UNIVERSAL
LIBRARY

OU_156221

UNIVERSAL
LIBRARY

OUP—24—4-4-69—5,000

OSMANIA UNIVERSITY LIBRARY

Call No. 330/N52T Accession No. P.G. 18096

Author Nevins, E.

Title Text-book of economic analys 1958.

This book should be returned on or before the date last marked below.

TEXTBOOK OF ECONOMIC ANALYSIS

TEXTBOOK OF ECONOMIC ANALYSIS

BY

EDWARD NEVIN
M.A. (Wales), PH.D. (Cantab.)
Lecturer in Economics,
University College of Wales, Aberystwyth

LONDON
MACMILLAN & CO LTD
NEW YORK · ST MARTIN'S PRESS
1958

Copyright © by Edward Nevin 1958

MACMILLAN AND COMPANY LIMITED
London Bombay Calcutta Madras Melbourne

THE MACMILLAN COMPANY OF CANADA LIMITED
Toronto

ST MARTIN'S PRESS INC
New York

PRINTED IN GREAT BRITAIN

FOREWORD

THE first need in writing a text-book is obviously a clear appreciation of the needs of the type of student for whom it is intended; the second is a rigorous shaping of the design of the book to meet those needs. In all elementary studies there is a danger that in attempting to cater for more than one group of readers an author may end by failing to provide adequately for any of them; this danger is especially great in a subject such as economics which is approached by an unusual variety of readers for a variety of purposes. This book is intended primarily for the student preparing for the examination in economics of the General Certificate of Education at the advanced level. As a result, it is severely academic in its general approach, and has as its aim that of supplying (in the words of the syllabus of one examining authority) 'an introduction to elementary economic theory . . . providing candidates with a firm and precise understanding of fundamental concepts . . . and their uses, rather than a vague and general understanding of analytical concepts and tools appropriate to advanced analysis.'

In its general outline, therefore, the book follows orthodox, Marshallian lines, and eschews the 'current problems' type of approach which many believe to be more suitable for students whose aim is to be able to follow contemporary economic issues in an intelligent way. Such an approach is certainly more interesting for the newcomer to the subject. Quite apart from the requirements of an examination syllabus, however (and these have exercised a guiding influence in the writing of this book), I for one remain unconvinced that there is any satisfactory substitute for the traditional approach to economic theory if an understanding of the basic principles at work in general economic affairs—as opposed to specific *ad hoc* issues of a particular moment—are to be fully understood.

Primarily in order to confine the book within tolerable limits of space, the descriptive matter contained in it has been reduced to

the minimum consistent with the requirements of the advanced level syllabus. This has the secondary advantages that teachers can clothe the theoretical bones with whatever descriptive flesh they consider most suited to the special needs of their students, and also that (economic theory having a somewhat slower rate of obsolescence than factual data) the book will date less rapidly than one containing a great deal of descriptive material.

Two other points are worth mentioning here. First, at the end of each chapter I have indicated works in which the subject-matter treated in the chapter is discussed more fully and deeply. The choice of references has inevitably been restricted by the fact that there is little point in referring students to other text-books of an elementary nature (which in general cover much the same ground and differ only in methods of presentation), and also by the limited access which readers of this book are likely to have to the various learned journals in which are immured many of the most valuable discussions of contemporary economic theory. Hence, the 'further reading' suggested is occasionally of a rather advanced nature and is suitable for teachers rather than the students themselves; in the great majority of instances, however, the latter could safely be encouraged to consult these references and would undoubtedly benefit immensely from them. As a precaution, the occasional advanced works have been marked with an asterisk.

Secondly, although most of the theoretical work covered by this book lies over ground on which there is almost complete agreement in contemporary economic theory, on certain issues—and especially the two topics of the rate of interest and employment theory—any kind of *communis opinio doctorum* is conspicuous by its absence. One procedure for meeting this situation is to attempt to present a kind of Cook's tour of all the conflicting doctrines, with the hope that the unfortunate student will be able to assess for himself the relative merits of the different theories and appreciate the inter-relationships existing between them. In my experience, and, more important, in that of others whose experience in these matters is far greater, such a procedure almost invariably leads to a state of complete and utter confusion in the mind of the student who (if he is to find this type of book of any use) is simply not at a sufficiently advanced stage of comprehension of the subject to be

reasonably expected to appreciate what is involved in these complex and esoteric controversies. In this book, therefore, I have quite deliberately employed a rigorously Keynesian presentation (i.e. in the *General Theory* sense) of the theory of the rate of interest and of issues involving the overall level of incomes, prices and employment. In the fullness of time, the student who pursues the study of economics will learn of the criticisms and alternatives which have been offered in connection with these matters, but I am firmly of the view that it is better for a student to master one approach (and, if he so chooses, to subsequently unlearn it when he is in a better position to grasp what is involved) rather than to be faced in his theoretical infancy with a whole generation of deep and bitter controversy under which he will almost certainly submerge in despair.

Finally, may I record here my gratitude to the various people who were kind enough to sacrifice some of their valuable time in order to read through and comment on the earlier drafts of the book, especially my colleagues Professor Arthur Beacham, Mr. J. H. Morris, Mr. P. E. Hart and Mr. L. J. Williams, and my fellow-researchers into the affairs of the Welsh economy, Miss Gwen Jones, Miss E. M. Alexander, Mr. Gareth Richards and Mr. G. H. Peters. To ask highly-trained and busy people to read right through an elementary text is an order of no mean tallness; the cheerfulness and efficiency with which the imposition was discharged constituted an exercise of great charity on their part. Needless to say, I must claim as entirely my own work any errors and all the imperfections which remain.

CONTENTS

CHAPTER | PAGE

FOREWORD v

PART I: INTRODUCTION

1. THE NATURE AND SCOPE OF ECONOMICS
   1. Scarcity and choice 3
   2. Defining economics 5
   3. Methods of economic analysis 7
   4. The objectives of economic analysis 10
   5. Relation to other sciences 12
2. SOME BASIC CONCEPTS
   1. Utility 14
   2. Wealth and welfare 15
   3. Capital and income 16
   4. Factors of production 18
   5. The margin 19
   6. Equilibrium 21
   7. Economic laws 22

PART II: THE THEORY OF DEMAND

3. DEMAND AND UTILITY
   1. Total and marginal utility 27
   2. The law of diminishing utility 30
   3. The concept of equi-marginal returns 31
   4. Demand curves 34
   5. Perverse demand curves 38
   6. Inter-relationships of demand 40
4. INDIFFERENCE CURVES
   1. Preference versus utility 43
   2. The analysis of price effects 50
   3. The analysis of income effects 51

CHAPTER PAGE

5. ELASTICITY OF DEMAND
   1. The concept of elasticity 54
   2. Factors governing elasticity of demand 57
   3. The significance of elasticity 61

PART III: THE THEORY OF PRODUCTION

6. LABOUR
   1. Introductory 67
   2. Division of labour 68
   3. Other factors in labour efficiency 70
   4. The size of the population 71
   5. The population of the United Kingdom 76
   6. The mobility of labour 80

7. LAND
   1. The supply of land 83
   2. The law of diminishing returns 85
   3. Sites and industrial location 89

8. CAPITAL
   1. Types and forms of capital 94
   2. Gross and net output 97
   3. Capital formation and consumption 99

9. ENTERPRISE
   1. The functions of the entrepreneur 104
   2. Optimum factor combination 106
   3. Large-scale output—increasing returns 107
   4. Small-scale output 111
   5. Industrial organisation 113
   6. The industrial structure of the United Kingdom 116

10. THE OPERATION OF SUPPLY
   1. Cost curves 124
   2. Average and marginal cost 128
   3. Supply curves 131
   4. The significance of time in supply 136
   5. Joint supply 141
   6. Opportunity cost 142

PART IV: THE THEORY OF PRICE

CHAPTER PAGE

11. THE NATURE OF MARKET EQUILIBRIUM
    1. Prices, commodities and markets 147
    2. Industry and firm: the demand curve 148
    3. Industry and firm: equilibrium conditions 150

12. PRICE UNDER PERFECT COMPETITION
    1. What is perfect competition? 154
    2. The demand curve under perfect competition 155
    3. Market equilibrium under perfect competition 156

13. PRICE UNDER MONOPOLY
    1. The nature of monopoly 160
    2. Sources of monopoly power 161
    3. Monopoly equilibrium 162
    4. Monopoly and public policy 166

14. PRICE UNDER IMPERFECT COMPETITION
    1. The nature of imperfect competition 171
    2. Market equilibrium 173
    3. The significance of imperfectly competitive equilibrium 176

15. THE PRICE MECHANISM
    1. The allocation of resources 180
    2. Optimum allocation 183
    3. Changes in supply and demand 186
    4. Taxation and rationing 189

PART V: THE THEORY OF INCOME DISTRIBUTION

16. WAGES
    1. The demand for labour 195
    2. The supply of labour 198
    3. The determination of wages under competition 200
    4. Relative wage-levels 202
    5. Time rates and piece rates 205
    6. Real wages and money wages 207
    7. The role of Trade Unions 209
    8. State intervention in wage-determination in Britain 211

CHAPTER PAGE

17. RENT
1. Rent as a payment for land 214
2. Rent as a surplus 216
3. Quasi-rent 219

18. PROFITS
1. Profits as an accounting surplus 221
2. Profits as a cost 222
3. The bearing of uncertainty 223
4. Profits and the allocation of resources 225
5. Monopoly profits 227

19. THE NATIONAL INCOME
1. Measuring the national income 230
2. National income accounts 235
3. National income and welfare 238
4. The distribution of the national income 240

PART VI: MONEY AND BANKING

20. THE NATURE OF MONEY AND BANKING
1. Why have money? 251
2. Requirements for an efficient money 254
3. The nature of bank credit 256
4. Bank credit as money 260

21. THE BRITISH BANKING SYSTEM
1. Historical development 263
2. The Bank of England 265
3. The British commercial banks 270

22. THE MONEY AND CAPITAL MARKETS
1. The Bill of Exchange 276
2. The nature and functions of the money market 279
3. Long-term securities 282
4. The new issue market 284
5. The Stock Exchange 285

23. THE THEORY OF THE RATE OF INTEREST
1. The nature of liquidity preference 289
2. Short-term and long-term rates 295
3. Other risks and uncertainties 297

CHAPTER PAGE

24. THE MECHANISM OF MONETARY POLICY
1. The weapons of the Central Bank 299
2. Central Bank powers in action 302
3. The effectiveness of monetary policy 304
4. Selective credit controls 307

PART VII: THE THEORY OF EMPLOYMENT

25. SAVINGS AND INVESTMENT
1. Aggregate demand and employment 313
2. The components of aggregate demand 315
3. Consumption behaviour 317
4. The determinants of investment 319
5. Savings and investment 322

26. THE PROCESS OF INCOME FORMATION
1. The nature of the income flow 325
2. The multiplier 328
3. Investment fluctuations and income 331
4. Income and the propensity to save 334

27. FLUCTUATIONS IN PRICES AND EMPLOYMENT
1. Inflation and deflation 337
2. The historical trade cycle 340
3. The problems of full employment 343
4. The budget as a stabiliser 347

28. INDEX NUMBERS
1. The problem of measurement 352
2. The nature of index numbers 353
3. The retail price index 355
4. Other index numbers 359

PART VIII: INTERNATIONAL ECONOMICS

29. THE SIGNIFICANCE OF INTERNATIONAL TRADE
1. The law of comparative costs 365
2. The terms of trade 369
3. International trade and employment 373
4. Free trade and protection 375

CHAPTER PAGE

30. FOREIGN EXCHANGE

1. The mechanism of foreign payments 379
2. The role of exchange rates 382
3. Fixed versus variable exchange rates 384
4. Official intervention in foreign exchange 389
5. International monetary institutions 392

31. THE BALANCE OF PAYMENTS

1. The balance of payments on current account 396
2. International capital movements 397
3. The balance of payments of the United Kingdom 400
4. Devaluation, deflation and import controls 402

32. BRITAIN IN THE WORLD ECONOMY

1. The inter-war years 406
2. The post-war problems 407
3. The Sterling Area 411

INDEX 416

PART I

INTRODUCTION

CHAPTER I

The Nature and Scope of Economics

1. SCARCITY AND CHOICE

The world in which we live is a desperately poor place. Two and a half thousand million human beings crowd its surface, each one needing food and shelter in order to do nothing more than preserve his own life. But parts of the world, it may be said, are not poor; in America the family garage has two cars and indoors the refrigerator contains ice-cream and chicken. Even in Europe the general standard of living is not bad; few people die of starvation or exposure, and even television sets are not uncommon. The sad truth is partly that Europe and North America are not the world, or even the greater part of the world, and that for the majority of the world's population human life exists on the border between bare subsistence and extinction. It is also true that the relatively rich communities such as Europe and North America contain enormous pockets of real poverty, and the world is a long way yet from the day when they will be finally removed. Even if all the good things of the world were placed in a gigantic common pool, and shared equally amongst all the human race, therefore, the world's population would still have scarcely enough food and clothing and warmth and shelter to keep itself alive.

This is not all. If a situation could be conceived of in which every adult in the world was able to produce as much material wealth as the most highly skilled operative in the most modern American factory—and the situation, of course, is scarcely conceivable for many years to come—it would still be found that the world was dissatisfied with its living standards. This is a familiar characteristic of human beings—their perpetual discontent with the material wealth they have and their incessant desire for more, or for better. When I was a boy, getting (usually) a penny a week for sweets and

the rest, I used to envy the wealthier boys whose allowance ran to twopence or even sixpence a week. If only I had as much as them, I used to think, I would be absolutely content. But it never worked out like that; when the time came and I could rely on sixpence a week, I found myself consumed with envy of the people who counted their expenditure in shillings and whose life contained all sorts of things which were still beyond my grasp. And so it has been all throughout history and with the whole human race. As the fundamental needs of life have been met, and a more or less secure physical existence has been achieved, the realisation of the possibilities of an easier and fuller material life has dawned and new wants—as real and as pressing as those which have been satisfied—have arisen. The lowest-paid worker in Britain today enjoys a standard of living which would have seemed incredibly high to a moderately wealthy man a couple of centuries ago; yet the worker is probably conscious only of the things he lacks in comparison with his neighbours. Even the much-envied prosperous American family with its two cars and deep-freeze is nowhere within sight of gratifying all its existing wants. By the time all those wants have been satisfied, new ones, as yet not conceived but soon to acquire enormous strength, will surely have arisen.

These two basic facts—that in its material wealth the world is very poor even by our present standards and that the material needs and desires of human beings appear to be capable of infinite growth and development—explain the existence of the science of economics. Since the goods and services which the human race is capable of producing are strictly limited, the problem must arise of choosing which wants are to be given priority and which are to be left unsatisfied. This problem—the *economic problem* as it is called—arises only because of the insufficiency of the world's wealth in relation to its needs. The way in which priorities are allocated, or the way in which the choice between the satisfaction of want A and want B is made, is of the essence of economics. If ever mankind found itself able to produce more material wealth than it knew what to do with—which, unfortunately, it never will—then the necessity for choosing between the satisfaction of one want and that of another would have disappeared; and with it would disappear the science of economics.

2. DEFINING ECONOMICS

Scarcity, then, is the foundation of economics. Can economics therefore be defined as the study of the way in which man attempts to overcome the problem of scarcity? To do so would make the scope of economics unnecessarily wide. To begin with, economics is a *social* science—that is, it is concerned with human activities which in some way influence other members of a community as well as the person undertaking them. For example, time is always a scarce factor: a man will always have to choose between various ways of using his time. But if he is choosing between spending a particular Saturday afternoon in bed or on a long walk, his decision will have virtually no importance for other members of society and is therefore outside the interests of the economist. On the other hand, if the choice is between an afternoon in bed and an afternoon at work then society *is* interested, because the total amount of wealth produced, and therefore the general standard of living, is going to be influenced by that decision. Consequently, the economist is directly interested in the factors which enter into that decision of whether to sleep or to work.

In general, this means that a choice or decision is within the scope of economics only when it involves some commodity or service which can be *exchanged.* The man deciding between bed and a walk cannot pass on to anyone else the benefit he derives from either; it is entirely personal and subjective. But he *can* pass to someone else the result of an afternoon's work in a factory—he can sell the goods he produces or the labour which helps to produce them. Economics is primarily concerned with exchange, including the processes which lie on either side of it—the production of the thing exchanged and the consumption of the thing exchanged. Anything which affects these processes is within the scope of economics.

Even with this reservation, a definition of economics which links it with the overcoming of scarcity as such remains too wide. The quantity of food which can be produced with a given amount of labour and land, for example, is not solely an economic question. It involves technological issues such as the application of chemistry to soil fertility, or mechanical engineering to agricultural processes.

On the other side, the consumption of food will be influenced by such factors as racial characteristics, climate and physiology. An economist obviously cannot be expected to be a master of all the sciences and technologies which have a bearing on even the simplest acts and occurrences of everyday human life. His interest in and special knowledge of all these things extends only as far as they influence, or are influenced by, the actual process of exchange. The chemical properties underlying a particular fertiliser are quite outside his ken, and, as an economist, he has no special interest in them. But he *has* an interest, and *is* concerned, if the application of that fertiliser, whatever its physical properties may or may not be, results in certain types of foodstuff losing some of their exchange value through increased abundance, or if some development in the market for chemicals, or labour, or land, stimulates or discourages the application of the fertiliser concerned.

Exchange is vital to economics, then, not only because it indicates that a particular human activity has a social, as well as an individual, significance, but also because the economist is competent to analyse only those aspects of individual and social activity which in some direct way influence or are influenced by the process of exchange. Not all acts of exchange are of economic significance, however; the economist can say no more than anyone else about the complex motives and emotions which make a boy decide for or against swopping a particular conker for another conker. The object being exchanged must be something generally accepted as scarce in relation to the need which society feels for it, and fortunately conkers do not come into this category. The exchange must be based on scarcity; and the overcoming of the scarcity must have a direct bearing on the process of exchange.

It follows that economics studies those acts and aspects of exchange which arise from efforts to overcome that scarcity of the material requisites of well-being which exists, and always will exist, on this earth. It is sometimes said that economics is concerned only with transactions which are, or can be, expressed in terms of money. As a simple and straightforward working definition this is sound enough, since in modern economic life money is used for almost all processes of economic exchange. But it is not completely true; for example, an economist would be concerned with ex-

changes carried out by means of barter. Or again, the fact that some members of the community sacrifice some of their limited time to take part in local government unquestionably has economic significance, since local government authorities have economic functions. But local government representatives are not paid for their work and therefore their allocation of time to it is not regarded as an economic activity. The reason is that human sentiments in themselves are incapable of measurement or precise assessment; it is possible to add up only sums of money paid from one person to another, or to compare one sum of money with another. So long as exchange takes place through the medium of money there is some measuring-rod, imperfect though it may be, against which to judge the relative importance of the various goods and services involved. But services which are not paid for cannot be valued in any objective way whatever, and the economist leaves them alone for that reason, not because they are inherently uninteresting to him. If a domestic servant receives a certain sum of money each week in return for her services, the economist can at least assume that the value of those services is roughly measured by the weekly wage. But who can assess the value of the housewife's services to her family in the home? The economist, at least, is too sensible to try.

3. METHODS OF ECONOMIC ANALYSIS

It has been seen that the economist sets out to analyse and explain those acts of exchange which are directly connected with the attempt to overcome scarcity of goods and services, and, in practice, with those exchanges which are carried out through the medium of money. This raises two questions. *How* does he proceed to this analysis? *Why* does he proceed to it? The latter question will be discussed in the following section: for the moment only the former need be considered.

Unlike the physical scientist, the economist cannot carry out controlled experiments under laboratory conditions, where all factors except the one under examination are prevented from operating, so that conclusions can be arrived at with a fair degree of certainty. The subject-matter of economics, as of all social sciences, is man—man, as Marshall puts it, 'in the ordinary business of life' —and man carrying out the ordinary business of life under labora-

tory conditions is a contradiction in terms. This method of analysis being denied him, the economist has recourse to two other methods —the deductive method and the inductive method.

By the *deductive* method is meant the process of logical deduction from a given proposition; the method is characteristic of, for example, mathematics. A hypothesis is taken as being true, and a process of reasoning is applied to it so as to bring out all the implications which it contained but which were not immediately obvious. For example, if it is postulated that the producers of goods will always act so as to obtain the greatest possible money profit, certain conclusions can then be derived from that postulate (they are in fact implicit in it) concerning the number of goods which they will sell in a given situation and the price at which they will sell them. Or again, it can be postulated that the more of a commodity a person consumes the less enjoyment he derives from each additional unit of it; from this hypothesis can be derived certain conclusions about the nature of demand.

The bulk of the theorising in economics is based on this type of procedure. It should be clearly understood, however, that the truth of any conclusions arrived at in this way will depend on three things; first, the correctness of the process of logical deduction, second the validity of the original hypothesis, and third the degree to which man himself acts logically in the context involved. The first need not cause much concern. If I say 'A is true, therefore it follows that B will happen', and my reasoning is in fact faulty, clearly my conclusion will be wrong. However, as a science establishes itself, by constant discussion and development errors of this kind are eventually discovered and put right. The basic theory set out in this book is, for the most part, the product of the analysis and criticism of many generations of scholars, and elementary logical errors are scarcely likely to have survived this process.

The second two limitations on the method are more serious. The problem of deciding upon a hypothesis having general validity is a difficult one for a science concerning itself with human nature in all its infinite complexity. The hypothesis that producers always seek to obtain greatest profit, for example, is falsified every day by producers who act in the interests of their workpeople, or society, or according to religious or political beliefs, rather than in the pursuit

of profit. And the conclusions drawn from a hypothesis can be true only to the extent that the hypothesis is sound.

Similarly, assuming that a sound hypothesis is found, conclusions drawn from it by rigorous logic will be true of the real world only if mankind also draws the logical inference from a given fact and acts upon it. A worker might accept as true the hypothesis that he would prefer to work for a high wage rather than a low wage, but a deduction that he will therefore move from factory X, where he is paid £10, to factory Y, where he could get £12, might be falsified in the event because the man fails to act rationally—he may instead be influenced by instinct or laziness and so on. How many of us as consumers are perfectly rational in our purchases? Where is the person who has never bought something 'on the spur of the moment', i.e. quite irrationally?

Because of these limitations to the usefulness of abstract reasoning, economists constantly have recourse to a second method of analysis, known as the *inductive* or *empirical* method. Here the appeal is to facts, rather than reasoning, and an attempt is made to arrive at conclusions from the known facts of actual life. Economic historians, for example, seek to find in the record of history explanations of important economic occurrences and developments. Statisticians collect data about production, prices, wages and employment, and the figures are analysed to find significant trends or inter-relationships. Inductive work of this kind can in itself give rise to generalisations: the collection of data concerning the household expenditure of various classes in the community, for instance, might indicate some relationship between income, or social class, and spending habits which could never have been arrived at by logic alone.

