "Studies in the Theory of International Trade. Harper Brothers, N.Y. 1937.

218. _____, "Cost and Supply Curves," in Clemence (46).

219. Wagner, H. M., "Recent Advances in Game Theory," American Economic Review, Vol. 48, 1958. p. 368.

220. Walras, Léon, Elements of Pure Economics. Homewood Irwin, 1954.

221. Ward, B., "The Firm in Illyria: Market Syndicalism," American Economic Review, Vol. 48, 1958. pp. 556-590.

222. Weintraub, S., Price Theory. Pitman Pub. Co., N.Y. 1949.

223. Whitin, T. M., The Theory of Inventory Management. Princeton, N.Y. 1957.

224. Whitin, T. M.,"Managerial Economics and the Firm," American Economic Review, Proceedings, Vol. 50, May 1960. p. 549.

225. Whittaker, E., History of Economic Ideas. Longmans, Green and Co., N.Y. 1940.

226. Wicksell, Knut, Lectures on Political Economy. Routledge and Sons, London. 1935.

227. _____, Selected Papers in Economic Theory. Allen and Unwin, London. 1958.

228. Wicksteed, P. H., Commonsense of Political Economy. Macmillan, London. 1910.

229. Wolfe, J.N, "The Representative Firm," Economic Journal. Vol. 64, 1954. p. 337.

230. Zeuthen, F., Economic Theory and Method. Longmans, Green and Co., London. 1957.