It is in combination with the deductive method, however, that this type of analysis is most useful. By it one can check a theory arrived at through logical deduction against the experience of reality, and so expose unrealistic hypotheses or faulty reasoning. Empirical analysis can never *prove* a theory, since the world is so complex that the data never admit of only one explanation; facts seldom 'speak for themselves' and never with one voice. But a theory can be shown to be *inconsistent* with the facts, and therefore wrong. On the other hand, deductive analysis must come before

inductive, since in empirical work a choice has always to be made from an infinite number of facts and figures, and this choice necessarily presupposes some theorising. If I am investigating the demand for eggs over the past ten years, and examine data concerning the price of eggs, the production of eggs and the price of bacon, then I have implicitly postulated a theory relating the demand for eggs to those factors, and I am now testing the degree to which the facts confirm that relationship. The two types of analysis are therefore interdependent, each being equally valuable and necessary.

One other distinction needs to be made clear at this stage. Contemporary economic analysis is usually classified into either *micro-economics* or *macro-economics*. The former is the branch of theory which concerns itself with individual units or particular sectors of the economy—such as the demand for a particular product or its relative value, or the output of particular firms or industries. Macro-economics, on the other hand, considers the economic system as a whole and the relationship to the whole and to each other of broad aggregates—such as the general price level, the rate of investment and changes in overall industrial production —and concerns itself with individual groups of producers or consumers only insofar as they influence the economy in its entirety. Broadly speaking, Parts II–V of this book are devoted to micro-economics, while Part VII is an introduction to macro-economics.

4. THE OBJECTIVES OF ECONOMIC ANALYSIS

Turn now to the second question: what is the point of the study of economics? Some subjects are worth study for their own sakes—because they are intellectually satisfying or because they help us to attain desirable physical or mental qualities. There are reputed to be mathematicians, for example, who find tears coming into their eyes at the mere mention of the binomial theorem, so overcome are they at the sheer beauty of the concept. However, though parts of economic theory have some virtue as pure intellectual exercises, few people would devote to its mastery for this reason the considerable effort necessary.

Economics is primarily a utilitarian subject, in the sense that its value lies mainly in the help it can give towards understanding the forces at work in the economic life of society. Those who possess

such understanding are better able—whether as professional economists, politicians, administrators, or as citizens and voters—to formulate intelligent and effective economic policy, and to appreciate the inter-relationship between the activities of workers or consumers and those of the rest of society.

To take the point a little further, what is it which is being sought in economic policy—that is, in annual budgets, monetary and banking measures, investment plans and wages policy? Put simply, economic policy (public or private, and whether standing on its own or merely part of broader policies) has three aims. First and most fundamental is the pursuit of *efficiency* in its widest sense. Economics, it must be remembered, is the study of the ways and means of overcoming scarcity, of making the best use of limited and inadequate resources. The first job of the economist is therefore to seek out the ways in which society's limited resources may be best used, given the framework of the social and economic structure and the priority attached to various needs.

Secondly, economic policy tries to increase the *stability* of economic life. In the past, countries such as Britain have suffered periodic slumps, involving widespread unemployment and idleness of productive capacity, with all the material and moral evils which follow from these things. To prevent the recurrence of mass unemployment is a clear objective of modern economic policy, but it calls for the understanding of the way in which the complex economic system works and the obscure causes of its periodic collapses in the past.

Finally, modern economic policy is inevitably concerned with *equity* so far as the material necessities and comforts of life are concerned. There is substantial agreement nowadays that pronounced inequality of wealth is socially and economically undesirable, and that the community as a whole has an obligation towards its weaker members who might suffer severe hardship if there was no special provision to prevent it. The task of economic policy in this respect is to find how this aim may best be realised, and how it may be reconciled with the other objects of policy—by no means always an easy task.

5. RELATION TO OTHER SCIENCES

From what has been said about the nature of economics and the aims of economic policy it will be obvious that the economist must be perpetually bumping up against the territories of other special studies and disciplines. For example, government economic policy (in the formulation of which the student of economics tries to join, or at least seeks to understand) is frequently enforced through legislative measures, such as the budget or control over local government expenditure. Inevitably, then, economics becomes mixed up with *politics*, but the economist must take the political mechanisms of his society as something outside his sphere of expertise. The task of understanding and improving the political machinery through which economic policy operates must be left to the political theorist.

More important, perhaps, is the fact that economic measures or problems may have political significance—that is, they may involve decisions on matters falling within the proper sphere of the government. For example, the economic situation may be such as to require an increase in total government revenue. Suppose that this could be obtained by raising the rate of income tax on either high incomes alone or on relatively small incomes. The economist, as such, cannot make the choice between these alternatives, since it raises the question of which section of the community should bear the additional burden, and this is a matter for the government to decide and answer for to the electorate. All that the economist can do in cases of this kind is to present the alternative courses and his assessment of their various effects; the final decision, being political, must be left to the government and therefore, ultimately, to the community as a whole. Decisions of a political nature, involving basic issues of fairness or justice or morality are called *value judgements*, and the economist is no more fitted to make them than anyone else. His concern is with means, not ends; the ends, as decided upon by society as a whole through the processes of representative government, he must take as given.

Similarly, economic affairs are frequently involved with *ethics*, and the economist cannot claim any special right to decide ethical issues. He is not concerned, as an economist, with the rightness or

wrongness of given actions; he is concerned only with their economic effects. As an individual, of course, he is as much concerned with the morality of human actions as anyone else, but he cannot pronounce on such matters with any special authority. Again, as has already been noted, economic activities such as consumption and production may well be bound up with almost every special study in existence, from psychology and physiology to chemistry and nuclear physics. Naturally the economist does not presume to trespass into these specialised fields of knowledge. He will not seek to enquire *why* people consume tea in one country and coffee in another, or *why* a combine harvester will harvest a given acreage in a fraction of the time taken by a horse-drawn reaper. He leaves to other specialists the task of explaining why these things are so; he takes them as facts of the situation just as much as climatic conditions or the inherent qualities of the race. He is concerned only with the interaction between these things and the activities of production and exchange; even so limited, he will find that his field of study is complex and far-ranging.

FURTHER READING

A. MARSHALL, *Principles of Economics* (8th ed., Macmillan, London, 1920), Bk. I, Chs. I–II.

* J. N. KEYNES, *The Scope and Method of Political Economy* (4th ed., Macmillan, London, 1930), Chs. I–IV.

A. C. PIGOU, *The Economics of Welfare* (3rd ed., Macmillan, London, 1950), Ch. I.

SIR H. HENDERSON, *Supply and Demand* (Cambridge University Press, London, 1932), Ch. I.

* L. ROBBINS, *An Essay on the Nature and Significance of Economic Science* (Macmillan, London, 1932), Chs. I–II.

CHAPTER 2

Some Basic Concepts

1. UTILITY

Words are frequently misleading, sometimes downright dangerous. The greatest confusion in their use usually arises from a failure to make clear the precise sense in which words are being used. Partly for this reason, and partly to avoid breaking an argument later on in order to explain terms having a special usage in economics, it is best to explain at this stage what exactly is meant by a few concepts which will be used constantly in this first journey into economic analysis.

Before any commodity can enter into the economic sphere it must be scarce—meaning by that *scarce in relation to our desire for it*. Earthquakes are rare, but they are not scarce phenomena in the sense that people would like to have more of them. The element of desiredness necessary for an economic transaction arises from the fact that commodities or services convey *utility*; the consumer derives from them some pleasure or satisfaction or benefit. It is perhaps unfortunate that this word 'utility' has been adopted in economics to stand for the satisfaction derived from goods and services by their consumers. In ordinary speech, the word 'utility' is taken (quite properly) to mean 'usefulness'; it is not long since Britain had 'utility' furniture and clothing, the term 'utility' then being used to connote serviceability and robustness of construction. The word utility is not used in this limited sense of 'usefulness' in economics. A consumer may derive utility from a thing which is useful, such as a garden fork or a fountain pen; but utility in the economic sense can be derived from commodities which are not useful in the ordinary sense—luxuries of all kinds, for example. In economics, the expression 'a commodity conveys utility' means merely that some people want it; *why* they want it, as has been seen, is not the concern of economists.

So far as economists are concerned, therefore, a consumer may be said to derive utility from a commodity even though the consumption of that commodity is regarded by society as useless or even positively undesirable. It is not generally felt in Britain that the consumption of opium is a desirable practice; but there are people who are foolish or unfortunate enough to want opium, and want it badly, and they are prepared to pay for it. Therefore people can be said to derive utility from opium, however deplorable the consumption of opium may be.

The opposite of utility is *disutility*—fortunately a word not in general usage and therefore having less ambiguity about it. When it is said that some commodity or activity gives rise to disutility, it is meant that people would prefer not to have it or indulge in it. When people go out to work, it would be said that they derive disutility from their labour, meaning that they would prefer not to work, and thus have to be rewarded in some way for acting in a manner contrary to their inclinations. (The general assumption in economics that people derive disutility from work, and therefore have to be paid to do it, is often falsified because in reality many people enjoy their work. This is one more instance of a hypothesis which is forced to over-simplify the complex forces underlying human activities.)

2. WEALTH AND WELFARE

For a commodity or service to enter into exchange, and therefore to fall within the purview of economics, it must be scarce and utility must be derived from it. If it is not scarce—that is, if there is more than sufficient for everyone to have as much as he wants—there will be no need for people to give something else in exchange for it. If utility is not derived from it, no-one will *want* to take it in exchange for something else. There is still a third requirement, however; it must be physically possible to exchange the commodity in question. My health is a real asset to me, and plenty of people would pay a great deal if I could somehow give them part or all of my good health in exchange for their indifferent health. But it just cannot be done; therefore my health, infinitely valuable though it is to me, has no exchange value—that is, it cannot be used to obtain goods and services through the process of exchange. The economic

problem consists of allocating resources between various competing uses; hence it cannot arise in connection with entities which cannot be transferred or exchanged. Since economics is concerned with exchange in all its aspects, 'value' to the economist means *exchange value*, and the word will be used only in that sense in all that follows. However treasured a thing may be to an individual (a souvenir, a quality, an experience and so on), if that thing cannot be exchanged then so far as the economist is concerned it has no value.

In economic analysis, the word *wealth* means the total value of things possessed by a person or group of persons. Those things may be tangible goods such as houses, factories, land or cars, intangible assets such as skill or business goodwill, or sums of money. So long as they have value—that is, exchange value—they constitute wealth. Hence economists speak of individual, or private, wealth and of the wealth of a country or group of countries. As will be noted in a little while, wealth can be regarded as either the total stock existing at any one moment or a flow being acquired and consumed over a period of time.

It is not difficult to see that wealth is not the same thing as *welfare*. By welfare is meant the general condition of well-being of a person or group of persons, including their mental and moral as well as their material well-being. The assessment of general social welfare has to take account of considerations such as the contentment of society with its government, the state of law and order, or the health and education of the people. Economic welfare is that part of social welfare arising from the wealth-producing and consuming activities of a community. Normally, a rise in economic welfare—people becoming materially better-off—will be associated with a rise in welfare as a whole. Normally, but not necessarily. A despotic government may force its subjects to work longer hours and produce more goods, for example, but the people may well think that the loss of their liberty (or even their leisure) more than offsets the improvement in their supply of goods.

3. CAPITAL AND INCOME

A second important distinction in connection with wealth revolves round the question of whether, in a count-up of a com-

munity's wealth, it is intended to calculate how much wealth exists at a particular *moment* of time or to measure how much wealth the community will produce, or has produced, over a particular *period* of time—whether a flashlight 'still' photograph is being taken, so to speak, or a moving picture. If the former approach is followed, then the wealth is a certain stock which exists and can be counted up and valued, and that stock is known as *capital.* It would consist of what is usually called *fixed capital*—factories, houses, roads, machinery, shipyards—or of *working* capital—raw materials, half-finished goods still in course of manufacture, or stocks of finished goods. All capital is wealth which will be consumed only some time in the future; the difference between working capital and fixed capital is that the former is destined to be consumed directly, as particular goods in which it is embodied are finished and sold in the market, while the latter will be consumed only indirectly as it contributes over the years to the production of a whole series of goods and is gradually worn out.

If the second approach is adopted, however, and the amount of wealth produced over a period of time is measured, *income* is being calculated. Unlike capital, the total amount cannot be visualised as one enormous collection existing at a particular moment, because income is a flow which passes into consumption or into additions to capital as it is produced. The economic system can be likened to a river pouring through a pipe into a reservoir and then out again into the houses of consumers. To measure the amount of capital the amount of water existing in the reservoir on a particular day would have to be calculated; to measure income the quantity of water which had passed in through the pipe during a given week, or month, or year would have to be calculated. The amount flowing into the reservoir during a particular week might be the same as the amount flowing out to consumers through the outlet pipe. In that case the volume of the water in the reservoir would be the same at the end of the week as at the beginning, except for any loss from leakage or evaporation. If more water flows in than flows out, then the reservoir level will be higher. Much the same is true in economics. If the inflow (income) is equal to the outflow (expenditure by consumers) then the capital stock will be unchanged except for a wastage economists call *depreciation.* (More will be heard of

this a little later.) If income exceeds consumer expenditure, then the difference will take the form of a greater capital stock. If consumers spend more than they earn, then the community's capital is getting smaller and smaller—and that is a process which can no more continue indefinitely than can the process of taking more out of a reservoir than is being put in.

4. FACTORS OF PRODUCTION

A community derives its income through the intelligent use of the resources with which it is endowed. Unfortunately manna does not rain from Heaven any more; everything society consumes or puts into its capital stock has to be *produced*. The resources which society employs in this production can be classified into four groups comprising resources of broadly similar characteristics. These groups are called *factors of production*, meaning (literally) the things that make the production of wealth possible. The factors are:

Land—meaning not only the fields and hills, but all the other resources endowed by nature: the sun, wind, rain, water-power and mineral wealth. The quality of land, of course, varies enormously over the globe, not only because of its geological qualities but also because of its position, the prevailing climate etc.

Capital—briefly, capital may be defined as the stock of wealth existing at any moment of time; it comprises all those goods which are wanted by their owners because of the contribution they will make to the production or consumption of wealth in the future. It would therefore include manufacturers' or wholesalers' stocks of finished products.

Labour—the exercise of human physical or mental powers in the creation of wealth. The part played may be indirect, such as the manufacture of a single small component in a complex finished product. Or the labour may be directed not to the production of any tangible commodity but to the performance of personal services—as is the case with a teacher, or nurse, or lawyer.

Enterprise—a special form of human activity which is sometimes included under the same heading as labour but which has such special characteristics, and plays such a special role in the economic system, that it is best thought of as a separate factor. It refers to the exercise of powers of *organisation* in the combining of the other factors, and to the supply of *enterprise* in seeking out, developing and applying new methods of production, new products or new inventions.

On these four factors, operating together, depends the production of wealth. Different societies, or even different generations, are endowed with each of the four in varying measure; the production of wealth in a society will be at its maximum when the factors are combined with one another in the most suitable proportions.

5. THE MARGIN

In almost all contemporary economic theory great reliance is placed on the use of the concept of the *margin*. This concept is an unfamiliar one, and it is essential that the student should be quite clear as to what it means. In the economic world adjustments are constantly being made by both producers and consumers. Producers are always increasing or decreasing their output in response to changes in demand, or using more or less labour or capital as new methods of production come along. Consumers, similarly, are constantly altering the pattern of expenditure as the prices of some commodities change relatively to those of other commodities, or as fashions die or are created. A great deal of economics, indeed, is devoted to the analysis of the effects of adjustments such as these on the economy as a whole.

Now although economic conditions are never completely stable, so that these adjustments of production or consumption are constantly occurring, the process of change is always a *gradual* one. An industry as a whole never comes into existence or goes out of existence *en bloc* and overnight. One or two firms expand or reduce their output, one or two new ones may be created or may move into the production of some different commodity. As a result, for both the individual firm and the industry of which it forms part, the choice is seldom or never between producing or not producing. The choice is usually between producing a few extra units of output or a few less units.

The same is true of consumers. Only seldom do instances occur of people suddenly giving up altogether the consumption of a particular commodity; the usual situation is that in which a person decides to buy rather less of this and rather more of that. This is especially true when it is remembered that economics is concerned with the total consumer demand for a commodity, rather than with

the demand of an individual. In a market where consumers number thousands, or perhaps tens of thousands, changes are bound to be gradual; even if some individuals give up the consumption of a commodity altogether, or if new buyers come in who previously bought nothing, the proportionate effect on *total* demand is bound to be small. Cases have been known in which consumer demand suddenly underwent a violent change upwards or downwards, but these are distinctly exceptional.

Given this inherent characteristic of the real world, where production and consumption are the outcome of decisions by a large number of separate individuals, economic analysis naturally concentrates on the places where the changes occur. If 99 people consume today the same number of apples as yesterday, but one man stops eating apples today, any effects on the price of apples can be connected with the one man who for some reason has altered his consumption; economic analysis can concentrate on this element of change and ignore the elements which have not changed. If 9 producers keep a constant output and the tenth expands his by 50 per cent, analysis will again concentrate on this one changing factor in the situation.

This frontier, so to speak, at which changes and adjustments are liable to occur is called in economics the *margin*. The particular commodities which someone starts to consume, or ceases consuming, are called the *marginal units*. The marginal apple is no different physically from any other apple; it is simply that apple which would not be bought if demand changed ever so slightly and left consumers buying one apple less than before; the man who would give up that apple would be the marginal consumer—the man who would reduce his consumption of apples by one if (for example) the price of apples rose very slightly.

Similarly, economists speak of the marginal producer—the man who is just induced to produce a certain commodity with the existing level of profits to be obtained from it, but who would give up producing it if his profits fell slightly. If a shoe manufacturer thinks that at the present price it is worth his while to produce 100 pairs of shoes over a given period, then the hundredth pair of shoes is the marginal unit. It is no different from any other pair; it could not be picked out of his production line. Nevertheless there is a pair of

shoes which he would *not* produce if prices fell a little, and that is the marginal pair.

This concept will recur time and again in the course of this study, in all kinds of contexts. As it appears in each connection it will be defined for the special purpose for which it is being used. But the essential nature of the concept remains the same in all contexts; it always refers to that last single unit in the total supply which a person thinks is just, but only just, worth exchanging, and which would cease to be exchanged if its price or cost altered very slightly.

6. EQUILIBRIUM

Another fundamental concept in economic analysis is that of *equilibrium*. In ordinary speech, the word means a state of balance, and this is close to its meaning in economics. A situation is said to be one of equilibrium if there are no influences tending to change any of the factors involved in it. A firm is said to be in equilibrium, for example, if the rate of profit which its owner is receiving is just sufficient to cause him to continue producing the present level of output without any increase or decrease. An industry is said to be in equilibrium if there exists no tendency for new firms to enter it or for firms already in it to leave. A market is said to be in equilibrium when *at the ruling price* the amount of a commodity being offered for sale is just equal to the amount which its consumers wish to buy.

Two possible kinds of equilibrium situation can be distinguished. In one, *stable* equilibrium, any temporary movement in the factors involved, arising from some outside cause, would be replaced after a time by a movement in the opposite direction until the initial situation was restored; the situation contains some inherent stabilising characteristic which will bring things back to their starting-place. On the other hand, certain situations may be of *unstable* equilibrium, which means that if for any reason the factors involved are subject to a change, the situation will depart increasingly from the initial position and will have no tendency to return to it; on the contrary it will tend to settle in some quite different position.

Examples of these two types of equilibrium in economic affairs will be discussed later, after some basic analysis has been carried

out. For the moment, the distinction is best illustrated by the traditional example of the egg. If an egg is laid on a flat table along its horizontal axis (i.e. laid on its side, so to speak), it will settle in a position of rest. If pressure is now applied at one end the egg will be moved away from that position, but it will automatically return to its original position (in a sort of pendulum fashion) if the pressure is removed. It is in a position of stable equilibrium. On the other hand, if the egg is placed upright on its vertical axis (a certain amount of denting of one end would be required for this) and then pushed slightly to one side, it will show no tendency to return to its original vertical position. On the contrary, it will rapidly move further and further from it until it arrives at an entirely new situation of rest. It was in a state of unstable equilibrium.

7. ECONOMIC LAWS

From a systematic analysis of a deductive or inductive nature what is called an 'economic law' is frequently arrived at—for example, discussion will later centre on the law of diminishing utility, the law of increasing returns and other generalisations. Unfortunately economics can have no pretensions to arriving at immutable and universal truths such as those resulting from the development of physics and chemistry. The law of gravity, or Boyle's laws, will always remain valid so long as the natural order prevails. But economics is concerned with human beings, and human beings, unlike inanimate matter, have minds and wills of their own, with the result that one can never be *certain* as to how they will react in any given situation. All that economics can do, and all that it attempts to do, is to indicate from its deductive or inductive analysis certain generalisations which are *likely* to be valid of human beings in the mass and in the long run. The 'laws' which are arrived at, in other words, are statements of *tendencies* that are likely to result from certain causes.

For example, the law of demand indicates that, generally, the lower the price of a commodity the greater the amount of it which will be consumed. That could not be relied upon to be true of any single individual; some people, for example, may already have enough of it and will not increase their consumption of it however cheap it becomes. For the whole body of consumers taken to-

gether, however, it is very likely to be true, since both reason and experience indicate that sooner or later some people are bound to increase their consumption of the commodity involved.

Although economic laws can never be treated as if they were as universal and as certain as physical laws, therefore, their significance should nevertheless not be under-rated. They rest on the behaviour of people in fairly large numbers; they express what the normal and natural reaction of people will be to economic forces. This is a weakness insofar as no single individual can be relied on to behave in precisely the way that the group as a whole is likely to behave. But it is a strength insofar as such laws rest on the firm foundation of human nature which, although unpredictable in the individual, is immensely powerful in the community as a whole. If a parliamentary law is passed which conflicts with an economic law (i.e. with the natural inclination of ordinary people in matters relating to their economic welfare) it is doubtful if the parliamentary law will long survive as an effective measure. The experience of rationing and price control during the last war, for example, shows that it is only with the aid of determined enforcement and a lively social discipline amongst the community that economic laws can be prevented from asserting themselves in the form of black markets, as they did in many countries. Even in countries where the effects of conflicts of this kind have been successfully repressed, it is fairly clear that as time passes the tendencies described by economic laws become increasingly powerful. Even if they cannot claim anything approaching universal validity, then, economic laws are firmly rooted in human nature or experience, and although human nature can be restrained, and sometimes altered, neither is an easy task.

FURTHER READING

A. MARSHALL, *Principles of Economics*, Bk. I, Ch. III, Bk. II, Chs. I–IV.
SIR H. HENDERSON, *Supply and Demand*, Ch. II.
* L. ROBBINS, *An Essay on the Nature and Significance of Economic Science*, Chs. IV–V.
* J. N. KEYNES, *The Scope and Method of Political Economy*, Ch. IX, Note A.
P. H. WICKSTEED, *The Common Sense of Political Economy* (Routledge, London, 1933), Vol. I, Bk. I, Ch. II.

PART II

THE THEORY OF DEMAND

CHAPTER 3

Demand and Utility

1. TOTAL AND MARGINAL UTILITY

In the previous chapter the word 'utility' was defined, for the purposes of economic analysis, as the satisfaction, or pleasure, or benefit derived by a person from the consumption of wealth. A moment's reflection on those words 'satisfaction', 'pleasure' and 'benefit' will be enough to show that they are vague and elusive terms. In particular, they are essentially *introspective* terms—that is, they relate to inner sentiments and emotions and not to things having a physical or material existence, although the phenomena they describe are none the less real for that. For example, Smith may derive intense enjoyment from wearing a tie which Brown can look at only with intense suffering; one man's meat is another man's poison, as we say. Utility is something experienced within a person and is not inherent in the physical commodity with the aid of which he experiences it.

Given its introspective nature there is obviously no way in which to measure *total utility*—that is, the total satisfaction derived by a person from a commodity—any more than any human sentiment or emotion can be measured. I may believe that my bad tooth is causing me more pain than yours is causing you, but there is no way of testing this belief; similarly, you can believe that you have enjoyed a certain film more than I have, but there is no way of proving it.

The position is not quite the same with *marginal utility*. By this expression 'marginal utility' is meant the addition to a person's total utility brought about through the consumption of the last unit of a commodity he consumes. If I use three gallons of petrol in a week, then my marginal utility is the difference between the total utility I derive from the three gallons and the total utility I would have derived if I had used only two. If I go to the cinema six times

each month, then my marginal utility from seeing films is the difference between the total utility I derive from six films and the utility I would derive from seeing five. As will be established in a little while, this is by no means the same thing as saying that the marginal utility of a commodity is the total utility a person derives from it divided by the number of units of it which he consumes.

The important difference between total and marginal utility is that although utility cannot be measured, it *is* possible to say that a consumer will derive the same marginal utility from the last unit he consumes of each commodity entering into his range of expenditure. This may seem rather odd; if utility cannot be measured, how can marginal utilities be known to be equal? The answer is that a person's behaviour in spending his income indicates how much he is prepared to pay for the last unit of a commodity which he consumes, assuming that he is acting rationally and seeking to obtain the maximum possible satisfaction from his limited income. This amount can be taken as a measure of the extra satisfaction which that last unit brings in to him—i.e. of its marginal utility to him. For example, suppose that when a certain type of chocolate is selling at 6d. a bar, a man buys three bars a week; from this it can be concluded that the marginal utility of chocolate to this particular man is roughly reflected by the price of 6d.—or, more precisely, that it is equal to the utility of the other things he could have bought with that 6d. If the marginal utility was *less* than that of the other things he could have bought with 6d., then presumably he would not have bought that third bar (and perhaps not the second either) but would have spent the money on other things; if the marginal utility was *more* than that of the other things 6d. could have bought, then presumably he would be buying more than three bars and spending 6d. less on other goods. The fact that he buys three bars, but no more than three bars, suggests that the marginal utility is just equal to the sacrifice represented by paying over 6d. and foregoing the other things which that 6d. could have bought.

Does it follow from this that the *total* utility derived from a commodity can be measured by the total amount of money spent on it? In the example of the man buying chocolate, can it be said that, since he spends 1s. 6d. a week on chocolate, the total utility he derives from chocolate is equivalent to 1s. 6d.? Unfortunately it

cannot. The price of the commodity can be taken as indicating marginal utility, but the consumer buys *all* units of any commodity at that price; total expenditure would reflect total utility, therefore, only if the utility derived from each unit of the commodity was equal to its marginal utility. There would be no foundation at all for assuming that this is the case; in fact there is good reason for supposing that it is untrue. This limitation—that only marginal utility can be taken as reflected in the price a consumer is willing to pay, and not total utility—will not prove particularly inconvenient for the level of analysis with which this book is concerned. The fundamental conditions for equilibrium of price and output can be worked out by reference to marginal utility, and total utility is not involved.

A more important weakness of this concept of utility emerges, however, if an attempt is made to compare not total and marginal utility but the marginal utilities of different persons. Although it can be said that the marginal utility derived from a commodity is reflected by the price which consumers are willing to pay for it, it does not follow that if two people pay the same price for a commodity the marginal utility derived from it is the same for each person. If I say that the marginal utility I derive from a bar of chocolate is measured by 6d., I mean that it is equal to the other things I could have bought with that 6d. But the utility represented by 6d. to me may not be the same as the utility represented by 6d. to someone else. To a poor man, 6d. may mean a great deal; to a rich man, it may not be worth the bother of picking up. Another consideration is that a man of education and culture may derive a high degree of utility from commodities which other people are unable to appreciate properly. For reasons of this kind, it is extremely doubtful if there is any sense in which a comparison can be made between the utilities enjoyed by different people. So far as market equilibrium is concerned, this point is not of any fundamental significance; but it is important in the context of judgements as to whether society is 'better off' or 'worse off' as a result of a certain act of economic policy. Judgements of this kind, assessing the total welfare of society or of different elements in society, must rest in the last resort on political convictions and can have no claim to rest on objective scientific analysis.

2. THE LAW OF DIMINISHING UTILITY

Personal experience, and observation of the behaviour of others, makes familiar a human characteristic which underlies the whole nature of demand. This characteristic is that in general, and if all other things remain constant, the extra satisfaction derived from the consumption of additional units of any commodity tends to decline as the quantity consumed increases. A second chair in my dining-room is not quite as essential as the first; a third is a little less useful than the second, and so on. By the time I am contemplating buying a tenth chair I am beginning to wonder whether there is much point in cluttering up the place with so many chairs.

This simple and familiar fact of life has been generalised by economists into *the law of diminishing utility* which states that, other things being equal and over any specified period, the utility derived by a person from the consumption of a commodity increases as his consumption increases, but not as rapidly. In other words, marginal utility gets smaller and smaller the more of a commodity a person has. It is important, however, to bear in mind the qualifications 'other things being equal', since the law may not hold if factors other than the rate of consumption are changing. What exactly is involved in this qualification?

First and most important, it is assumed that *income* is constant, since the satisfaction derived from a commodity will naturally depend on the number and quality of other commodities which the consumer has at his disposal. Hence, if over any period a person's income goes up or down, the marginal utility derived from a commodity may undergo a change as a result, and the person may find that the marginal utility derived from a unit of the commodity has increased or decreased since the beginning of the period concerned.

Secondly, it is assumed that the *nature* of the successive units consumed is constant. This means not only that they must be physically identical (i.e. one bar of chocolate must be identical in blend, quality etc. with every other) but also that the consumption of successive units does not affect the consumer's basic attitude towards them. For example, certain drugs are habit-forming, and a person may become addicted to them as his consumption of them increases; naturally, his marginal utility will not necessarily decline

under these circumstances. Similarly, consumption of a commodity may result in a person acquiring a taste for it, and his enjoyment may increase as his consumption increases. Or the nature of the consumer's attitude towards a commodity may change; for example, as a man's collection of a particular type of rare stamp increases, further acquisitions may become more and more not mere stamps but the means to attaining the world's biggest collection of something. Finally, if the period of time involved is very long in relation to the time taken to consume the commodity, successive units of consumption may be so far separated as to lose all connection with each other.

Summing up, then, the law of diminishing marginal utility holds only if incomes and tastes remain constant. If either changes, the law may become inapplicable; it will be only a temporary invalidity however, since it will come back into operation as soon as the change has ceased and a new pattern of incomes and tastes has been established.

3. THE CONCEPT OF EQUI-MARGINAL RETURNS

From these two concepts of marginal utility and diminishing marginal utility economists derive a basic theorem describing the way in which consumers with limited resources (i.e. a finite income) will divide their expenditure as between the infinite number of different goods and services they could enjoy. This can be done only if it is postulated that consumers wish to get the maximum utility from their limited income, and that they act rationally in seeking to attain this end. It is possible to argue that in real life people do not always have the aim of getting the greatest enjoyment of material things which is open to them; people often act in a way which is contrary to their best material interests—for example, because of religious or political convictions. Similarly, it can be argued that people are by no means rational in their behaviour, but allow their judgement to be influenced by irrational forces (such as superstition or strong emotions) or are in ignorance of the most logical course open to them.

There is a substantial amount of truth in these arguments, but the economist would justify the retention of his basic postulate of utility-maximisation on two grounds. First, although people may

not always be strictly rational in either their motives or their actions insofar as economic affairs are concerned, it is only by the process of logical deduction that a good deal of analysis can be carried out; if the assumption of rationality is abandoned, then systematic analysis becomes impossible and economists would be reduced to stating that there is an infinite number of possibilities in any given situation, a conclusion of singularly little usefulness. Secondly, although the assumption of rationality on the part of everyone all the time may not always be realistic, experience shows that it is realistic for *most* people and for *most* of the time. Few generalisations about human behaviour can hope to be true of all times and all people; the economist, being concerned with society as a whole rather than with particular individuals, believes that his assumption concerning human economic behaviour is certainly the most realistic generalisation which can be made.

Given this assumption that in general people seek to maximise their utility and act rationally towards that end, there can be established the generalisation that people will distribute their expenditure as between various commodities so as to ensure that the marginal utility obtained from every commodity purchased is in the same relationship to price. This is usually known as the law of equi-marginal returns. The example given in Table I may help to make it clear. In this example it is assumed that a person is trying to decide how to distribute an income of 46s. amongst butter, cheese and onions, and it is assumed also that the marginal utility derived from each successive unit of each commodity can be measured by the person concerned in terms of some undefined

TABLE I: *Equi-marginal Returns*

	Marginal utility derived from each pound		
Pounds consumed	*Butter* (6*s. a lb.*)	*Cheese* (4*s. a lb.*)	*Onions* (2*s. a lb.*)
1	25	20	21
2	18	15	17
3	12	11	14
4	10	8	10
5	9	6	7
6	8	5	4
7	5	2	3
8	3	1	2

unit. In this example, he would buy 3 lbs. of butter, 4 lbs. of cheese and 6 lbs. of onions, because in this way the ratio of the marginal utility derived from each commodity to its price would everywhere be equal (2 to 1)—12 to 6s. for butter, 8 to 4s. for cheese, and 4 to 2s. for onions. The significance of this particular distribution of expenditure (i.e. that which brings marginal utility everywhere into the same ratio to price) is that it maximises the total utility derived by the consumer from a given income. If he distributed his expenditure in any other way his total utility would necessarily be reduced. By buying one more pound of butter and one pound less of cheese and onions, for example, the consumer in Table I would be losing more utility from the cheese and onions he sacrifices (8 units from the former and 4 from the latter) than he would gain from the extra pound of butter (10 units). *Given the law of diminishing marginal utility, in fact, this requirement that marginal utility should be everywhere in the same ratio to price must be met if the maximum utility is to be obtained*; increasing consumption of one commodity at the expense of another after this point has been reached necessarily means that the marginal utility gained from each extra shilling spent on the one is smaller than that sacrificed through spending a shilling less on the other. The fact that in practice we cannot measure utility in no way affects the validity of this generalisation; we do not need to know the magnitude of something in order to know that it becomes smaller through the process of adding to it units which are smaller than those which are simultaneously being taken away from it.

What is important here, it should be remembered, is the *proportion* in which expenditure is distributed amongst the commodities. In whatever units the marginal utilities are measured, the essential point is that total utility will be maximised when the *ratio of marginal utility to price* is equal for all the commodities; at the optimum pattern of expenditure marginal utility may be everywhere twice, or ten times, or twenty times price. In the example given just now, the ratio of marginal utility to price was everywhere 2 units = 1 shilling. If the person had been spending 64s. rather than 46s., he would have bought 5 lbs. of butter, 5 lbs. of cheese and 7 lbs. of onions, so that marginal utility would have everywhere been $1\frac{1}{2}$ times price.

4. DEMAND CURVES

It has been seen that people desire goods and services because they derive utility from them, that the utility derived from consumption will vary as between different commodities and different people, and that in general the additional utility derived from any commodity by any one person diminishes as his rate of consumption rises. In economic analysis it is convenient to express all these things in the form of a *demand curve* for each individual commodity, such as that shown in Fig. 1. This curve is a line connecting the various possible prices charged for the commodity (measured along the vertical axis) with the amounts of the commodity which people will be prepared to buy at each different price. For example, the curve shows that if price was OP_1 consumers would purchase a quantity OM_1, whereas if price fell to OP_2 consumers would purchase a total quantity OM_2.

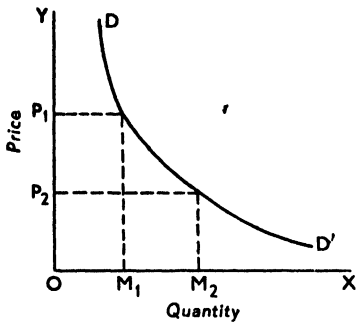


FIG. 1

A question which immediately arises when we draw such a curve is this: why should it slope downwards from left to right, suggesting that the lower the price of a commodity the greater the quantity bought in the market? The basic characteristics of consumer behaviour which have already been discussed contain the answer to this question. The discussion began with a recognition of the tendency for human beings to derive less and less additional satisfaction from a commodity as their consumption of it increases—i.e. the tendency for marginal utility to fall as the quantity consumed increases. It was then postulated that consumers act in such a way as to *maximise* the utility they obtain, and it followed that this maximisation is secured when the marginal utility of each commodity a person consumes is everywhere in the same ratio to its price. Now, given the law of diminishing marginal utility, it must follow that demand curves will normally fall from left to right, since if a consumer will derive from an increase in his con-

sumption of any given commodity a smaller marginal utility than he now enjoys, he will only increase his consumption if price falls so as to offset the decline in marginal utility.

Or look at it this way. At any moment a consumer is maximising the utility derived from a given income; the marginal utility he derives from each commodity is everywhere in the same proportion to price. If, all other things remaining equal, the price of one commodity falls, its marginal utility/price ratio will now be greater than that of any of the other commodities. The consumer will therefore be able to get from it a greater marginal utility *in terms of the sacrifice involved* (i.e. per penny of its price) than from any other commodity. To maximise his utility, therefore, he will switch expenditure from other goods, whose marginal utility/price ratio is relatively low, and spend more on the commodity whose price has fallen, where the marginal utility/price ratio is relatively high. And he will go on switching expenditure to this relatively cheap commodity until (as his consumption of it rises) its marginal utility falls again into the same relationship with its price as that of other commodities. Conversely, if the price of a commodity rises, the relationship between its marginal utility and its price will have *fallen* below that of all the other commodities, and it will pay a consumer to switch his expenditure away from that commodity and towards the others. Normally, then, a rise in the price of a commodity will reduce the quantity consumed, and a fall in its price will cause an increase in consumption—i.e. its demand curve will slope downwards from left to right.

In practice it is seldom possible to draw up a schedule of this kind showing the amounts purchased (i.e. the 'demand') at every conceivable price, since in the real world only a limited range of prices has usually been experienced. It may be possible to draw up a more limited schedule of prices and quantities demanded, however, from the statistics of different markets or by experiment. For example, it may be known that for a certain type of motor-car the numbers demanded at recent prices have been as follows:

Price (£)	*Quantity demanded* (*000s*)
650	150
600	300
550	400
500	550

From these data we can draw up the demand schedule shown in Fig. 2, although it is necessary to *assume* that the curve joining the four observed points is a smooth one—the quantity demanded at a

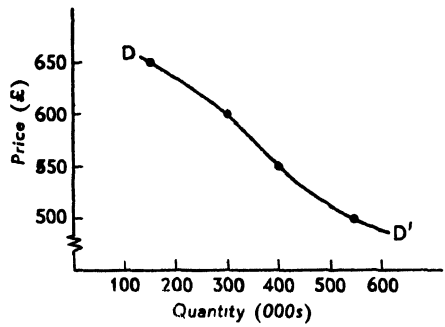


FIG. 2

price between £600 and £550 is not *known*, but it can be assumed that the rise of the quantity demanded from 300,000 at £600 to 400,000 at £550 will be a gradual and continuous one.

Even if it was statistically possible to construct a demand curve of this kind, it could be drawn up only on the assumptions which underlay the law of diminishing utility (since the normal demand curve is no more than a graphical expression of that law)—that is, that income and tastes are constant. If either of these changes, then it will be necessary to draw up an entirely new demand curve. For example, let the curve D_1D_1' in Fig. 3 represent the demand schedule for oranges at a particular moment. Now imagine that the public suddenly becomes more conscious of the nutritional value of oranges; a completely new curve, D_2D_2', will need to be drawn, showing a greater quantity than before demanded at any given price. A similar shift would occur if people in general became

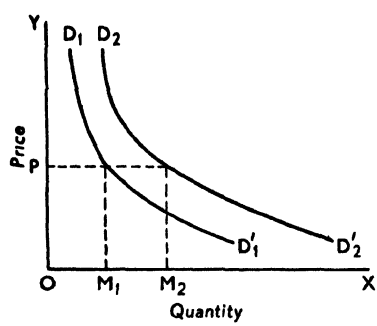


FIG. 3

richer and could afford to buy more oranges at any particular price.

Another factor assumed constant in the construction of these demand curves is the price level of goods which are either near substitutes for or closely competitive with the commodity in question. As will be seen in a later section of this chapter, some commodities may be related to one another in an especially close manner, so that a change in the price of one commodity may cause a shift in the whole demand schedule for another commodity.

Put briefly, then, the demand curve for a commodity is drawn up on the assumptions that the incomes of consumers are constant, that their tastes are constant, and that the price level of closely related commodities is constant. It is important therefore to distinguish between a change which occurs in the consumption of a commodity as a result of a change in its price and one which occurs because of a movement in one of these factors which are assumed to be constant. In the former case the movement in consumption is measured by moving along a single demand curve; it is usual to speak then of a *contraction* or *extension* of demand, according to whether the quantity consumed is reduced through a rise in price or increased through a fall in price. In the latter case it is necessary to draw up an entirely new demand curve, and economists speak of a *shift* in demand, or an increase or decrease of demand.

In all this discussion the quantity purchased by consumers has been related to a sum of money, the price. In fact consumers are always comparing the marginal utility of a commodity with that of the other commodities which they could obtain with the resources available to them. To obtain a commodity, consumers have to sacrifice not so much a sum of money, which usually has little or no inherent value, but the other things they *could* have bought with that money. Money, as is shown at a later stage, is merely the medium through which the exchange of goods is most conveniently carried out. The consumer is always assessing the relative attractiveness of all the commodities he could conceivably purchase, and ideally along the vertical axis should be measured not money but the utilities which a consumer is sacrificing in order to consume a particular commodity. It is impossible to do this, since the number of alternative commodities available is infinite, but if

it was possible to do this it would make clearer this important fact that consumers are always seeking to bring into equality the marginal utility derived from each commodity and the marginal utility they sacrifice by buying it rather than some other commodity.

5. PERVERSE DEMAND CURVES

It has been seen that demand curves normally slope downwards from left to right, reflecting the fact that a fall in the price of a commodity will usually stimulate an increase in its consumption while a rise in its price will usually reduce consumption. Is it possible for *perverse* demand curves to exist—that is, for this normal price-quantity relationship to be reversed? Instances of this can be conceived of, but it will be found that in each of them the assumptions underlying the demand curve are in fact falsified. The first, and most important, case is that of so-called *inferior goods*. By this is meant goods whose consumption falls as income rises; they are often commodities of a relatively low quality which form an important element in the expenditure of the poor. The best-known examples are staple food-stuffs such as bread and potatoes. Because of their bulk, such foods are important in the diet of poorer people; as people become richer they prefer to buy the more expensive and attractive foodstuffs (for example, dairy products or confectionery) and so reduce their consumption of these coarser foods. If such commodities form an important element in a person's expenditure, a rise in their price can cause a significant reduction in his *real* income—i.e. his money income in terms of the goods and services which can be bought with it—since the purchase of a given number of, say, loaves of bread (which are essential and *must* be bought) now costs more, so that he is left with less to spend on other things. As a result, he will have to cut down his expenditure on relative luxuries, such as cakes or meat, and will need to buy more bread in order to fill up the nutritional gap which this causes. The rise in the price of bread may therefore result in an *increase* in the quantity of bread consumed.

It will be noticed that this odd result follows because a change in price causes a change in real *income*, and the law of diminishing utility, on which normal demand curves are based, specifically postulates an unchanged income. What is happening in the case of

these inferior goods is a shift of the demand curve as price rises and real income falls—from D_1D_1' to D_2D_2', from D_2D_2' to D_3D_3', and so on, as shown in Fig. 4 (*a*)—rather than a movement along an upward-sloping demand curve, DD', such as that shown in Fig. 4 (*b*), which would result from simply connecting up the various prices and the quantities purchased at those prices.

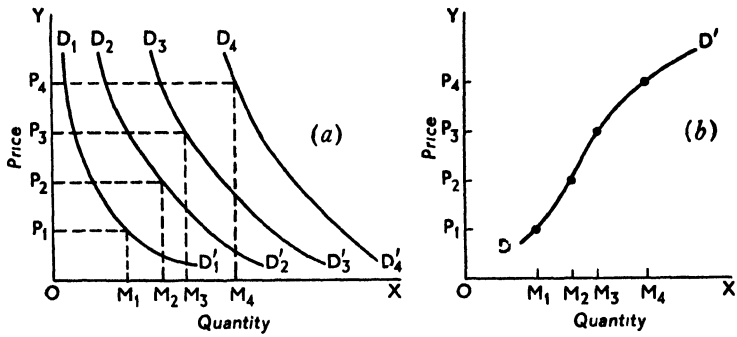


FIG. 4

Similar considerations apply to the two other situations in which there may seem to be perverse demand curves. In one, an increase in price gives rise to expectations on the part of consumers that *further* price increases are likely, so that they rush to buy a commodity before its price rises even more—a sequence of events frequently underlying Stock Exchange transactions in securities. Here again, at first sight, is a case where a greater quantity is purchased at a high price than at a lower one; once again, however, it is one in which the whole state of demand is changing, this time because of consumers' expectations as to future prices. A series of demand curves shifting to the right or left is more appropriate than a single upward-sloping curve.

In the remaining type of situation, the commodities involved are demanded specifically because of their high price and therefore exclusive nature—especially precious stones and other means of human ostentation. If these commodities were reduced in price so that they became within the means of a large number of people, they would thereby lose their exclusive character and hence their appeal. Apparently a fall in price would result in a fall in the

quantity demanded, but the situation is in effect one which involves different commodities; although the commodity is physically unchanged by a change in price, its character so far as the consumer is concerned *has* changed, and it is therefore a different commodity from the economist's point of view. As in the other cases, therefore, new demand curves (of normal slope) are really involved, rather than a single perverse curve.

6. INTER-RELATIONSHIPS OF DEMAND

Before concluding this examination of some fundamental characteristics of the demand curve, a word is necessary about interactions which may occur between the demands for commodities which are physically distinct. As was noted earlier, a demand curve is drawn up on the assumption that the prices of other goods are constant at a time when a change is taking place in the price of the commodity whose demand is being analysed. In real life prices are constantly changing, so that this assumption seldom holds. This would not matter too much if all commodities were related to one another only in the sense that all of them compete for the limited income of consumers. If I spend £1 more on books, the demand for other commodities is bound to be affected, since I have £1 less to spend on other things. But if there is no *special* relationship between commodities, the effects on the demand for commodities in general of a change in the demand for one in particular can safely be ignored; even if the change is large for the one commodity (say a doubling of total expenditure on it), the total number of commodities in existence is virtually infinite; if the effects of this change in the demand for one commodity are spread evenly over all other markets, it will not have any significant effect on any one of them.

Occasionally, however, inter-relationships exist which result in an uneven spread of such changes, so that the markets for physically separate commodities are intimately linked with one another. Three important inter-relationships of this kind are worth noting. The first is known as *joint demand*; this exists when the satisfaction of a want involves two or more physically separate commodities used in conjunction—the demand for petrol is closely linked to the demand for motor-cars, and similar links exist with pipes and pipe-

tobacco, pens and ink, knives and forks etc. The effects of this inter-relationship are easily visualised. A shift in the demand for one product, or a movement along its demand curve as a result of a change in price, will cause a shift in the same direction in the demand for the related product without any change in the latter's price. Commodities linked in this way are known as *complementary* goods. (The opposite situation prevails, of course, with goods which are directly *competitive*—electric cookers and gas cookers, cigarettes and pipe-tobacco, rail travel and motor-coach travel etc. Here, a movement in the demand for one product because of a change in its price will cause an *opposite* movement in the demand for its substitute, without any change in the price of the latter).

Secondly, there is the case of *derived demand*, which exists when a commodity is demanded not for its own sake but by virtue of its contribution to the manufacture of another product. Few people want steel as such; the demand for it is derived from the demand for the commodities made of steel—cars and washing machines and railway lines. Similarly, the demand for bricks is derived from the demand for houses and factories, and that for raw cotton is derived from the demand for dresses and shirts. This relationship is basically similar to that of joint demand, and a similar connection will exist between the markets concerned.

The final type of inter-relationship is known as *composite demand*. This exists when a single product is wanted for a number of different uses, as contrasted with joint demand where several products are wanted for a single use. Instances of composite demand are found with, for example, sugar (used for direct consumption, or baking, or jam-making, or even fire-lighting), sheep (for wool, or meat, or skins) and coal (households, gas-making, electricity generation, ships' bunkers). The interaction between different markets in cases such as these operates through supply rather than demand. An increase in the demand for one type of use will tend to draw supplies of the product away from its other uses, causing a relative scarcity in other markets and therefore influencing price and consumption in those markets.

In dealing with all these types of demand, then—and they cover a wide range of the whole economic system—it must always be borne in mind that the assumption of the independence of demand

curves by no means always holds. The demand for any given commodity will depend on its price; but it may also depend on the prices and supply of, and demand for, many other commodities having at first sight no connection with it.

FURTHER READING

MARSHALL, *Principles of Economics*, Bk. III, Chs. I–III, V–VI.
HENDERSON, *Supply and Demand*, Ch. III.
WICKSTEED, *The Common Sense of Political Economy*, Vol. II, Bk. II, Chs. I–III.

CHAPTER 4

Indifference Curves

1. PREFERENCE VERSUS UTILITY

In discussing the concept of utility it was made clear that there are serious limitations to its usefulness in the real world because, being essentially an introspective concept, it did not lend itself to measurement. It can reasonably be assumed that a person is able to assess for himself the relative marginal utilities of different commodities, and that the price he is willing to pay for different commodities is therefore a guide to the marginal utility he derives from them. But the fact remains that the utility derived by two people from a commodity which they buy at the same price may be widely different, so that the whole concept of utility is inevitably a somewhat vague and difficult one.

Primarily to place the theory of consumer behaviour, or demand, on a less nebulous basis, the technique of *indifference curve analysis* has been evolved. This analysis rests on the fact that, while utility cannot be measured, it is possible to observe from experience and experiment the *preferences* which people display when choosing between different commodities. Consider the case of a man buying a sandwich in a railway refreshment room. There are cheese sandwiches or tomato sandwiches, both being sold at the same price, and he decides to have a cheese sandwich. What conclusion can be drawn from this? It is certain that nothing can be said about the absolute magnitude of the satisfaction he derives from the cheese sandwich, since this is inherently immeasurable. But it *can* be said that that satisfaction—whatever it is—is *greater* than the satisfaction he would have derived from the tomato sandwich. Why else would he have chosen cheese rather than tomato? It is assumed always, of course, that people act rationally, that they use their resources so as to obtain the greatest possible benefit from them.

Having made that assumption, it follows that if a man chooses to employ a certain amount of his wealth on buying one unit of commodity A rather than one of commodity B, he must derive more satisfaction from A than from B. In other words, even if nothing can be said about how much utility a person enjoys from any given commodity, it can still be said that that utility is greater or less than the utility he would obtain by the expenditure of an equal amount of money on some other commodity which he does *not* buy.

Now in actual life such preferences are being expressed every day by everybody. A woman spends her clothing allowance on a new hat rather than on a new scarf; a man buys more tobacco rather than a new book. Every purchase in a shop is an expression of preference as between different commodities, implying that greater satisfaction is derived from the goods which are bought than would have been derived from those which could have been bought but are not. The comparison need not be only between goods selling at the same price, of course. If I go once to the theatre instead of using the money to go twice to the cinema, then it follows that I derive from one visit to the theatre more satisfaction than I would from two visits to the cinema. (*How much* more, of course, no-one can say.)

If people are rational, their actual expenditure behaviour can be interpreted so as to establish their scales of preference as between various commodities, and this will indicate which of two comparable expenditures gives the greater satisfaction. More than that, their expressed preferences as seen in actual purchases can be used to discover from which sets of commodities consumers derive *equal* satisfaction. Consider again the case of the man choosing his sandwiches. If he had expressed *no* particular preference as between cheese and tomato, and had left it to the waitress to give him whichever she cared to, then it could have been concluded that the man derived equal satisfaction from either kind of sandwich, although once again nothing need be said about the absolute magnitude of that satisfaction. Similarly, I may prefer a visit to the theatre to two visits to the cinema, I might have no particular preference between one visit to the theatre and *three* visits to the cinema, and I might definitely prefer four visits to the cinema to

one to the theatre. Assuming that I am acting rationally, it follows that I derive *more* satisfaction from one theatre visit than from two cinema visits, *less* satisfaction from one theatre visit than four cinema visits and *equal* satisfaction from one visit to the theatre and three to the cinema. All this can be concluded without any necessity for knowledge about *how much* satisfaction I derive from either form of entertainment or about the relationship of that satisfaction to the price paid for it.

Indifference curve analysis uses as its basis this latter fact that if a person has no especial preference as between a given amount of one commodity and a given amount of another—i.e. if he is *indifferent* as between these alternatives—then he derives an equal degree of satisfaction from the two sets of commodities. By experiment it is possible to discover what combinations of goods have this characteristic for any particular person. He can simply be asked 'Would you prefer 10 of commodity A or 1 of commodity B?', 'Would you prefer 10 of A or 2 of B?' and so on, until the quantity of commodity B is reached at which the person concerned expresses himself indifferent as between 10 of commodity A and the given number of B. The questioning can then be recommenced, taking as the alternatives 9 of commodity A and various quantities of B, until once again the quantity of B which has equal attraction for the consumer as the postulated amount of A is arrived at. And so the process can go on, the questioning being repeated for every possible quantity of commodity A. If quantities of commodity A are now measured along one axis of a graph and quantities of commodity B along the other, all the combinations of A and B which leave the consumer indifferent between them (i.e. all the pairs of quantities from which he derives equal satisfaction) can be plotted. If all these points are now joined, the resulting curve is known as an indifference curve; it shows the quantities of the two given commodities between which the consumer expresses no especial preference.

A simple example may make this clear. Suppose that there are only two goods in existence—apples and oranges—and that a person has a fixed income given to him in the form of ten apples. He is now asked how many apples he is prepared to exchange for one orange —i.e. he is asked what number of apples plus one orange is equally

satisfactory to him as ten apples and no orange. Then he is asked how many apples he will sacrifice for another orange—again, what number of apples together with two oranges give him the same satisfaction as ten apples and no orange. The questioning is continued in this way, his answers being noted down, until the number of oranges needed to induce him to part with *all* his apples has been found. A schedule showing the different combinations of apples and oranges which give the consumer equal satisfaction could then be drawn up. It might be like this:

No. of apples	*No. of oranges*
10	0
9	1
8	2
7	3
6	5
5	7
4	10
3	13
2	17
1	22
0	27

As between all these possible combinations, the consumer will be *indifferent*—he will feel himself no better off and no worse off whichever combination he has. If all these points are now plotted on a graph—measuring apples along the vertical axis and oranges along the horizontal axis, as in Fig. 5—the result will be an indifference curve, *II'*, a schedule showing the various combinations of commodities which give the consumer equal satisfaction.

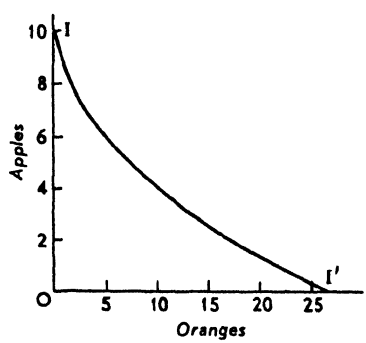


FIG. 5

So far the discussion has proceeded as if there were only one indifference curve for each person, but in fact there will be an infinite number of them, corresponding to each possible level of the consumer's satisfaction. The previous example was one of a consumer indicating every

equally satisfactory combination of apples and oranges between 10 apples and no oranges at one end of the scale, and no apples and 27 oranges at the other end. But the process could be repeated on the assumption that the consumer began with 11 apples rather than 10, and again with a starting position of 12 apples, or 20. Each time a new schedule of equally-satisfactory combinations would be obtained, and an indifference curve could be constructed from each. The final result would be, as in Fig. 6, an indifference 'map' showing the equally-satisfactory combinations with each level of total income, $I_1I'_1$,$I_2I'_2$ and so on.

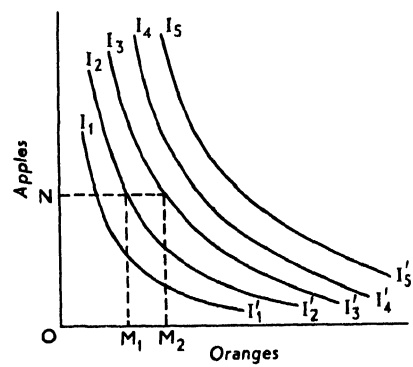


FIG. 6

An indifference curve to the right of another indicates combinations each of which yields a higher total satisfaction than any combination on a curve to the left. For example, $I_3I'_3$ indicates the combinations of goods giving the same satisfaction as (say) ON apples and OM_2 oranges. Every point on $I_2I'_2$ indicates a level of satisfaction equal to that derived from ON apples and OM_1 oranges. Since the consumer must derive greater satisfaction from a combination comprising a given number of apples and a larger number (OM_2) of oranges than from the same number of apples and a smaller number (OM_1) of oranges, every point on $I_3I'_3$ must represent a higher level of satisfaction than any point on $I_2I'_2$.

Similarly, it is clear that indifference curves cannot intersect. If they did it would mean that a consumer was indifferent as between

a point on a high indifference curve and one on a lower curve (i.e. the point at which the curves intersect and which represents the same combination of goods on each). This would contradict the previous finding that an indifference curve must represent at *every* point a higher level of satisfaction than every point on a curve to the left of it. Hence it is impossible for indifference curves to intersect.

This apparatus can show why a consumer will benefit from exchanging goods rather than consuming only a single commodity. Suppose the consumer has an income of *ON* apples, and that his indifference 'map' is that shown in Fig. 7. By definition his satis-

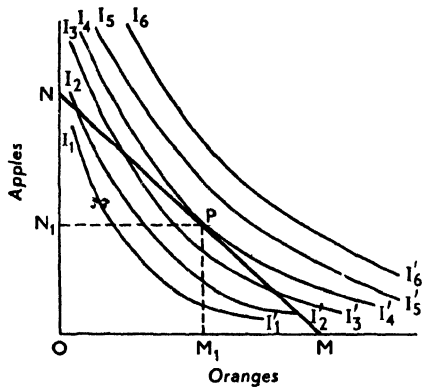


FIG. 7

faction will be maximised by achieving a combination of apples and oranges which lies on the highest possible indifference curve. Suppose also that in the world as the consumer finds it the current rate at which apples are being exchanged against oranges is in the ratio *ON* : *OM*—i.e. that *ON* apples can be exchanged in the market for *OM* oranges. This exchange rate can be shown on the graph by connecting the point on the vertical axis representing *ON* apples with the point on the horizontal axis representing *OM* oranges. This line (the *budget line* as it is called) will just make a tangent with an indifference curve. (Since there is an infinite number of curves in an indifference map, as has been seen, the budget line is *bound* to be a tangent to one of them.) That point of tangency, *P*, indicates what combination of apples and oranges the

consumer will choose, given both his relative preference for apples and oranges and the current rate of exchange.

Why is this? The explanation is that by entering into exchange the consumer has been able to reach a point on the indifference curve $I_4I'_4$. Given his initial stock of apples and the current apple-orange exchange rate, the budget line NM shows all the possible ways of consuming his income which are open to him. It is not possible for him to reach a point on any curve higher than (i.e. to the right of) $I_4I'_4$. But he *can* reach $I_4I'_4$, and by obtaining a combination on that curve he will, by definition, obtain more satisfaction than from any other combination which would lie on a lower curve. Hence the point P—where the budget line is tangential to an indifference curve—indicates the combination of apples and oranges from which (given his initial income and the exchange rate between apples and oranges) he will derive the maximum satisfaction.

This conclusion can be related to the earlier analysis of consumer behaviour based on the concept of marginal utility. The slope of an indifference curve at any point indicates the terms on which a consumer is prepared to exchange one commodity for another. Let Fig. 8 represent a magnified section of an indifference curve. The slope of the section AB indicates that in return for 10 apples (given his initial stock of 70) the consumer will require an additional 30 oranges if he is to feel neither better nor worse off. If a smaller and smaller section of the curve is taken, and a smaller and smaller number of each commodity, the slope at one point only can be conceived of, and this will indicate the rate at which the consumer is willing to exchange *one* unit of one commodity for the other—i.e. his *marginal* rate of exchange of apples for oranges, or what is usually called his *marginal rate of substitution* of apples for oranges. If the budget line is a tangent to the indifference curve, this means of course that both lines have the same slope at that point—i.e. that

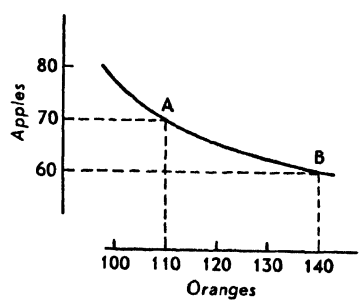


FIG. 8

the marginal rate of substitution is *equal* to the relative prices of the two commodities. This is exactly parallel to the earlier finding that in equilibrium a consumer will distribute his expenditure so that marginal utility is everywhere in the same proportion to price.

Another parallel can be found with utility analysis. The shape of indifference curves is usually taken to be that of the curves shown in Figs. 5–8—i.e. convex to the origin. This implies that as his stock of one commodity diminishes a consumer will require greater and greater quantities of the other commodity to compensate him for sacrificing further units of it. This *diminishing marginal rate of substitution*, as it is called, is the indifference-curve equivalent of the law of diminishing marginal utility. If it was desired to generalise this tendency from the case of only two commodities to the case of all commodities, then 'all other commodities' (i.e. total income) can simply be substituted for one of the commodities along the horizontal or, more commonly, the vertical axis.

In this way, then, the major conclusions about consumer behaviour—diminishing marginal satisfaction and equi-marginal returns—can be established without any employment of the rather nebulous concept of utility. They all follow from the basic assumption that if a consumer chooses one group of commodities rather than another, the former gives him greater satisfaction, and, conversely, that if he is indifferent as between two groups of commodities they give him equal satisfaction. Unlike utility analysis, the rest of the analysis can be based on objective evidence—i.e. that of the action taken by consumers in different situations.

2. THE ANALYSIS OF PRICE EFFECTS

With the aid of indifference curves a demand curve of the familiar kind can be constructed. Suppose that a consumer has a constant income equivalent to OP apples and that his indifference map is similar to that shown in Fig. 9. When the relative price of oranges is such that a quantity OA can be bought with OP apples, the budget line will be PA which just touches an indifference curve at point X_1. The consumer will therefore purchase OM_1 oranges.

Now suppose oranges become cheaper, so that a quantity OB of oranges can be bought for the equivalent of OP apples. The budget line swings across to PB, which just touches an indifference curve

at X_2. At this point the consumer will purchase OM_2 oranges. Similarly, if oranges become even cheaper, so that a quantity OC can be bought with the equivalent of OP apples, the budget line becomes PC and the consumer will purchase OM_3 oranges.

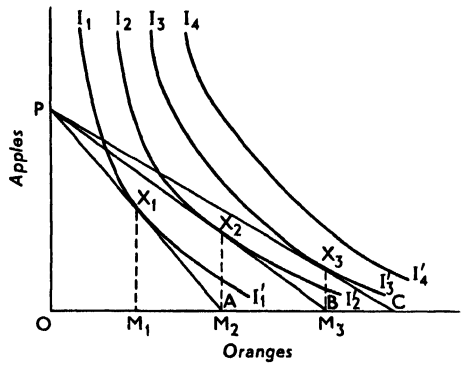


FIG. 9

Theoretically, by continuing in this way the quantity of oranges purchased by the consumer at each level of relative prices can be discovered. If the results so obtained for every consumer in the community are added together, the total demand for oranges at each relative-price level can be calculated. In this way a market demand curve for oranges can theoretically be built up.

3. THE ANALYSIS OF INCOME EFFECTS

A final illustration of the use of indifference-curve analysis is provided by considering a change in the income at a consumer's disposal. Suppose the initial situation is one in which the consumer has an income equivalent to 20 apples or 30 oranges (i.e. oranges have an exchange rate of 3 for 2 apples) and an indifference map similar to that in Fig. 10. From this map it can be deduced that in the initial situation he will buy 15 oranges and keep 10 apples. When the consumer's real income increases to 26 apples, the budget line has an unchanged slope, since a constant relative price of oranges in terms of apples is assumed, but it shifts upwards and to the right. It now touches $I_2I'_2$, and at that point the consumer will retain 12 apples and exchange the other 14 for 21 oranges. Let his

income increase again to 30 apples (=45 oranges). He will now attain curve $I_3I'_3$ and will exchange 16 apples (keeping 14) for 24

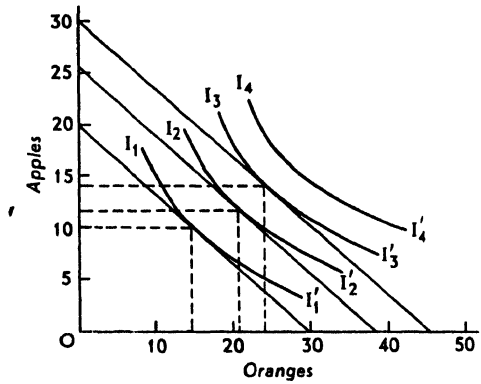


FIG. 10

oranges. Setting these results out in the form of a schedule gives:

Income (in terms of apples)	*Demand for:* *Apples*	*Oranges*
20	10	15
26	12	21
30	14	24

and by continuing in this way the effects on demand of changes in income with a constant level of relative prices can be traced out.

What of the case of inferior goods? Here an indifference map of the kind illustrated in Fig. 11 showing a consumer's preference as between, say, cake and bread is encountered. As the consumer's total income rises he moves from curve $I_1I'_1$ to $I_2I'_2$ and then to $I_3I'_3$. Over this range his consumption of both bread and cake increases. After this point (at which he is presumably feeling well-fed), the 'perverse' income effect begins to be felt. A further increase in his total income, bringing him on to curve $I_4I'_4$, results in a fall in his consumption of bread from OX_3 to OX_4, and a relatively large increase in his consumption of cake. Similarly when he moves on to curve $I_5I'_5$ his consumption of bread falls even further, to OX_5.

All the major conclusions arrived at through the use of the con-

cept of utility, therefore, can be established through indifference curve analysis, which is based on the actual evidence of preferences revealed by consumers in the market and assumes nothing more

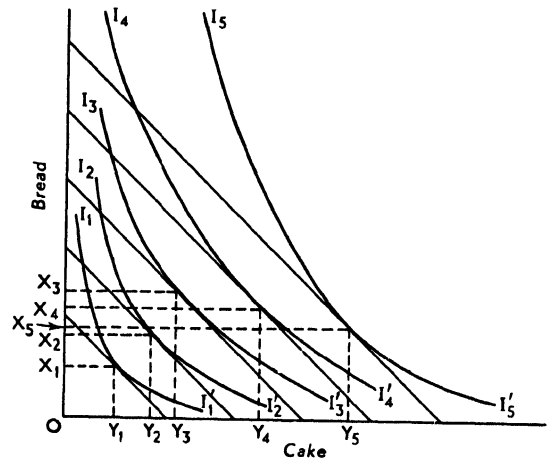


FIG. 11

than that human choice has a rational basis. Naturally this strengthens the analysis considerably. The economist need have no fear that all his conclusions are based only on the vague and immeasurable concept of utility; indifference analysis does not require to attach a quantitative magnitude to the satisfaction derived by consumers from commodities, but simply accepts the preferences expressed in the market place.

FURTHER READING

* J. R. HICKS, *Value and Capital* (2nd ed., Oxford University Press, London, 1946), Part I, Chs. I–III.

CHAPTER 5

Elasticity of Demand

1. THE CONCEPT OF ELASTICITY

In a previous chapter the discussion centred on the reasons why a 'normal' demand curve is drawn sloping downwards from left to right, reflecting the basic law of demand that, other things being equal, more of a commodity will be purchased at a low price than at a high price. It is now necessary to continue the discussion a little further so as to decide the meaning and significance of the *rate* at which a curve slopes downward—i.e. to discover why some demand curves may be relatively steep and others nearly horizontal.

The steepness of the demand curve at a particular point reflects what is called the *elasticity* of the demand for a particular commodity at that price; by elasticity is meant simply the degree to which demand responds to price. Consider the demand curves shown in Figs. 12 (*a*) and (*b*). In Fig. 12 (*a*) the curve is fairly steep, so that a

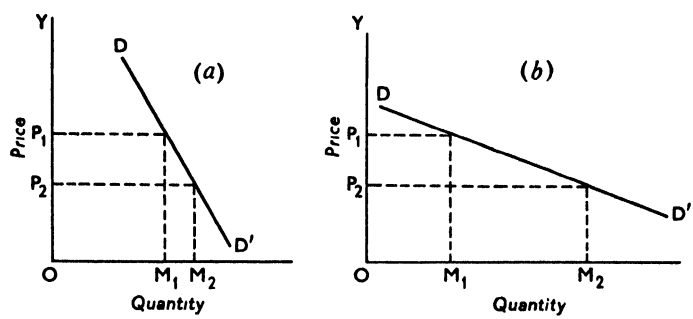


FIG. 12

fall in price from P_1 to P_2 results in a relatively small increase in the quantity purchased, from OM_1 to OM_2. Contrast this with the

curve shown in Fig. 12 (*b*); here the curve slopes much less steeply, and a similar fall in price, from P_1 to P_2, results in a relatively large increase in the quantity purchased, from OM_1 to OM_2.

The concept can be made a little more precise than this: the expressions 'relatively small' and 'relatively large' are very vague after all, and convey little. The simplest way in which elasticity of demand can be measured in unambiguous terms is by reference to the total amount of money which consumers spend on any given commodity at different prices. This total, of course, is the product of the number of units purchased and the price at which they are sold. When the price of a commodity falls, the total amount of it sold will normally increase. If the increase in sales is more than sufficient to offset the fall in price, then the total amount of money spent on it will *increase*, and demand is then said to be elastic. If the increase in sales is relatively small, however, and insufficient to offset the reduction in price, the total expenditure on the commodity by consumers will *fall* as the price is reduced, and demand is then said to be inelastic. Finally, if the increase in sales is just sufficient to offset the reduction in price, so that the total amount of money spent on the commodity remains constant as the price falls, then elasticity of demand is said to be equal to unity. Thus if the quantities of a commodity sold at various prices were as follows:

Price (*Shillings*)	*Sales* (*000s*)	*Total Amount spent* (*000s.*)
10	10	100
9	12	108
8	14	112
7	16	112
6	17	102

elasticity of demand would be said to be greater than unity at the prices of 10s. and 9s., equal to unity at 8s. and less than unity at 7s.

A rather more refined, but essentially similar, procedure is to express the actual change in the quantity purchased as a *proportion*, and similarly express the change in price as a proportion. Elasticity of demand can then be defined very clearly in terms of these proportions. If the two proportions are *equal* (e.g. a 10 per cent reduction in price causes a 10 per cent increase in the quantity purchased) then dividing the relative quantity change by the relative

price change gives an answer of 1, and elasticity can be spoken of as being *equal to unity*. If the quantity purchased increases by a smaller proportion than that of the reduction in price (e.g. quantity purchased increasing 5 per cent when price falls 10 per cent) then elasticity is said to be *less than unity*, and by dividing the proportionate change in quantity by the proportionate change in price its numerical value can actually be calculated (0·5 in the example). In this respect, the method is superior to the total outlay method, since the latter indicates only whether demand is greater or less than unity; it cannot measure *how much* greater or less. Similarly, an elasticity *greater than unity* is spoken of when the proportionate change in quantity purchased is bigger than the proportionate change in price. (It should be emphasised that the two methods will yield identical values for the elasticity of demand only if the price change being considered is *infinitesimally* small.)

It follows from all this that the fact that a demand curve has a constant slope does *not* imply that elasticity of demand is the same throughout the length of the curve, since the measurement of elasticity involves the *initial* price and quantity purchased as well as the *changes* in each. For example, suppose that the price of a certain commodity whose demand curve is a straight line falls from 10s. to 9s. and that the quantity purchased increases from 50 to 100 units (see Fig. 13 (*a*)). The formula

$$\frac{\text{Proportionate change in quantity purchased}}{\text{Proportionate change in price}}$$

gives an elasticity of $50/50 \div 1/10$, which equals 10. If the price now falls a further 1s., the quantity purchased rises from 100 to 150, so that the elasticity of demand at that price is given by $50/100 \div 1/9$, which equals $4\frac{1}{2}$. By continuing in this way it will be found that the elasticity of demand is different at every point on the curve—when price falls from 2s. to 1s., for example, the elasticity of demand is found to be only 0·22. It follows, then, that elasticity can be spoken of only as the elasticity at a particular point on the curve (i.e. at a specified price) or over a specified range of prices.

It is possible, however, to conceive of a demand curve which has the same elasticity throughout its length. Such a curve would have the shape of that shown in Fig. 13 (*b*); it is what the geometers call

a rectangular hyperbola, which is a curve such that the rectangle formed by abscissa, ordinate and the axes (e.g. $OP_1Q_1M_1$ or $OP_2Q_2M_2$) has a constant area at all points of the curve. The area of this rectangle, of course, is equal to the product of the lengths of two sides, one of them measuring the price and the other the

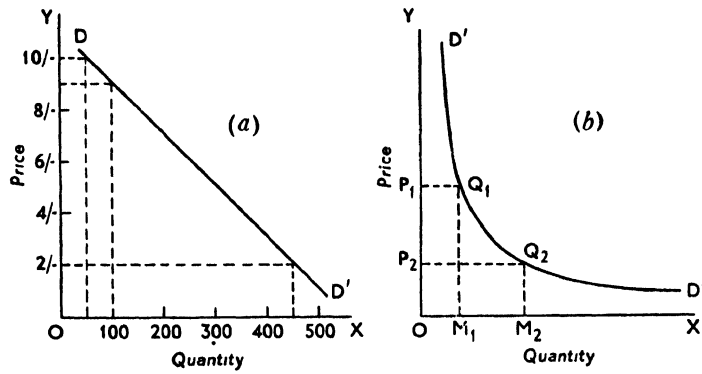


FIG. 13

quantity demanded. It therefore measures the total amount spent on the commodity at any price, since the number is being multiplied by the price at which each is purchased. The rectangular hyperbola is therefore a demand curve which indicates that the total amount spent on the commodity is unaffected throughout by a change in price; as has already been seen, this means that elasticity of demand is everywhere equal to unity.

Summing up, it can be said that the concept of elasticity of demand is concerned with the relative responsiveness of the demand for a commodity to changes in its price. Two limiting cases can be formulated. If a change in price has no effect whatever on the quantity purchased, then demand is said to be infinitely inelastic at that price. If a small change in price has an infinitely large effect on the quantity purchased, then demand is said to be infinitely elastic at that price. It is a useful exercise to draw a graph illustrating these two extreme cases.

2. FACTORS GOVERNING ELASTICITY OF DEMAND

Having discussed the ways in which elasticity of demand may be measured, it is necessary to consider the factors which tend to make

the demand for a particular commodity more, or less, elastic than the demand for other commodities. It is useful to bear in mind in this connection the fact that a change in the quantity of a commodity sold on a market may come about through *either* a change in the average quantity purchased by each consumer *or* a change in the number of consumers, although both effects may (and usually will) occur simultaneously. The immediate task is to discover why in some cases a change in price will have a marked effect on the average amount of a commodity consumed or on the numbers consuming it, while in other cases the effect will be only slight.

The first, and most important, factor is the extent to which fairly close *substitutes* can be obtained for a commodity. If a commodity can be easily replaced by others, then a small rise in its price will have a relatively large effect on its consumption, as people switch to other commodities which are fairly similar but have not changed in price. For example, most people regard lamb and beef as reasonably close substitutes. If lamb becomes dearer, therefore, it may not induce people who eat only lamb to buy much less of it but it will induce a good many people to substitute beef for lamb on the Sunday table. Hence, the demand for lamb will be relatively elastic.

Conversely, a commodity having no close substitutes is likely to have an inelastic demand, since it is not possible for a switch to or from other commodities to occur on any substantial scale. If the cost of sending a letter through the post increases, for example, people may be stimulated to some extent to use telegrams or telephones instead. These are not very close substitutes, however, and the falling-off in demand which will occur as a result of such substitution is unlikely to be significant in relation to the total.

This possibility of substitution is at the root of the usual assumption that the demand for necessities is likely to be inelastic while the demand for luxury goods will be elastic. By definition, a necessary commodity is one for which there is no adequate substitute, while a luxury is something which can be replaced fairly readily by something else. The adequacy or inadequacy of substitutes, however, is largely a matter of upbringing and habit, rather than a question of physical similarity or dissimilarity between commodities. A person of modest means, for example, would consider milk

a reasonably close substitute for cream, and a change in the relative prices of milk and cream would probably cause significant changes in his consumption of each. By contrast, a person accustomed for many years to drinking only cream in his (or her) coffee would probably regard milk (for this purpose, at any rate) as a very imperfect substitute and would replace cream with milk only reluctantly as cream became relatively expensive. Again, tobacco is by any physical standard a luxury; it is in no way necessary to human existence. Nevertheless the British public has become addicted to tobacco, with the result that fairly large increases in price have had only small effects on consumption. What is a luxury and what is a necessary commodity, then, turns largely on whether there are *thought* to be adequate substitutes, rather than on the physical nature and quality of commodities.

A second major factor influencing the elasticity of demand for a commodity is the number of uses to which it can be put, which is in a sense the inverse of this other factor of the extent to which it can be substituted for a particular use. Generally, a commodity which has several uses will have a relatively elastic demand because of the range of markets in which a price change can exert its effect. Electricity, for example, can be used for lighting, cooking, space-heating and motive-power. Even though a fall in price had only a small effect on its consumption for each purpose, the aggregate effect of a small increase in each might amount to a substantial change in total demand. Conversely, there is not much to be done with wheat except make flour of one kind or another from it, so that a price change must have a substantial effect on this single use if total demand is to be significantly changed. The fact that a commodity has only a single use, then, will tend to reduce the elasticity of its demand.

A third factor governing elasticity is the *relative importance* of a commodity in the total expenditure of consumers. If the purchase of a commodity forms a significant proportion of a person's total expenditure, either because its price is high or because the amount consumed is large, then his demand for it is likely to be more sensitive to price changes than if the expenditure involved had been only a minor element in his total spending. A person in the habit of visiting a cinema every night of the week, for example, is likely

o cut out one or two visits a week if cinema prices rise 50 per cent, imply because the extra cost involved is a substantial one. On the ther hand, a person who visits a cinema only once or twice a year s likely to be unmoved by the same price change. A doubling of he price of newspapers or salt is unlikely to seriously reduce lemand because, despite the frequency with which these commodities are purchased, their price is so low that the total amount pent by each person on them is relatively small. On the other and, a doubling of the price of, say, cake would have a much igger effect on household expenditure if consumption was not educed.

The discussion is moving close here, of course, to what was called earlier the *income effect*. If the expenditure on a commodity forms a large part of total expenditure, that commodity is likely to be one which makes an important contribution to the standard of life enjoyed by consumers. A rise in its price will therefore have a significant effect on the consumers' real income, and an income effect is likely to be felt—i.e. the whole demand curve will shift. This may reinforce or offset the direct effect of the change in price. Consumption of the commodity may be reduced not only because it has become relatively expensive but also because demand has shifted as a result of lower real income; on the other hand, if the commodity is an inferior good the tendency of consumption to fall with a rise in price (the 'substitution effect') may be more than offset by a shift of the demand curve to the right (the income effect). These two possibilities are illustrated in Fig. 14. In Fig. 14 (*a*) the direct price, or substitution, effect is to reduce consump-

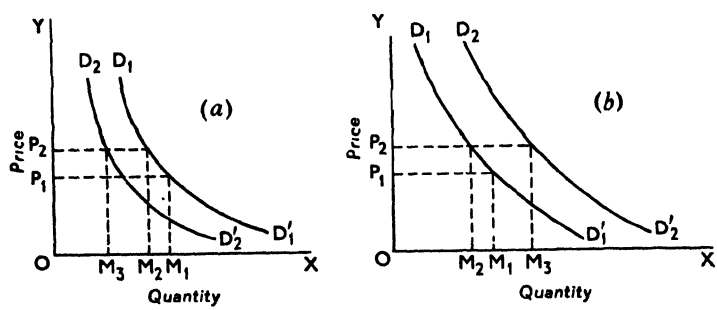


FIG. 14

tion from OM_1 to OM_2 when price rises from P_1 to P_2; the shift in the demand curve from $D_1D'_1$ to $D_2D'_2$ as a result of the fall in the real income of consumers, however, reduces total consumption even further to OM_3. In Fig. 14 (*b*) is shown the case of the inferior good; here, with a given demand curve $D_1D'_1$,the rise in price would cause consumption to fall from OM_1 to OM_2, but the income effect, illustrated by the shift of the demand curve to $D_2D'_2$, results in total consumption rising to OM_3. Thus economists speak of the *income-elasticity* of demand as well as of elasticity with respect to changes in price; the income-elasticity of demand is the responsiveness of the demand for a commodity at a given price to changes in the real income of consumers.

3. THE SIGNIFICANCE OF ELASTICITY

It is not difficult to see why a knowledge of the size of the elasticity of demand for a commodity would be extremely useful. A manufacturer contemplating an expansion of his output, for example, will need to make some estimate of the level to which his price will have to be reduced in order to sell the extra volume of goods he is thinking of producing. If elasticity of demand is less than unity, he will actually find himself selling a larger quantity of goods in return for a smaller total sales revenue, a development which he is unlikely to regard favourably. Conversely, if the elasticity of demand for his product is greater than unity, it will mean that through a relatively small reduction in price he can secure a large increase in his sales. The market research specialists are constantly trying to obtain precisely this kind of information, even though they may not speak of the concept of elasticity of demand as such.

Another case in which a knowledge of the elasticity of demand could play an important part in policy decisions is that of indirect taxation. For example, consider a commodity on which the government is levying a tax. If the rate of tax is raised, will total tax revenue increase? This will depend on the elasticity of demand at the ruling price. If it is less than unity, the quantity consumed will not decline in as great a proportion as the rise in the rate of tax, and total tax revenue will increase; on the other hand, if the demand is highly elastic, purchases may be reduced in a greater proportion

than the increase in the rate of tax, and the result will be that the government obtains a smaller revenue from a high rate of tax than from a low one—clearly a possibility needing to be taken into account.

In the light of all this, it would appear that, other things being equal, a government would be well advised to levy indirect taxes on commodities having an inelastic demand. In practice other things are not always equal. Many commodities have an inelastic demand, as has been seen, because they are considered to be necessary for a tolerable standard of life, and precisely because of this a government would hesitate, on social or political grounds, to make such commodities more expensive by taxing them. Nevertheless, this consideration of elasticity explains why commodities such as tobacco, wines and spirits contribute a very important share of the government's total revenue from indirect taxation.

Given the obvious usefulness of a knowledge of the elasticity of demand for different products, can elasticity be measured in practice? The answer is usually that it cannot. It might seem straightforward to discover from the statistics of production and sales of various commodities how demand has responded to price changes in the past. To translate such calculations into a demand curve, however, is another matter. An increase in consumption following on a fall in price *might* be the result of that fall in price. But it might also be due to a change in tastes (which is not measurable) or to changes in the prices of other goods (which are constantly occurring) or to changes in the income of consumers (which are also never absent). It is doubtful, therefore, if a clear distinction can ever be made between changes in consumption occurring because of a movement *along* a given demand curve (which is what price-elasticity involves) and those occurring because of *shifts* in the demand curve. Nevertheless, most producers usually have a fairly shrewd idea of the likely effects of a small change in their current price, although their ideas will admittedly be rough and ready and based on 'hunch' or instinct rather than statistical calculation. In any case, the fact that elasticity is seldom accurately measurable in practice does not make it a useless or unrealistic concept for the analysis of market behaviour. By employing it, economists are merely using in a clear and precise way a concept

which plays an important part, although in a less well-defined manner, in the decisions of producers and the determination of market price.

FURTHER READING

MARSHALL, *Principles of Economics*, Bk. III, Ch. IV.

* H. SCHULTZ, *The Theory and Measurement of Demand* (University of Chicago Press, Chicago, 1938), Ch. I.

PART III

THE THEORY OF PRODUCTION

CHAPTER 6

Labour

1. INTRODUCTORY

Economic activities can be broadly classified into consumption, production and exchange. The preceding Part was concerned with the factors underlying the rate of consumption of one commodity in comparison with that of other commodities. The next problem to be dealt with, therefore, is that of analysing the forces which determine how wealth is produced. All wealth is created, as was noted earlier, by factors of production; these are labour, land, capital and enterprise. Certain essential characteristics of each must be established before the analysis can proceed to the consideration of the way in which they come together in the creation of wealth.

Labour must be regarded as a very special agent in production, since (unlike land and capital) it is inseparable from the personality of its supplier. No producer employs labour in his factory or office; he employs men and women. Although people sometimes speak of 'hands' in an enterprise, human beings are a good deal more than mere hands. When land or capital are left unused, for example, there is merely waste; but when men and women are unemployed there is more than just waste involved—there are all the moral and social evils associated with unemployment. Although economists must for many purposes treat human labour as merely one of several factors contributing to production, it is important to remember that men cannot be regarded as a special kind of machine to be put to work and nothing more. The concentration on the productive activities of human beings, which is called labour, should not be taken to imply a belief that this is the only aspect of human economic activity with which the economist need be concerned.

2. DIVISION OF LABOUR

In considering the role of human beings in the process of production, the first and most striking fact which is to be observed in the modern economy is that of the *division* or specialisation of labour—that is, the employment of the overwhelming majority in narrowly defined occupations or even in narrowly defined tasks. There are farmers, factory workers, doctors, teachers, bus drivers, and so on; few people attempt to supply *all* their own economic needs, or even the greater part of them. People do not (most of them) try to build their own houses, grow their own food, and make their own clothes. Indeed, they specialise even within their particular occupations; the factory-worker specialises in a particular process or the use of a particular machine, the farmer usually specialises in certain branches of agriculture, the teacher specialises in certain subjects, even medical men specialise in certain branches of medicine—although only to them, for some reason, is the title of 'specialists' actually given in ordinary parlance.

Why is the human contribution to production organised in this peculiar (and at first sight rather unnatural) way? The answer is that experience has long since shown that specialisation leads to an immense increase in the amount of wealth which any given number of people can produce in any given time. There are a number of reasons for this. First, and most important, is the familiar fact that specialisation leads to a high degree of *skill and dexterity*—practice makes perfect, as the saying has it. The constant repetition of particular activities accustoms minds and muscles to the procedures involved, and enables them to be carried out rapidly and with less effort. Compare the speed and accuracy of the professional typist with the amateur, the drill of the trained soldier with that of the raw recruit, or the tennis of someone who practises frequently with that of a person who plays once in the proverbial blue moon. And increased speed and dexterity mean, in economic life, increased production and increased wealth.

A second advantage of the division of labour is the fact that it enables substantial *time* to be saved. On the one hand, a man turning from one job to another even in the same factory inevitably loses time in doing so and in settling in at the new job. On the other

hand, the time taken to train a man to become proficient at a wide range of tasks is naturally longer than that involved in training him for one particular task. An unskilled person can be trained to watch and operate a particular machine in the average factory in a few weeks; the training of a fully qualified turner or electrician or toolmaker will take six or seven years.

Thirdly, specialisation allows people to exploit their own individual *aptitudes*. Some people are naturally clever with their hands and some are the reverse; some take readily to indoor work, others cannot stand it; some move and think rapidly, some are naturally slow-moving. Through specialisation there is a greater chance, to put it no higher, that people can find jobs for which their particular talents are best suited.

Finally, by breaking up complex processes of production into separate tasks, specialisation encourages the *use of machinery* and its introduction. The use of machinery is encouraged because the breaking-up of production into separate tasks permits a person to devote his whole time to that task, producing a much larger output of components than if he were trying to carry out every stage in the process of manufacture. Since a machine is economically used only when its capacity is being employed fairly fully (which means when it is producing large numbers of units), the division of production into stages, each involving a high output, enables a machine to be economically employed. If a man made every part of a shoe himself he might produce only two or three pairs a week, and at this level of output it would scarcely be sensible to use machinery at any stage. If he concentrates only on, say, cutting out the sole of each shoe his output of soles might reach several hundred a week, so justifying the introduction of machinery to help him. What is more, the breaking-down of production into separate processes means than an engineer can develop a high degree of familiarity with every detail of a process. As a result he may be able to see ways in which tasks could be carried out mechanically—something he could not do without that breaking-down of particular processes which specialisation permits.

Division of labour, then, has considerable advantages, and the high standard of living now enjoyed by the population of a country such as Britain is probably due more to it than to any other single

factor. But society cannot afford to be unaware of its dangers and disadvantages. In the first place, this high degree of specialisation makes each person dependent on millions of other people completely unknown to him and quite beyond his control. Britain specialises in manufacturing, which means that it relies even for food, without which life is impossible, on countries thousands of miles away; perhaps it is only war which really brings home the enormity of what is involved in that. Again, if that part of the population which specialises in driving railway-engines for some reason cannot or will not drive engines, then the entire economic life of the country finds itself threatened with dislocation. Further, specialisation increases the risk of unemployment for the individual; the narrower his training and experience, the more difficult it may be for him to find alternative employment if he is forced to seek it.

More generally, there is the danger that a narrowness of specialisation will cause a narrowness of personality and outlook. A man or woman engaged all day and every day on some wholly mechanical and repetitious task finds little scope for creative impulses or the exercise of mental faculties. Can a community employed in this way on monotonous and mechanical tasks, largely under factory conditions, really develop to the full its intellectual and moral qualities? It is an important question. In answering it, many would point out that the infinitely higher living standards and shorter working hours enjoyed by the present generation in comparison with the people of, say, three hundred years ago make for a fuller and more wholesome human life than was even conceivable when men found themselves hard put to it to attain even bare subsistence. Present generations, they would say, have in return for specialisation an infinitely more varied and satisfying life than their predecessors. Ultimately, only opinions can be expressed on an issue of this kind; the important thing is that society should be aware of, and guard against, the dangers which over-specialisation in any field may involve.

3. OTHER FACTORS IN LABOUR EFFICIENCY

It would be wrong to imagine that the extent of specialisation is the only factor determining the productive efficiency of a nation's

labour force. The amount of capital with which labour is equipped, or the nature of the land on which it is engaged, naturally play an extremely important part. The significance of such factors will be considered at some length when the discussion passes to the subjects of capital and land.

A great many other considerations have a bearing on the efficiency of labour as such. Some races are endowed to a higher degree than others with certain qualities favouring economic efficiency—qualities such as adaptability, inventiveness, perseverance and energy. Physical strength as such has played an important part in the past but is diminishing in importance with the spread of mechanisation. A country's climate may be an important factor in this context, since a temperate climate tends to be more conducive to the exertions of economic life than a tropical climate, for example.

The general conditions under which a nation's labour force lives, as well as its innate qualities, will also affect its efficiency. An underfed, badly housed and inadequately clothed population can hardly be expected to provide a fully efficient labour force. Nor is it merely a question of the physical requirements of life. A first-rate labour force will develop only in conditions which offer adequate personal freedom, variety of life and scope for ambition. A slave is seldom a really efficient worker.

Many elements therefore contribute to the efficiency of a country's labour force. Some of them are scarcely within human control—innate qualities, or climate—but many can be provided, or at least encouraged, by adequate welfare provision and by careful design of education and vocational training. As the benefits of such measures accrue to the community as a whole, and seldom to a single individual or institution, the responsibility for them is generally agreed to lie primarily with the government. The use of resources in them can raise the wealth-producing capacity of a community just as surely, if somewhat less dramatically, as some revolutionary development in the field of technology.

4. THE SIZE OF THE POPULATION

The effective labour supply available to a community depends partly on the efficiency of the average unit of labour. But obviously

it also depends on the *amount* of labour available to it; since the overwhelming majority of the adult population of any country is engaged in economic activity as suppliers of labour, this amounts to saying that it depends on the size of the population.

If a nineteenth-century economist had been asked what were the factors which govern the size of the population of any country, his answer would probably have been influenced by the arguments put forward with great force in 1798 by an English clergyman, Thomas Robert Malthus. The essence of these arguments was that population tends to outrun the means of subsistence available, so that positive checks on population growth such as wars and disease were necessary (unless people themselves corrected this tendency for population to expand with undue rapidity) to confine the numbers of humanity within the limits imposed by the available food supply. Implicit in this argument was the belief that a rise in the standard of living would stimulate an increase in the size of the total population, while a fall would have the reverse effect.

History has shown that this distinctly pessimistic view of the economic destiny of the human race is not well founded. Britain, for example, has a population at the present time which is roughly five times as big as it was at the time when Malthus wrote; yet that vastly increased population is living at a much higher standard of subsistence than was the common lot in 1800. The reason is that Britain has been able to obtain supplies of food from countries all over the world which were little more than undeveloped colonies—and certainly not major food exporters—when Malthus was considering the problem of feeding Britain's population. Not only have new lands been exploited; agricultural methods and equipment have undergone tremendous development in the past century and these have raised agricultural productivity to levels which would have seemed inconceivable to earlier generations. At the same time, a tendency has emerged for people to have smaller rather than larger families as their standard of living rises (largely because the burden of raising children rises with it).

Does this experience mean that from an economic point of view society can be indifferent to the possibility of over-population? This would be far from the truth; economists conceive of an *optimum* population—that population which, given the amounts of

land, capital and enterprise available to a community, is such as to maximise output per head in that community. So long as the average output per head is rising as population increases, then the community is *under*-populated; if average output falls when population increases, the country is *over*-populated. The optimum cannot be thought of as some absolute number; it is closely related to the available supplies of the other factors of production and will change from time to time as the effective supply of the other factors changes.

Bearing in mind these considerations, it is clear that vast areas of the globe—especially in India and China, but perhaps also in certain parts of Eastern and Western Europe—are over-populated, in this sense that average output would rise if the amount of other factors was raised in relation to the amount of labour. This could be done, *either* by reducing the size of the population, holding the supply of other factors constant, *or* by increasing the amounts of the other factors which are combined with the existing labour force.

In Britain, and in many other Western countries, a very different prospect is faced—i.e. that of a declining population—and this possibility has given rise to serious concern. Why is this? Surely a smaller population would mean that there would be more wealth available, and therefore a higher standard of life, for each individual? Unfortunately the matter is not as simple as that. A declining population gives rise to a number of difficulties. The first, and most obvious, is that a fall in the size of the population will raise average output per head only if the country was over-populated (in this economic sense) in the first place, which is by no means always certain. Secondly, a falling population implies an increase in the average age of the community; since the birth-rate is lower than the death-rate, the older age-groups tend to become a relatively larger element in the population. This ageing can have serious effects. Since older people are generally less adaptable than the young, the economy may become less able to absorb the constant change and adjustment involved in technological development. Further, mobility of labour is likely to decrease; shifts of labour from a declining to a growing industry are effected mainly through a cessation of recruitment to the one and increased

recruitment to the other, and this process will inevitably be hampered if the number of young people entering employment is diminishing.

A third disadvantage of a declining population is that the smaller size of the market and of the labour supply may result in the loss of the gains which, as will be seen later, normally arise from large-scale production. (The rapid economic development in Britain during the eighteenth and nineteenth centuries was only made possible by the rise in the country's population.) Fourthly, a population in which the 'active' age-groups (i.e. people of working age) were becoming proportionately less important, and the retired or dependent section becoming more important, would be faced with a serious burden of welfare services. The provision of pensions, health services, homes for the aged and so on ultimately depends on the productive capacity of the employed population. If the number of old people was increasing while the size of the occupied population was diminishing, greater demands would be falling on the output of a smaller number of producers. Certainly the burden of taxation to finance these social services would bear increasingly on the active members of the community, and this might seriously discourage them in their efforts.

Fifthly, an ageing population will involve a shift in the pattern of demand away from commodities associated with the younger age-groups (houses, children's clothes and toys, schools etc.) and towards the goods and services demanded by older people (services such as entertainments and travel, and commodities like fashionable clothing, less necessary types of food, cars and so on). This would involve the problem of re-shaping the economy to meet the changing structure of demand; it also carries a danger of increased instability, since the demand for many of the commodities consumed by adults is subject to far more variation from changes in taste and fashion than the basic requirements of a young population

Finally, and allied to this, is the problem of investment outlets. A declining population implies a diminishing market and scale of production, so that the need for investment in factories and equipment will naturally be reduced. More important, a large part of investment nowadays is devoted to what is sometimes called *social capital*—assets such as houses, roads, schools and hospitals. These

things are much needed when the population is growing; a declining population will need little, if any, new capital of this kind, however, since it will inherit a stock of it which was adequate for the larger population previously existing. As will be shown at a later stage, a high rate of investment is of critical importance if a relatively rich country is to keep its resources fully employed; this closing-up of avenues of investment may therefore create some serious problems.

For all these reasons a declining population is a matter for some concern. This is not to say that it could not also bring important advantages. If the population was initially in excess of the optimum for the country concerned, naturally the wealth available per head would increase if population fell. In a country like Britain, where a large population is crowded into a relatively small area, a reduction of numbers would provide an opportunity for removing overcrowding of houses, towns, schools and the countryside as a whole. It might also help Britain's foreign trade position, since a smaller demand for food and raw materials (which Britain imports from other countries on a large scale) could be wholly reflected in a reduction of imports, home output being unchanged. In this way the proportionate reduction in imports might be much greater than that in the population. For example, consumers in the United Kingdom spend at the present time something of the order of £4,000 million a year on food, of which roughly £1,200 million is imported. If the demand for food fell by 10 per cent, so that total requirements fell to £3,600 million a year, but home food production was unchanged, Britain's import needs would fall to £800 million a year. A fall of 10 per cent in the demand for food would therefore cause a fall of one third in import needs.

In all this discussion, of course, only the economic aspects have been touched on. Population changes have significance from the strategic and political points of view; since 'population' means human beings, the subject also has fundamental moral and social aspects. In discussing these matters the economist can claim no special authority or knowledge, but obviously these considerations must not be overlooked.

5. THE POPULATION OF THE UNITED KINGDOM

It will be instructive now to examine briefly the characteristics of the British population. The trend of the total population of the United Kingdom in the past century may be seen from Table II. Between 1851 and 1911, the total rose rapidly—its decennial

TABLE II: *Population of the U.K.*, 1851–1951
(*million*)

1851	22·3
1871	27·4
1891	34·3
1911	42·1
1931	46·0
1951	50·2

Source: *Annual Abstract of Statistics, 1955* (H.M.S.O. London, 1955), Table 6.

increase averaged between 11 and 14 per cent (i.e. 1·1 to 1·4 per cent per annum). After 1911, however, the rate of growth showed a marked decline to a decennial average of somewhat less than 5 per cent, and in recent years it has fallen to 3–4 per cent. The main factors at work in this rapid increase in the second half of the nineteenth century and the slower increase in the first half of the present century have been the birth- and death-rates; on balance the net movement of people into and out of the United Kingdom reduced its population by about 4 million during the period 1871–1931, while during 1931–51 there was a net inflow of about 440,000 people immigrating into the United Kingdom. The death-rate has continuously declined (from an annual rate of 22 per 1,000 total population at the beginning of the 1870's to about 11 per 1,000 at the present time) although its decline in recent years has been less rapid than in the nineteenth century, contributing to a minor degree to the recent slowing down of the expansion in total population. The decline in the birth-rate has been of much greater influence—it fell from 35 per 1,000 in the early 1870's to a little over 16 per 1,000 in the early 1930's, and in recent years has been running at just under 16 per 1,000.

The probable future trend of population in the United Kingdom is a matter of some dispute. If the birth- and death-rates continue

at their present level, then the population would continue to grow at about 5–6 per cent every ten years. A Royal Commission on population reporting in 1949 estimated that the population of Great Britain would be somewhere in the range of 41–57 million by the end of the present century, according to which assumptions were made concerning future movements in birth- and death-rates. The official British demographers are now inclined to believe that the population of the United Kingdom will reach a peak of about 53½ million in the 1980's; although a gentle decline is probable thereafter, the population is not expected to fall much below 53 million by the close of the present century.

The change in the age and sex structure of the British population during the present century may be seen from Table III. It will be noted that there are more females than males in the British population, and the numerical superiority of the former has increased somewhat during the present century.

TABLE III: *Age and Sex Structure of the Population of the United Kingdom*, 1901–51

Age-group (years)	1901		1951	
	000s	*% of total*	*000s*	*% of total*
0–14	12,422	32·5	11,326	22·6
15–29	10,808	28·3	10,326	20·6
30–44	7,493	19·6	11,125	22·2
45–59	4,638	12·1	9,557	19·0
60–74	2,343	6·1	6,112	12·1
75+	531	1·4	1,779	3·5
Total - - -	38,235	100·0	50,225	100·0
Of which Males - -	18,491	48·4	24,118	48·0
Females -	19,744	51·6	26,107	52·0

Source: *Annual Abstract of Statistics, 1955* (H.M.S.O. London, 1955), Table 7.

This latter development probably owes something to 'artificial' causes; two wars have taken a heavy toll of men serving in the Armed Forces, while those serving overseas at the time of the

Census would in any case be excluded. The most important reason for the excess of females over males, however, is the tendency for women to live longer than men; at the present time the average newly-born child in Britain has a life expectancy of 67 years if it is male but 72 years if it is female. (The male's greater exposure to occupational strains and hazards is probably the major explanation of this.)

The age distribution is of greater economic significance. The ageing of the British population during the present century can be seen from Table III. In 1901, nearly 61 per cent of the population was under the age of 30, whereas by 1951 this proportion had fallen to a little over 43 per cent; conversely, the over-sixties accounted for 75 of every 1,000 in 1901 but 156 of every 1,000 in 1951. The significance of such a trend has been discussed at some length in the previous section; it is, of course, the inevitable consequence of a falling birth-rate combined with a falling death-rate.

The geographical distribution of the population is shown in Table IV. The most interesting feature of this table is the shift in the distribution of the population away from what may be called the extremities of the United Kingdom—Scotland and the North of England, Wales and Northern Ireland—and towards the 'central' area covered by London, the East and South of England

TABLE IV: *Geographical Distribution of the Population of the United Kingdom*, 1911–51

Region	1911		1951	
	000s	*% of total*	*000s**	*% of total*
London and East	11,206	26·6	14,004	28·0
South and South-West	4,373	10·4	5,670	11·3
Midlands	5,915	14·1	7,801	15·5
North and North-West	12,156	28·9	13,685	27·3
Scotland	4,760	11·3	5,097	10·1
Wales	2,421	5·7	2,599	5·1
Northern Ireland	1,251	3·0	1,371	2·7
Total	42,082	100·0	50,225	100·0

* Slight discrepancy in totalling is due to rounding.

Source: *Annual Abstract of Statistics, 1955* (H.M.S.O., London, 1955), Table 10.

and the Midlands. This 'central' area accounted in 1911 for just over 51 per cent of the total population; forty years later the proportion was nearly 55 per cent. This movement reflects to some extent the increasing urbanisation of the British population, and the continuous drift of population away from rural areas. In 1911, about 24 per cent of the population lived in rural areas, whereas in 1951 the proportion had fallen to 20 per cent. At the present time, indeed, little short of 40 per cent of the total British population is to be found in the seven great conurbations of Greater London, West Midlands, West Yorkshire, South-East Lancashire, Merseyside, Tyneside and Clydeside.

The shift in the geographical distribution of the population, of course, is closely connected with its redistribution amongst different industries in response to changes in the structure of the economy. The marked changes which have occurred in the occupational distribution of the British population in the past century may be seen from Table V, which refers to the occupied popula-

TABLE V: *Distribution of the Occupied Population of England and Wales*, 1851–1951.

(% *of Total*).

Industry	1851	1901	1951
Agriculture and Fishing	21·4	8·5	4·7
Mining and Quarrying	3·9	5·5	4·4
Manufacturing	32·9	29·1	34·1
Of which:			
Textiles and clothing	*22·2*	*14·1*	*7·7*
Metals, machinery etc.	*6·1*	*9·1*	*17·5*
Other manufacturing	*4·6*	*5·9*	*8·9*
Manufacture of food, drink and tobacco	1·7	2·2	3·3
Building, decorating, contracting	5·0	7·1	6·2
Transport and communication	5·4	9·7	7·7
Commerce, finance and distribution	5·3	9·0	13·8
Personal service	14·6	14·6	7·5
Central and Local Government	1·5	2·6	8·2
Other	8·3	11·7	10·1
Total	100·0	100·0	100·0

Source: Adapted from data given by E. A. G. Robinson, 'The changing structure of the British economy', *Economic Journal*, Vol. LXIV, No. 255, September 1954, p. 459, Table III.

tion of England and Wales. The decline of some traditional occupations—especially agriculture and the textile industries—is especially striking. In 1851 these two industries accounted for something of the order of 44 per cent of the total working population; by 1951 the proportion had fallen to only 12 per cent. This is a very large part of the explanation of the relative decline of the numerical importance of Scotland and the North, Wales and Northern Ireland, since these areas were especially dependent on agriculture and/or the textile trades.

Offsetting this decline is the growth of the proportion of the population engaged in industries which now play a dominant role in the British economy—and in particular the engineering trades and metal manufacturing. More people are also now engaged in what could be called the administration of output, in the realms of wholesale and retail distribution and financial and commercial services generally. The increasing importance of the government, Central and Local, as an employer is also brought out by the occupational statistics. These changes have constituted the 'pull' for population which was being pushed out of its native regions by the decline of older industries.

The United Kingdom, then, has a population which has more than doubled itself in the past century and which continues to grow. Its density of population per square mile, in fact, is one of the highest in the world amongst countries of any comparable size—527 compared with 200 in France, 358 in Italy, 50 in the United States and a mere 6 in Canada. It is a population which is ageing, however, and Britain therefore faces all the problems which that implies. Furthermore, the British population is highly urbanised and highly industrialised, and is therefore dependent for a large part of its food on supplies from abroad. That, too, as will be seen at a later stage, raises immensely important and difficult problems.

6. THE MOBILITY OF LABOUR

For an economy it is not only the size and efficiency of the labour force which matters; it is also important that labour should be used where it is most needed, and that its distribution should change smoothly in response to developments in the pattern of production and demand. This raises the question of the extent to which the

labour force has the quality of *mobility*—the power to move from one place to another, or one industry to another, whenever such movement is called for.

Mobility is of two kinds. First there is geographic, or *lateral*, mobility—the movement from one area to another. Generally, such movement is hampered by a number of factors. Human beings tend to develop an affection for their own neighbourhoods and to build up social relationships in their area through clubs, societies, unions and churches; naturally they are reluctant to move to different neighbourhoods where they will need to start rebuilding these relationships anew. The costs and inconvenience of movement constitute further obstacles, especially in conditions of housing shortage. Parents are reluctant to disrupt the education of children by removing them from one school to another, for example. Again, people are often unaware of opportunities existing outside their own district. Most often of all, sheer inertia militates against movement. When the mobility of labour as between different *countries* is considered, of course, the obstacles become greater and more numerous. Language problems arise; differences may exist in working conditions and social security systems; the costs of movement are much greater.

The second type of mobility can be termed occupational, or *vertical*, mobility—the movement of people from one industry or trade to another. The specialisation of labour sometimes means that a man cannot use his training and experience in more than a limited range of industries. Trade unions and professional associations frequently impose restrictions on entry with the aim of preventing too much competition for jobs, so holding up the level of wages or salaries. High costs of apprenticeship or education (in terms of time as well as money) may prevent people from entering an occupation for which they are well suited. Movement from one social class to another—which is frequently involved in movement from one occupation to another—is difficult, and achieved by only a small fraction of the population.

A high degree of labour mobility is obviously desirable for an economy which seeks to reap the benefits of a progressive technology or to sell in world markets where the pattern of demand is never stable. How, then, can mobility be increased? Lateral

mobility can be encouraged by an efficient and comprehensive system of employment exchanges, by government grants to assist in training and resettlement or to meet costs of movement, and by a stock of houses which leaves some margin for movement. In Britain, government policy has operated in most of these ways for some considerable time now. Yet the innate human revulsion from the uprooting of house and home remains, and the experience of the 1930's indicates that labour is reluctant to move even when faced with the prospect of prolonged unemployment in its own area. The continued housing shortage in the post-war years is an added factor of considerable importance. Direction of labour under statutory powers is one remedy, but it is one which is generally agreed to be incompatible with the democratic way of life and unacceptable for any but the most desperate situations, such as that of war.

On the occupational side, a good deal of progress has unquestionably been made in countries such as Britain through the opening of educational opportunities to all classes of the community, irrespective of means. It would be too much to say that complete equality of educational opportunity now exists in Britain, but certainly there is considerably more opportunity for the children of relatively poor parents to train for and enter skilled trades and professions than was the case even a generation previously. Restrictions imposed by trade unions and professional associations still exist, and hence prevent complete freedom of entry, while movement between social classes can never be an easy process. The strength of each of these factors, however, is perceptibly declining as time goes by, and there is a reasonable chance that in a generation or two they may have become relatively insignificant.

FURTHER READING

E. HUBBACK, *The Population of Britain* (Penguin Books, London, 1947), Parts I and III.

I. BOWEN, *Population* (Cambridge University Press, London, 1954), Chs. III–VI, IX–X.

Report of the Royal Commission on Population. (Cmd. 7695, H.M.S.O., London, 1949.)

P. SARGANT FLORENCE, *Labour* (Hutchinson, London, 1949), Chs. I–II.

CHAPTER 7

Land

1. THE SUPPLY OF LAND

When land is spoken of in economics, rather more is meant than when the word is used in ordinary conversation. Land in the sense of a physical area of the earth's surface is naturally the major element involved, because this area is an indispensable element in the production of foodstuffs and raw materials. The responsiveness of land to attempts at such production varies widely from one region to another, and this has important consequences for the pattern of economic life in various regions.

Other elements in the nature and quality of the earth's surface are of considerable importance, however. Physical environments vary in their degree of suitability for economic activity; *climatic conditions* for example, may not be conducive to high productivity in agricultural or, indeed, industrial pursuits, while on the other hand they may constitute a region's major asset, as in the case of tourist centres. Again, the *position* of land may be good or bad, in that it may be accessible or inaccessible to other trading areas, or international shipping routes, or major markets. The *superficial nature* of land is important; mountainous country is more difficult and expensive to develop than flat country in which communications are easily established.

A third set of factors is involved in what land may *contain* below its surface. The distribution of mineral wealth throughout the world has played an important part in economic progress, and will continue to do so. What is important in this context is not merely the actual quantity of mineral wealth which land may contain but the ease with which it can be extracted. Many coal seams in Britain are becoming exhausted, for example, not in the sense that all the

coal has been taken out of the ground but in the sense that the coal which remains is too inaccessible or thinly dispersed to justify the costs of mining it. Under this heading could also be included the potentialities of different areas of land for the supply of *power*; mountainous country may score in this respect even though it is less valuable in other ways. Indeed, paradoxical as it may seem, 'land' from the point of view of the economist may include the sea, inasmuch as the extraction of fish (or other things) from the sea is a form of economic activity which displays many of the characteristics of agriculture or mining.

Bearing in mind all these aspects of land from the economic point of view, it will readily be appreciated that it is difficult to distinguish in practice between those qualities of a particular area of land which are natural to or inherent in it and those which have been given to it through the application of human resources. One field may be more fertile than another not because it is in some sense an inherently better piece of land but because it has been ditched and drained more efficiently, or because more fertiliser has been used on it. At any moment of time these two contributory factors in the yield of land—the natural element and the added element—are always inextricably tangled up with one another, so that in practice it is usually impossible to attribute some of the yield to the one element and the remainder to the other.

Nevertheless it is easy to distinguish between them *in theory*, since they are distinct concepts. It is important that this distinction be drawn in economic analysis, because it is only the first type of quality—the natural and inherent powers of the soil—which arises from land itself. The qualities added to it originate from labour or capital, and must be treated as products of those factors. Having made this distinction it is possible to discuss the characteristics of land as an economic factor; there are two which are peculiar to land and this justifies the treatment of land in economic analysis as a separate factor of production.

The first of these is that, unlike the other three factors, land is *fixed in supply*. A small amount of land may be reclaimed each year from the sea, or lost to it through erosion, but the net change involved is so small that it can be neglected. This apart, the total area of land available to mankind as a whole is a fixed quantity,

although the total amount actually in use may vary from time to time. The *effective* amount of land—that is, the amount measured in terms of productive capacity—may also vary, as the fertility of land is exhausted through unwise cultivation or increased through technical development; in the 1920's and 1930's, for example, a large area in the Tennessee Valley of the United States was first reduced by inefficient cropping and erosion to a dust-bowl, and then restored to a high degree of fertility by scientific culture, irrigation and afforestation. But those qualities of land which are independent of human activity—its position, climate and whatever innate qualities exist which can be neither destroyed nor increased by any human agency—are fixed in supply. It was because of this characteristic that the law of diminishing returns, which is discussed in the next section, was first formulated in connection with land.

The second peculiarity of land is that it has no *cost of production*, another feature not shared by the other factors. To produce labour certain costs have first to be incurred—the rearing, educating and training of human beings. Capital and enterprise also have a cost of production which has to be incurred if those factors are to be brought forth. But land has no such cost for society as a whole; the fixed quantity was given to mankind by its Creator, and no-one ever needed to expend wealth or effort to bring it forth. An individual using someone else's land has to pay for it, of course; it has a cost *to him*. But this would be only a transfer between individuals, and society as a whole has incurred no costs. (All this refers only to the innate qualities of land; the qualities added by human effort naturally have a cost of production.) It follows that it is always profitable to use land, however small the return it brings; because its cost to society is zero, the return must be greater than the cost, and hence there must be a social gain. This characteristic of land, like its fixity of supply, has given rise to a concept which has subsequently been extended to other factors—in this instance, the concept of *rent* as a surplus. This concept will be discussed at length in a later chapter.

2. THE LAW OF DIMINISHING RETURNS

From history, observation and experiment it has become established that if a farmer is able to choose the amount of land he

cultivates, the land being free in the sense of having no price (as it would be for settlers in a new country, for example), he will apply his labour and capital to the area which gives him the highest average return. If to that area he subsequently applies more labour or more capital, or both, the average output per unit of labour or capital will sooner or later begin to diminish. His total output will rise, that is, but eventually *not* in as great a proportion as the amounts of other factors being employed.

This *law of diminishing returns* is best illustrated by means of a simple diagram, as are a good many economic concepts. Consider an example of a farmer producing a single crop, wheat, with a fixed acreage of land. In the diagram (Fig. 15) the number of men whose

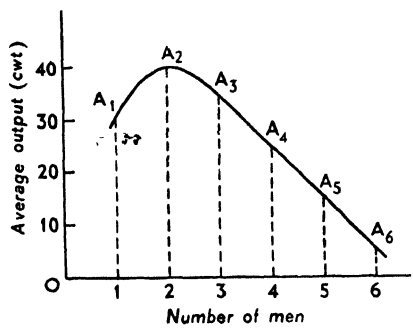


FIG. 15

labour is used on the given acreage is measured along the horizontal axis; on the vertical axis is measured the average yield of wheat per man—the total output divided by the number employed. If the farmer used only his own labour, he would attain an output of 30 cwt. (A_1). If he employed an additional man, however, believing that the area was too big for one man to cultivate efficiently, total output would be 80 cwt., giving an average of 40 cwt. per man (A_2). The average return would have risen, and the amount of labour employed would be in just the right proportion to the acreage being cultivated. If he now applied further units of labour the law of diminishing returns would begin to operate. If he employed another man, total output would rise (say to 105 cwt.) but not to the same extent as the increase in the labour force;

average output would fall to 35 cwt. (A_3). As the amount of labour increases, so the returns per unit of labour will continually decrease.

In view of the fixity of the supply of land, this feature of agricultural production was regarded by nineteenth-century economists as one of considerable importance. It implied that, if population increased, the additional food supplies required could be produced only by applying further units of labour and capital to the fixed supply of land, and therefore only under conditions of diminishing returns—less and less food would be produced per head of the population. The law was extended to other uses of land. As further units of labour and capital are applied to mines, for example, the average output will tend to fall as seams are worked out and mining has to go deeper and deeper (although mineral deposits are wasting assets in the sense that, unlike fields, their fertility cannot be maintained indefinitely by careful cultivation). *Building land* is similarly subject to diminishing returns; after a certain point, as labour and capital are applied, sites become crowded and congested and each building added reduces the average value of the existing buildings.

As a generalisation concerning economic activity, the law is unquestionably valid. Some important points about it should be noted, however. First, the formulations of the law by the nineteenth-century writers explicitly or implicitly assumed that the state of technological knowledge remained unchanged while the additional factors were being applied to the fixed area of land. If this was not the case, average output per man might well increase rather than diminish as the scale of production expanded. If six men are employed in cultivating an area which was previously worked by three men, average output may well decline *if the same methods of production are employed.* But if new techniques or equipment are developed simultaneously, the six men may well have a *higher* average output than the three. The output curve shown in the diagram would be shifted bodily upwards by the improved production methods, although its shape would remain unchanged. This is in fact what has happened in history, so that the gloomy predictions of an increasing population being driven towards starvation by diminishing returns have not been borne out.

Technical development has more than offset the law of diminishing returns, and average output in agriculture has risen rather than fallen. But in the short run, of course, when significant technical developments cannot be looked for, the law can and will come into operation.

Secondly, in formulating the law it is assumed that there is no change in the *quality* of the units of the variable factors being applied to the constant factor, and also that the fixed factor is capable of being spread over different amounts of the variable factors. If the quality of the variable factors was not constant, or if the fixed factor could be used only in conjunction with a specific quantity of the other factors, then the law might not hold.

The final point about the law is that it is not uniquely associated with the factor of land; it can, and does, apply to other factors. The nineteenth-century writers formulated the law in connection with land because of their awareness that land was fixed in supply and therefore likely to be the constant factor to which varying quantities of other factors would be applied. But under certain conditions other factors may be fixed in supply, at least for a time, and diminishing returns will be experienced if a constant amount of *any* factor is used in conjunction with increasing amounts of other factors. If a farmer commences from a position in which the right amount of land and labour are being combined and then proceeds to spread his fixed labour force over a greater and greater area by taking new land into cultivation, average output per unit of resources used will diminish in exactly the same way as when increasing amounts of labour were applied to a constant area of land. It is not land as such which gives rise to diminishing returns; it is the *fixity* of factor supply. In the long run only land is completely fixed in supply, but in the short run (i.e. as a temporary state of affairs) other factors may be fixed in supply and will thus give rise to diminishing returns. The proposition is essentially simple; it amounts to saying that if I have a bag containing six pears and six oranges and I double the number of pears in it, I do not double the total amount of fruit in the bag. Generalising the law, it can be said that in all productive activities there exists some combination of the factors of production which results in the maximum average output per unit of the resources used; *if additional amounts of factors are*

now applied to a fixed quantity of one factor, total output will increase in a smaller proportion than the amounts of variable factors being used.

3. SITES AND INDUSTRIAL LOCATION

One important quality of land, as has been noted, is its geographical position, apart altogether from its fertility. For this reason, a piece of land may have a value out of all proportion to its agricultural potentialities. An acre of land in London's West End is worth a good deal more than hundreds of acres of first-class farming land; this is because for some purposes an acre of poor land in Piccadilly is worth ten thousand good acres in Cumberland.

Site values may be derived, as in that instance, from the favourable position of the land in relation to potential demand. They may originate in the relative *isolation* of land, if residential building is involved. Most important of all, however, different areas have widely different values as sites upon which production can be carried on. Hence it is necessary to analyse the various factors which influence the value of land as an *industrial* site. Why is it that one piece of land will command a high price as a site while another, of equal or greater agricultural value, will command only a low price? What attracts industry to one area rather than another?

Probably the most important factors historically which have led industry to particular areas are those of *power* supplies and *raw materials*. The woollen industry, for example, one of Britain's oldest industries, was originally dependent on water power (i.e. before the application of the steam engine) and as a result it tended to be carried on in relatively hilly areas (such as the West Riding) where fast-flowing streams were available. Similarly, when steam power become widely adopted, manufacturing industry was attracted to regions such as the North of England and South Wales where ample coal supplies were available.

The importance of raw materials is self-evident. For industries concerned with the *extraction* of such materials, such as coal, slate and metal ores, one area will obviously have advantages over another. Such industries can be located only in areas where the minerals are deposited in the quantity and geological conditions necessary for their extraction to be worth-while. For industries

using raw materials, the advantages of some areas over others may be equally compelling. 'Heavy' industries, such as the manufacture of iron and steel, consume large quantities of coal and mineral ores, and it is advantageous for such industries to be operating close to the supplies of these materials; to transport them from the area in which they are mined to another area where they are used in production would add considerably to production costs. During the great development of these industries in Britain during the eighteenth and nineteenth centuries, therefore, the coalfields of the country became also centres of manufacturing industry. The industries consuming large quantities of iron and steel, such as shipbuilding and engineering, are in their turn attracted to the areas in which iron and steel are produced. Nowadays, when large quantities of raw materials are imported into Britain from abroad many industries find it advantageous to operate near *ports*; flour milling, for example, takes place predominantly at the ports where foreign wheat is imported, since the wheat can be taken directly into granaries from the holds of ships with a minimum of expensive handling operations.

On the other hand, there is a similar saving of transport charges if industries are close to the major *markets* for the finished product; these may not be, and usually are not, the same areas as those from which the raw materials are drawn. For an industry making goods sold to consumers rather than to other producers, it will be advantageous to be located near the great centres of population; London attracted many of the new industries making consumer goods (for example, electrical goods and furniture) during the 1920's and 1930's mainly for this reason. Hence an industry may be pulled towards different areas—one where its raw materials are to be found and another where its market lies. The outcome will depend on the nature of the product. If the raw materials used up are bulky and the finished product relatively cheap to transport (as is the case with the 'heavy' industries) the industry will tend to locate itself close to raw material supplies. If the final product is more expensive to transport, because of its bulk, its fragile nature or the need for rapid delivery, the market will exercise the greater pull—as with biscuits, or electrical goods, or bread. For industries whose product *cannot* be transported—service trades, for example, such

as hotels, laundries, garages, or professional services—the locality of the market will entirely determine the localisation of the industry.

These are all factors which may attract an industry to a particular area in the first instance. When an industry has established itself in a certain area, however, cumulative forces may be set up which tend to retain it in that area. *Subsidiary trades* may establish themselves around an industry in a particular area, so increasing the advantages of that area for a new firm setting up in the industry —marine engineering has grown up especially near centres of shipbuilding such as the Clyde, and firms specialising in the manufacture of textile machinery have been drawn towards Lancashire, where their major customers are established. This may extend to institutions specialising in the handling of an industry's raw materials or finished product, such as the Liverpool Cotton Exchange for the Lancashire cotton industry or the Bradford firms specialising in dyeing and finishing for the Yorkshire woollen industry.

Again, an area in which an industry becomes concentrated tends to develop a skilled *labour supply* with special knowledge of that industry, so that a new firm would find there a pool of skilled labour which it would be unlikely to find elsewhere. (On the other hand, in times of acute labour shortage, a firm might be attracted to an area where a labour supply is easy to obtain because there was no industry there previously.)

It is easy to see how these factors tend to be cumulative in effect. Once an industry is established, subsidiary or specialised trades are set up in the area, increasing the attractiveness of the area for that industry. The growth of a skilled labour force accustomed to the industry operates in the same way. Hence an industry may be attracted to an area by its acquired advantages long after the 'natural' forces which attracted the industry initially have ceased to operate. The town of High Wycombe became the centre of the furniture industry, for example, primarily because it was close to the beech trees of the Chilterns which originally supplied the industry with its main raw material. The industry no longer relies on local wood, but the area still attracts new firms because it contains specialists in furniture machinery and processing, a labour

force traditionally skilled in furniture making and schools and technical colleges paying special regard to training for the furniture industry.

Although these are the major forces affecting the location of industry in different areas, other factors may operate in particular cases. Good *harbour facilities* are obviously of vital importance to shipbuilding and repairing trades; the qualities of the local *water supply* and the peculiarities of the local *climate* are reputed to have determined the specialisation of Burton in brewing and Lancashire in spinning; a low *rent* may attract industries to relatively undeveloped areas; occasionally, the local availability of *capital* may attract an industry—a well-known aircraft company initially established itself in Yorkshire in the 1930's, for example, because a local landowner was prepared to put capital into the venture in an attempt to relieve unemployment amongst the people of his area. Again, the owner of a firm may be drawn to a locality for purely personal reasons of a non-economic, and even irrational, nature.

The advantages to an industry of a high degree of specialisation within an area are obvious enough. Unfortunately, the experience of the depression of the 1930's made clear that such specialisation also had grave social dangers. An undue dependence by a community on a limited range of industry can result in widespread poverty and hardship if that industry encounters bad times. If industry is diversified, a depression in one may be offset by prosperity in another, and the area as a whole will not be unduly affected. In one-industry areas there is no alternative employment if depression hits that industry, and the whole region is seriously affected.

Mainly because of this danger, government policy in Britain since 1945 has favoured the diversification of industry, especially in areas which before the war were dependent on one or two basic industries and were especially hit by the depression of the 1930's. Under the *Distribution of Industry Act* of 1945 these 'development' areas, as they are called, received special treatment. The government was empowered to build factories or to promote trading estates in them so as to attract firms to those areas as tenants; it could give financial assistance to such firms, and in the days of controls over building and raw materials such firms tended to be

favoured in precedence to those operating outside the development areas. Under various *Town and Country Planning Acts* of 1943, 1947 and 1955 power has been given to local government authorities, acting with the advice of the central government, to control the development of land throughout the country so as to prevent further industrial building in areas in which it is considered that industrial development has proceeded far enough. The problem in all these attempts to avoid the dangers of undue concentration and specialisation is that the forces attracting industry to particular areas are often powerful. If state intervention is necessary to induce an industry to locate itself in some area other than that it would itself have chosen, the implication is that as a result the industry will be located in areas where its costs of production are higher than they would otherwise have been. For goods which the community itself consumes, this higher cost represents the price of a better balance in the distribution of industry, and can be borne as such. For industries supplying export markets, however, the extra cost is more serious; foreigners will see no reason why they should pay for improvements in British industrial distribution. The crucial problem here, then, is to reconcile the need for a balanced distribution of industry with that for the highest possible level of industrial efficiency.

FURTHER READING

A. MARSHALL, *Principles of Economics*, Bk. IV, Chs. II–III.

R. COHEN, *The Economics of Agriculture* (Revised ed., Cambridge University Press, London, 1949), Ch. III.

E. A. G. ROBINSON, *The Structure of Competitive Industry* (Cambridge University Press, London, 1951), Ch. X.

A. BEACHAM, *Economics of Industrial Organisation* (3rd ed., Pitman, London, 1955), Ch. VII.

CHAPTER 8

Capital

1. TYPES AND FORMS OF CAPITAL

In a first preliminary discussion of capital in Chapter 2 it was said that capital was the wealth of a community regarded as a stock existing at any moment of time, unlike income which can be measured only as a flow over time. The essential feature of capital in practice is that its full value is enjoyed only indirectly and after it has contributed to the production of goods and services which are themselves consumed. There is one major exception to this. A house renders utility directly to its occupier, and is not wanted for the contribution it can make to the production of other goods. Why then are houses classed as capital, as they invariably are? The answer is that a house is a long-lived asset and continues to render its services for many years after it has been produced. If a house was treated as being consumed in the year it was produced (which it clearly is not) the amount consumed in that year would be violently exaggerated and consumption in later years would be underestimated since the house would not then be included but would still be entering into consumption. Strictly speaking, other durable consumer goods such as furniture and motor-cars should be treated as capital for the same reasons. Like houses, these things yield their full utility only over a period of years. The difference is only one of degree; houses generally last much longer than these other goods.

If it were possible to take a flashlight photograph of the economic wealth of a community at any moment of time, all these things *would* be included and counted in the total capital stock. In fact, this procedure is not possible year by year, and in the usual calculations of annual additions to or reductions of the total stock of capital, houses are the only commodities included which render service

directly to consumers. It is possible to estimate the total change in the value of the houses in an economy, but the task of valuing changes in the total stock of furniture, pianos, cars etc. would be a gigantic one. An increase in the number of cars used by industry for production purposes, however, *would* be counted as an addition to capital, not because a car is a durable commodity but because, like other industrial capital, cars used by an industry are valuable for the indirect contribution they make to the production of the final output of that industry and are not wanted for their own sakes.

For practical purposes, then, and not through the application of strict logic, annual additions to the total stock of capital belonging to a community are usually classified under personal, industrial and social capital. As has been seen, under *personal* capital is included only new dwelling-houses. Under the heading of *industrial* capital are included fixed and working capital, the distinction between these two, as was seen in Chapter 2, being that fixed capital can be used many times whereas working capital is embodied directly into final output and can therefore be used only once. Fixed capital comprises the plant, machinery and buildings used by industry of all kinds for the purposes of production; working capital includes, on the one hand, stocks of finished goods in the hands of producers, wholesalers and retailers and, on the other hand, raw materials and semi-finished goods still awaiting embodiment in finished goods. Under *social* capital are included all the assets belonging to the community as a whole rather than to particular persons or industries—for example, roads, schools and hospitals.

Attempts have been made in the past to estimate the value of the total stock of capital of the United Kingdom (including major items of personal property other than houses, and including land also). Some widely accepted estimates are those of Mr. H. Campion relating to the early 1930's. These suggested a distribution of the following kind:

	£ million	%
Personal capital - - -	5,900	35
Industrial capital - -	9,000	53
Social capital - - -	2,100	12
	17,000	100

The estimate for social capital, however, includes no allowance for roads and military equipment; there is no real sense in which a value can be placed on these things. Indeed, this consideration indicates the limited reliance which can be placed on any valuations of the total national capital. After all, the value of an economic asset is the price which it will fetch in an open market, and there is little meaning to be attached to the value of roads and schools and hospitals (or indeed the entire assets of any particular industry) in this sense.

It will have been noted that in this discussion the word capital has been identified with physical assets—houses, factories, roads and so on—and no mention has been made of stocks of money or claims such as mortgages or securities. From the point of view of the economy as a whole, money or claims to money are not capital, although in ordinary speech people are inclined to speak of 'capital' when they really mean 'money'. To an individual, money is capital because it represents a claim on goods and services; the community as a whole, however, cannot enrich itself by merely producing claims on itself. Money or securities of all kinds represent claims on other members of a community, just as a debt is an acknowledgement of such a claim, and for the community as a whole such claims and debts naturally cancel themselves out. As will be seen later, money or monetary instruments are wanted not for themselves but for the goods and services which can be obtained with them, and an increase in the amount of money in a society in no way makes that society richer; only an increase in its stock or supply of goods and services can do that. But claims on *other* societies (i.e. foreign countries) *are* capital, because a society can add to the total supply of goods available to it by exercising those claims. Printing more pound notes cannot make Britain as a whole better off, because pound notes represent a claim on Britain itself; but a sudden gift of dollars *will* make Britain as a whole better off, because they can be used to add to the supply of goods available to Britain's population. A country's stock of claims on other countries, then—gold and dollar reserves and titles to physical assets abroad—is part of its capital; it is the only monetary item which can be included as part of the national capital. For the world as a whole, such claims cancel themselves out with the corresponding debts or liabilities, of course, and the world's capital stock will include only physical assets.

2. GROSS AND NET OUTPUT

If a statistician is trying to estimate how much wealth a country has produced over a given period—say a year—it is obvious that he cannot add together raw materials used to make final products *and* the final products themselves. If he did, he would be counting the same thing twice. For example, if Crusoe produces a hundredweight of wheat in a year, all of which is used to make flour which in turn is used to make, say, fifty loaves of bread, he could not say that his total output of wealth is 1 cwt. of wheat *plus* 50 loaves of bread. The bread contains the wheat, and in producing the one he has used up the other; only the bread represents the net amount of wealth he has produced during that year. Final products *and* the resources used up in making them cannot be added together if the true production of wealth is to be measured.

In the same way, statisticians cannot count as wealth produced in a given year any goods, or parts of goods, which are in fact drawn from a stock with which producers began that year. Just as a producer may use up in manufacturing a commodity resources which have been produced during the same period—Crusoe and his wheat—so he may use up resources which have been inherited from a previous period—i.e. capital. To the extent that capital has been used up, of course, a society is poorer afterwards than it was before, and this consumption of capital has to be subtracted from total output to find how much new wealth that society has created. For example, if a country starts the year with a stock of 1,000 pairs of shoes, consumes 500 pairs in the course of the year and ends with a stock of 900 pairs, its true output of shoes during that year must have been 400 pairs, even though its consumption was 500.

For capital in the form of stocks of goods this is obvious enough. In the case of fixed capital, however, it is less obvious, although the point is equally important. Say a country produces 400 pairs of shoes and ends the year with the same plant and machinery as it began; assuming its stocks are unchanged, is its true output of wealth 400 pairs of shoes? The answer is no: in the course of the year the machinery has been in use, and will have been worn out to a greater or lesser extent. Even if it has not deteriorated from this wear and tear, technical developments are almost certain to have

made it to some degree obsolete. Whichever is the cause—and usually both factors are operating—the fixed capital will be worth less at the end of the year than at the beginning; it will have *depreciated.* Before the true value of output can be assessed, then, some allowance must be made for this depreciation associated with it. Unless sufficient resources are set aside each year to replace the fixed capital which is being used up or is wearing out (the setting-aside being known as *amortisation* of capital) the society concerned will one day find that its equipment has reached the end of its life and its powers of production will be seriously affected.

The output of an economy can therefore be measured on a *net* basis or a *gross* basis, according to whether a deduction has or has not been made for capital depreciation. In practice it is difficult to measure the extent to which capital has been worn out over a particular period, especially if an allowance is made (as it certainly should be) for the intangible process of obsolescence and if the estimates are complicated by changes occurring contemporaneously in the general level of prices. (The replacement of a machine purchased five years previously may require considerably more than its original cost if the prices of capital goods have risen in the meantime.) An indication of the differences between gross and net output in the United Kingdom in recent years, however, is given by the following estimates presented in the 1956 Blue Book on National Income and Expenditure (in £ million):

TABLE VI

	Gross National product	*Capital consumption*	*Net national product (National income)*
1950	11,636	1,010	10,626
1951	12,793	1,163	11,630
1952	13,928	1,301	12,627
1953	14,858	1,368	13,490
1954	15,909	1,427	14,482
1955	16,784	1,558	15,226

It will be seen that the correction necessary to allow for depreciation —about 9 per cent of gross output—is a by no means insignificant item.

3. CAPITAL FORMATION AND CONSUMPTION

Standards of living vary tremendously amongst the different societies in the modern world, and the primary cause of this, of course, is that the productive capacities of different countries are widely different. Why should one set of people be able to produce more wealth per head than another? A good many factors can be mentioned in explanation, such as differences in the skill and strength of the labour forces, or in the natural resources available, and the contrasting social and political histories of various races. But pride of place in such an explanation would invariably be given to the fact that the amounts of capital with which different societies are equipped vary tremendously. One coal miner produces more coal than another. There may be many reasons for this; but in practice the main reason may well be that one is winning coal with a pick while the other has power-operated cutting and loading machinery to help him. In almost all productive processes machinery can carry out operations at an infinitely greater speed, and for a longer continuous period, than the strongest and most skilful labourer.

If the use of large amounts of capital is the major road to high productivity, why have not all countries acquired as much capital to assist their labour force as have the highly-developed economies? To answer this question it is necessary to appreciate what is involved in the acquisition by a community of capital equipment—*capital formation* as it is called. By definition, capital goods are such as to yield their full benefit only over a considerable period after their construction. If resources are devoted to the production of such goods it is therefore necessary for a community to resign itself to sacrificing for a time the satisfaction it could have enjoyed immediately if the resources had produced consumer goods rather than capital.

Consider a simple example. Crusoe, on his island, had a limited amount of corn. Without the slightest danger of over-eating he could have eaten the lot in the first few months. Instead, he set some aside and planted it. In order to provide the seed for planting, he had to cut down his rate of consumption, which was not pleasant; but in the fullness of time he had his reward in the form of a

good crop. The situation is essentially the same in a complex economy; if labour and other resources are used to add to the stock of capital, then the society must sacrifice the consumer goods those resources *could* have produced and which it could have enjoyed immediately. If there are unused (unemployed) resources which are put to work on capital formation—road building, or constructing factories—a society can add to its capital without reducing its consumption; even here, however, the element of sacrifice is present, since the society is still foregoing the consumer goods which those unemployed resources could have produced if society had chosen to employ them differently.

This is the crucial factor in raising the productivity of what are nowadays called under-developed countries. Present needs and satisfactions—i.e. current consumption—must be to some extent denied if capital is to be produced. But if a country is very poor it may be *impossible* to deny present needs; there is little point in providing people with capital if, in the process, they all die of starvation. This is a compelling consideration for countries possessing a democratic system of government, at any rate. If a government regards the welfare of its citizens as secondary to the greater glory of the State, however—and is in a position to disregard the feelings of its citizens—the economic development of a country may be pressed forward at an accelerated pace by holding down present consumption by brute force and disregarding whatever human suffering may follow as a consequence. Hence, a high rate of capital accumulation was achieved in the Soviet Union during the 1920's and 1930's through deliberately incurring the cost of appalling hardship and outright starvation on a vast scale. The different standards of values accepted in most Western countries, and their different political systems, of course, make policies of this kind impossible for them.

Most countries must attain a certain standard of wealth, therefore, before they can feel themselves in a position to set aside resources for capital formation; yet in many cases a standard of production sufficiently high for this can only be attained *after* they have acquired more capital. This vicious circle can be broken only by outside help from richer countries. If a wealthy country gives or lends wealth to a poorer country, then the latter can devote this aid to

capital formation without reducing its level of consumption. For example, Pakistan, say, can carry out a programme of irrigation works only by using labour previously at work producing food. Without that food many of its people would starve, so Britain and other countries lend to Pakistan the food with which the population can be fed while the capital construction is under way. When it is finished, Pakistan's food output will rise significantly—certainly by more than enough to repay the loan over a period of years.

A distinction has to be made, however, between *maintaining* the community's capital stock and *adding* to that stock. In the previous section it was argued that production uses up capital just as it uses up labour and raw materials, and unless resources are set aside to replace this used-up capital the community will be consuming more than it is currently producing. If it continues to do this, of course, it will eventually find its wealth-producing capacity seriously reduced, and it will be substantially poorer as a result. Now, if the capital stock is large it may be a long time before this using-up process seriously affects production; the maintenance of a house may be neglected for years before the rain pours through the roof in sufficient quantities to make life really uncomfortable. If there is a sufficiently grave reason, then, a community may for a number of years consume more than it is really producing, through the expedient of not using resources for the maintenance of capital but using them, instead, to produce goods which are immediately consumed. During a war, for example, when the outcome of the year or two years ahead is the vital consideration and the more distant future can be left to look after itself, the maximum contribution to the war effort will be secured by using on war production resources which should have been devoted to maintaining the nation's capital intact. In due course all this neglect of capital will need to be made good, and this will then involve a greater sacrifice of current consumption than would otherwise have been necessary. But the assumption is that in the meantime the war will have been won, and that is the vital consideration. That is partly why the British public had to endure rationing and shortages long after the second World War; Britain had to make good the capital depreciation which had occurred throughout the war years.

Just as economists distinguish between gross and net output to

allow for the real but usually unseen element of capital consumption in the course of production, so they distinguish between gross capital formation (or *gross investment*, as it is often called) and net capital formation (or *net investment*). The total use of resources in making capital goods is gross investment. Some of this investment, however, is merely replacing the capital the economy is currently using up (i.e. depreciation) and is not *adding* to its capital at all. Only by deducting this element from total investment is it possible to get a true indication of the amount by which the stock of capital has been *increased*; this amount is called net investment. It is the rate of net investment which indicates how rapidly a community is adding to its productive capacity. In the United Kingdom in recent years, it is estimated that net investment has amounted to only 40–45 per cent of the total amount spent on capital goods (i.e. gross capital formation).

One important point to remember is that one element of capital comprises claims on other countries—claims in the form of either foreign currencies (or gold, which will purchase them) or the ownership of physical assets (mines, plantations, factories etc.) in foreign countries. Adding to such claims constitutes investment, just as much as adding to the stock of plant and machinery at home. Investment may therefore be *domestic* (adding to capital within the country) or *foreign* (adding to claims on other countries). This latter type of investment has been of especial importance in the history of the United Kingdom.

It is clear from what has been said that a high rate of investment is a necessary condition for increasing economic strength; at the same time, it involves the sacrifice of present consumption, which is never a pleasant experience. How is this conflict between future wealth and present enjoyment resolved? In a country such as Britain, where capital is owned by both the State and private individuals, it is the outcome of decisions taken by both sets of people. The government (including nationalised industries) decides how much should be set aside for the increase of socially-owned capital; the owners of private businesses decide how much they will add to their capital, guided largely by the price of capital in relation to the wealth they expect it to bring in (a process which will receive a good deal of discussion at a later stage). These decisions concerning

investment play a crucial role in determining the overall level of prices and employment, and this too will need to be investigated in detail later on.

FURTHER READING

MARSHALL, *Principles of Economics*, Bk. IV, Ch. VII.
HENDERSON, *Supply and Demand*, Ch. VIII.
* F. A. HAYEK, *The Pure Theory of Capital* (Macmillan, London, 1941), Part I, Chs. I, V and VI.

CHAPTER 9

Enterprise

1. THE FUNCTIONS OF THE ENTREPRENEUR

A pack of cards thrown on a table will not deal itself into hands; a pile of eggs, butter and flour thrown into a basin will not form itself into a pudding; similarly, a collection of land, labour and capital will not form itself into a wealth-producing relationship. For any group of individual entities to be placed in some purposeful and systematic association with one another, a controlling intelligence must be at work. For cards, it is provided by the dealer, for the pudding, by the cook; for the production of wealth, by the *entrepreneur*. This last word is perhaps a strange choice (its translation into English is particularly misleading) but it has now an accepted place in the technical jargon of economics. By the entrepreneur is meant the person (or group of persons) who brings the other factors of production into an organised relationship, decides on the quantities of each factor to be employed, and on the produce which they are to be used in making. He provides the factor called *enterprise*.

There would be a great deal to be said for treating this factor, enterprise, as a form of labour and the supplier of this factor, the entrepreneur, as merely one special type of labourer. Labour has been defined as any exertion of the human mind or body directed to the production of wealth, and this process of assembling and organising factors certainly comes within that definition. For the purposes of analysis, however, it is helpful to treat this particular form of labour as a separate factor (just as it is helpful to treat land separately rather than as a form of capital) because of the rather special forces which operate on it and of the distinct role which the entrepreneur plays in the economic system of a country.

The essential point about the entrepreneur is that he incurs the

cost involved in obtaining the services of the other factors, which is known beforehand, in the hope of a return from the sale of the product, a return which lies in the future and is uncertain. The other factors—land, labour and capital—are engaged for a specific rate of payment which is agreed before their services are used; the entrepreneur has no such guaranteed return, since his judgement as to what a product will cost to manufacture and what it will fetch when it is sold may prove to be wide of the mark. In this possibility lies one of the characteristics which distinguish enterprise from the other factors; although the other factors may receive a small reward, only the entrepreneur can receive a *negative* reward—that is, make a net loss. If the judgement of an entrepreneur proves to be defective, he may end not only with no payment in return for his exertions but—if the costs incurred exceed the price at which a product has to be sold—with less wealth than that with which he began, some of it having been lost in his unsuccessful productive venture.

The function of an entrepreneur, therefore, is to shoulder the risks and uncertainties involved in assembling the other factors in an efficient productive relationship and in making the decision as to the type of commodity those factors should produce. Now, if the world was a static sort of place where the amounts of various goods and services which people consumed remained more or less the same from month to month, and where production methods also remained unchanged, this job of organising and directing production would be a matter of administrative routine. In fact the world is not like that at all. The demand for different commodities is constantly changing as fashion and habit change, or as people find themselves worse off or better off, or as new products or new substitutes come along to tempt the consumer away from established buying habits. Modern production requires resources to be laid out (i.e. costs to be paid out) months or even years before the finished article reaches the shop window, and the task of anticipating correctly how many units of a product will be sold, and the price they will fetch, four or five months hence is by no means an easy one. Nor are production techniques constant; an entrepreneur may buy machinery and organise his productive flow only to find a few months later that a new device has been invented, or new process discovered, which

makes his methods out-of-date and expensive in relation to the methods now open to his competitors. The primary job of the entrepreneur in a dynamic world is therefore to assess and bear the *uncertainty* inherent in the modern economy—uncertainty as to future demand, prices and productive methods. It is possible to insure against certain business risks which operate with sufficient regularity to permit an insurance company to calculate their probability—losses such as those occurring through storms, fires, or theft —but the uncertainties attaching to future demand or supply conditions are of their nature unpredictable, and the entrepreneur must undertake the responsibility of assessing and bearing them.

2. OPTIMUM FACTOR COMBINATION

The entrepreneur, then, has to decide what is to be produced and in what quantities; if his judgement proves to be wrong, the reward he receives will be small and may even be negative. Having decided what to produce, however, he still has the problem of the way in which the commodities are to be produced. Very few commodities have only a single method of production; almost all can be produced in a large number of different ways. Catching fish might seem as straightforward a form of economic activity as any, but the methods by which fish can be caught range from, at one extreme, the man with a rod and line (i.e. much labour and relatively little capital) to the modern trawler with complex radar and echo-sounding equipment (i.e. much capital and relatively less labour) at the other.

On what basis does the entrepreneur decide the factor-combination with which his commodity should be produced? It is assumed that the entrepreneur, like the owner of any other factor of production, seeks to obtain the maximum possible monetary reward from his economic activities. If he decides to produce a certain quantity of output and to sell at a certain price, he will seek the factor-combination which results in the production of the goods at the lowest *cost*. His reward consists of the difference between what he has to pay for the factors he uses (i.e. the cost) and what the product brings in; the lower the cost, the greater his reward.

This combination of factors which results in the production of a commodity at the lowest cost per unit is called the *optimum*, or best,

combination. It is the collection of land, labour and capital which, with the directing and organising control of the entrepreneur himself, constitutes the best (i.e. cheapest) way of producing the article concerned—it is the productive method which involves the smallest sacrifice of the community's scarce resources in the production of a given amount of wealth. Naturally this optimum will involve different proportional combinations of the various factors for different types of commodity; some commodities are produced most cheaply with a lot of land and relatively little labour and capital, while others are best produced with large amounts of capital but relatively little land and labour. The optimum combination for any given commodity will also vary from time to time as new techniques and inventions occur. Finally, the optimum is closely related to the relative costs of the various factors; one combination may be the optimum for a particular commodity when land is relatively plentiful, and therefore cheap, but a completely different combination may become the optimum if land becomes dearer relative to labour.

3. LARGE-SCALE OUTPUT—INCREASING RETURNS

The law of diminishing returns—output per unit of resources decreasing as the scale of production increases—was first formulated in connection with agricultural production. Output per unit of resources employed is the obverse of average cost per unit of output. If on a farm 500 units of produce result in a year when ten men are employed, it can *either* be said that average output per man is 50 units a year *or* that the average cost of a unit of produce is $\frac{1}{50}$ of a man-year. The two things are the same. Hence the law of diminishing returns can be expressed in the form: after a certain point, the application of further factors of production to another factor which is fixed in supply will result in rising average costs of production.

Obviously this law follows as a matter of definition if an optimum factor-combination is postulated. If such an optimum exists (as it must, at least theoretically, although it may be difficult to discover it in practice) any increase in production secured through a different factor-combination must necessarily result in higher costs. The law of diminishing returns supposes that a farmer begins from, or reaches, the optimum factor-combination—i.e. that at

which the average cost per unit of output is lowest. If he now adds more labour and capital, but employs the same amount of land, clearly he is altering the proportions in which his factors are combined and is therefore moving away from the optimum; his average production cost must rise.

Diminishing returns (or rising average cost) arises, therefore, because a producer is moving *away from* the optimum combination of factors in expanding the production of the commodity with which he is concerned. It follows that the reverse situation—i.e. that of *increasing returns*, or falling average cost—will arise if a producer moves *towards* the optimum combination as he increases his output. As production increases, so each unit produced has a lower average cost of production if the proportions in which the various factors are being employed come closer and closer to the optimum. It is generally believed, on the basis of experience, that in manufacturing industries, as opposed to agriculture, this situation of falling average costs as output expands usually arises, and economists have come to speak of the *economies of large-scale production*, or the law of increasing returns. What are these economies which are reaped by a producer as his scale of output increases?

They are best classified under the two headings *internal* and *external*. Internal economies are those which result from the expansion of an individual firm and are enjoyed only by that firm; they may be technical, marketing or financial in nature. By technical economies are meant reductions in the cost of the manufacturing process itself. Increased *division of labour* is a major source of economies under this heading; with a small output it may not be possible to employ one man wholly on one process or part of a process, but, as output rises and the labour force grows, a greater degree of specialisation, with all its advantages, may become possible. This may apply to management as well as operatives; the small firm may have only one manager who has to concern himself with all aspects of management from labour relations to sales promotion, but a big firm may be able to employ specialists in these activities. A second possibility is the increased use of *machinery*; a high-speed drilling machine, for example, would be largely idle in a firm whose weekly output could be handled by it in an hour, and it would therefore be uneconomic; its possibilities might be fully exploited,

however, in a bigger firm. Not only will the use of machinery be more widespread in a large rather than a small firm; frequently, equipment is cheaper to construct and operate (per unit of output) as its size increases. Doubling the area of the sides of a furnace, for example, will result in considerably more than a doubling of its capacity. A large lorry needs no more drivers than a small one, and so on. Further, a large plant may be able to take advantage of the *linking of processes*—that is, to arrange production in such a way that the work is carried on in a continuous sequence, avoiding the wastes involved in breaking production up into separate stages. A large steel mill, for example, may be able to convert ore into pig-iron, iron into crude steel and crude steel into finished steel strip without any break in the sequence, so reducing the transport costs and costs of heating and re-heating incurred if the process were carried on in its separate stages by smaller plants.

The *marketing economies* may be of buying or selling. On the buying side, gains will arise if the orders for raw materials are large enough to justify the employment of a skilled buyer. Buying in large quantities is in any case cheaper than in small lots, since the costs of handling and transport rarely increase in the same proportion as the quantities purchased; further, sellers frequently offer more favourable terms to important customers placing large orders. Selling, also, is generally less expensive per unit in the case of large quantities, since a selling organisation of a given size is often necessary whatever the volume of sales handled. Commercial travellers may secure a large order with the same expenditure of time and effort as that involved in securing a small order; the costs of book-keeping and administration are frequently the same for a large as for a small sale. A high turnover of goods may justify specially-designed transport facilities or sales outlets which handle the selling process with maximum efficiency and effectiveness.

The *financial economies* are fairly obvious. A big firm is usually regarded by investors as less risky than a small one, so that they are often willing to lend capital to it at a lower price. Banks are less likely to refuse accommodation to a large concern in times of credit stringency; important customers are too valuable to be lightly refused help. The costs of raising funds on the capital market do not increase proportionately with the size of the sum borrowed, so that

the average cost of borrowing in this way is smaller for a large-scale borrower.

All these economies accrue to an individual firm as it grows, and are known as internal economies. By *external economies* are meant the gains arising from the growth of an *industry* and which accrue to all firms in that industry, irrespective of their size. When an industry becomes big enough, for example, *disintegration* may occur—that is, one process may be split off and performed on a large scale by a specialist firm which can gain all the economies of large-scale operation in that particular process. All firms in the industry will now be able to have this process carried out for them at a low cost instead of attempting to meet their own needs by carrying it out themselves on a small scale and at high cost. The subsidiary trades referred to in Chapter 7 may spring up when the industry has become big enough to support them, and all firms will benefit from this. Specialist buying and selling agencies may be established; waste-products on a sufficient scale may justify the formation of firms utilising them, so reducing total costs; research and information bureaux may supply services for an industry which no individual firm could afford to provide from its own resources. Frequently such economies are associated with the specialisation of a given area in a certain type of activity, but they will not necessarily take this form.

All these economies, both internal and external, really arise from the fact that some of the factors of production are *indivisible*—they are available only in lumps, so to speak. If labour were completely divisible into small units, the optimum division of labour could be adopted however small the labour force, a fractional amount of labour being employed on each separate process. In fact, however, one man (or woman) is the smallest unit of labour which exists, and a man may therefore have to be employed on several processes so long as the scale of output is too small to justify his specialisation on one. Similarly, all the technical economies arising from the increased use of machinery are denied the small-scale producer because he cannot install a fraction of a machine; he can use the machinery only when his output is big enough to keep at least one machine in fairly full employment.

The indivisibility of factors of production lies behind both in-

creasing and diminishing returns, then. The necessity for relatively large units of factors (especially capital) in modern industry often means that the optimum factor-combination can be attained only by an enterprise operating on a large scale, while the use of a fixed quantity of one factor (land) in combination with increasing amounts of other factors in agricultural activities often results in diminishing returns. If it were possible for factors to be obtained in any required quantity, however small, then there would never be any need for an entrepreneur to adopt a factor-combination other than the optimum for his particular trade, and production would be carried on at all levels of output at the same average cost (i.e. under conditions of *constant returns*). In the real world, of course, this is seldom possible.

4. SMALL-SCALE OUTPUT

In view of the fact that large-scale industrial production often results in important economies which reduce average costs of production, how is it that small-scale enterprises survive? That they do survive is shown by the fact that in 1949 nearly 60 per cent of firms operating in manufacturing trades in the United Kingdom employed less than 11 people (although they provided little more than 5 per cent of the total employment in manufacturing). Several factors underlie this stubborn survival of small-scale enterprise. The most important is probably the limitation of the *size of the markets*; there is no point in producing on a large scale if only a small proportion of the output can be disposed of in the market. Many markets are severely limited by transport considerations—people like shops to be close at hand, bread cannot always be transported very far, garages and laundries and hotels have to be established in the areas where they are needed. Milk-processing costs might be reduced to a minimum if all the milk in the country was handled by one central establishment, but for this industry the effective market in fact is the local market and it is pointless trying to organise it on a national scale. Similarly, the production economies of one central brickworks might be very great, but these would be more than offset by the heavy transport costs involved in taking the finished product to the large number of scattered localities where it is wanted.

Secondly, the *nature of the industry* may be such as to require small-scale production; in agriculture and building, for example, activity is necessarily carried on over a wide geographical area and cannot be centralised.

Thirdly, the small firm (being less complex in its organisation and having less extreme specialisation of capital and labour) is frequently *more flexible* in its output, and can often respond more quickly to changes in demand. The change-over from one model to another in a mass-production car factory may involve the dislocation of output for several weeks or months, and will certainly involve enormous sums of capital; a small and less specialised plant could probably absorb changes in design without any serious inconvenience. In markets where fashion is liable to dominate, or where demand is likely to oscillate for any other reason, this consideration may be of considerable importance. It is obviously crucial where each finished product is made to individual requirements (as in tailoring or repair work) rather than to some standardised pattern.

Finally, forces of a technical nature are sometimes at work *limiting the continued growth* of a firm. A firm may become too complex for any single entrepreneur to control efficiently; funds may not be available to finance further growth; expensive selling and advertising campaigns may be necessary if a firm is to dispose of output by breaking into a competitor's market; and entrepreneurs are as subject to human inertia as most other people, with the result that they may not wish to expand their output (and therefore their responsibilities) even though they could increase their profits by doing so.

Small-scale enterprise survives, and is likely to go on surviving, then, despite the apparently overwhelming technical advantages of large-scale production. In any industry, and at any time, there will theoretically be an *optimum size* for firms; this will embody a compromise between all these conflicting forces—technical factors tending to favour the large-scale firm, and market (and other) factors tending to favour the small firm. It is seldom possible in practice to identify this optimum size, because it will depend on factors such as the quality of the entrepreneur (it is the amount of enterprise combined with the other factors which matters, not the num-

ber of entrepreneurs), and will in any case vary from time to time and industry to industry. All that can be said is that such an optimum can always be conceived of—the size of firm which is such that, taking account of all the forces at work, its costs of production are reduced to a minimum.

5. INDUSTRIAL ORGANISATION

Who *is* the entrepreneur? Theoretically the answer is clear—the person (or group) ultimately responsible for production policy decisions and bearing the uncertainty inherent in the process of producing in advance of demand. In practice it is by no means always easy to be quite sure who actually performs these functions. This fact is brought out by a brief survey of the various forms in which economic enterprises are organised; such a survey will also make it easier for this theoretical discussion of the factor of enterprise to be related to the actual economic world.

The simplest form of industrial structure to be found in Britain is the *one-man business*, the firm which is owned and controlled by a single individual who takes an active part in the productive process but who may employ other people to assist him. This type of firm is most often seen in trades which of their nature favour small-scale output—agriculture, retail trade, professional services etc.—but is occasionally found in some branches of manufacturing industry. Here at least there is no difficulty in identifying the entrepreneur. A similar type of organisation is that of the *partnership*, where two or more individuals own and control the firm. In general, such firms do not have limited liability—that is, there is no limit beyond which the private wealth of the partners cannot be claimed by creditors of the firm so as to secure payment of a debt—although there are exceptions to this rule. Once again the problem of identifying the entrepreneurs is not very great, since the responsibility for policy decisions and the burden of uncertainty obviously lie with the individual partners. In practice, however, some partners may take little active part in the running of the business, so that they implicitly transfer their entrepreneurial powers to the active partners.

The most important form of business structure in this country is the *registered joint-stock company*, and the vast bulk of private

industry is organised in this form. With a few special and relatively unimportant exceptions, such companies register under the provisions of the various Companies Acts, comply with the requirements of those Acts, and are granted limited liability. This means that the persons who invest funds in the company, and therefore own it—the shareholders—cannot be held responsible for any debts incurred by the company except to the extent to which they have contributed funds. For example, if a person holds shares to the value of £1,000 in a registered company, he could lose his £1,000 if the company was forced into bankruptcy, but a creditor would have no claim on the rest of the shareholder's wealth even if some of the company's debts remained unpaid.

These companies may be of two types. A *private joint-stock company* is one in which at least two but not more than 50 persons hold shares. It is private in the sense that it is not required to publish an annual balance-sheet; on the other hand it is not permitted to appeal to the public at large to contribute new capital. More important in practice is the *public joint-stock company*. Such a company is required to hold an annual meeting of shareholders and publish an annual balance-sheet, as well as to comply with a great many other legal requirements aimed at the protection of the investing public from mis-information and fraud. On the other hand, they can offer securities for sale to the general public and so attract funds for the business.

These securities may be of three main types. *Debentures* are acknowledgements of a debt amounting to a specified sum on which a fixed rate of interest is paid annually; often they carry a promise of capital repayment by or on a certain date. Persons who purchase debentures are in no sense shareholders of the company; they are simply creditors of the company entitled to the fixed annual interest payment and sometimes capital repayment on a specified date. If a company fails to meet these contractual interest or capital payments, for whatever reason, the debenture holder can have the company declared bankrupt and have its assets sold in order to regain his capital.

The second main type of security is known as the *preference share*. Unlike the debenture holder, the holder of this type of security is not a *creditor* of the company—he is a *shareholder*. This means

that he has a claim to any profits the company may make, but has no claim on the company if it fails to make any profit and is unable to pay him any return on his investment as a result. Like the debenture, however, the preference share bears a specified rate of dividend which must be paid before anyone else can share in the profits. The debenture interest has to be paid before profits are arrived at; the preference shareholder then has first claim on the profits, and until the specified rate of dividend has been paid to him no-one else can share in the profits. Some preference shares are *cumulative*, which means that if profits in any year are insufficient to enable the company to pay the full rate of dividend on its preference shares, then the deficiency must be made good in subsequent years before any other claimants to profits can receive anything. Usually, preference shareholders have no voting powers in the control of the company so long as their specified rate of dividend has been paid out.

The third type of security is the *ordinary share*. The purchaser of an ordinary share has no guaranteed rate of dividend whatever; he is simply entitled to share in whatever profits are made by the company over and above those needed to meet the prior claims of the holders of preference shares. The ordinary shareholder therefore bears the greater part of the risks of the enterprise; the return on his investment may be zero for years on end, while, on the other hand, it may be very high indeed if the company succeeds. In normal times, the ordinary shareholders attend an annual general meeting at which the policy of the company is explained by a board of directors elected by them to manage the company on their behalf. If the shareholders are not satisfied, they may refuse to re-elect the board of directors and elect new directors from amongst their number.

When an attempt is made to identify the entrepreneur in a private or public joint-stock company, the task is frequently difficult. In theory, the uncertainty inherent in economic enterprise lies unambiguously with the shareholders; so does the ultimate responsibility for the policy decisions necessary to the running of the company. In practice, however, the degree of true control exercised by the shareholders of a company is often very small; provided a company continues to pay a reasonable dividend, most shareholders are content to let the board of directors look after the affairs

of the company and do not even trouble to attend the annual general meetings. Hence the management of a large established company tends to take on an identity of its own, with the active directors and salaried managers regarding the general body of shareholders as a set of creditors to be placated but who in no real sense participate in the policy-making. Ownership and control, therefore, the two ingredients of entrepreneurship, become divorced and the identity of the entrepreneur often tends in practice to be somewhat nebulous.

This is even more so with other types of industrial organisation. *Co-operative societies*, for example, have a large number of small shareholders who theoretically control the societies but who in practice take little active part in policy-making. *Public corporations*, such as the Central Electricity Board, the B.B.C., the Gas Boards etc., or *nationalised* industries (coal-mining, railways etc.) are theoretically owned by the citizens of the whole country, but it is stretching the imagination to treat every voter as an entrepreneur in these industries.

6. THE INDUSTRIAL STRUCTURE OF THE UNITED KINGDOM

The division of enterprises as between various industries in the United Kingdom can be seen fairly easily from the employment data given in Table V of Chapter 6 above. The relative importance of different industrial sectors in the British economy is most clearly appreciated, however, by considering the contribution made by each to the country's total annual product, and data bearing on this are given in Table VII. The most important single sector in present-day Britain is that of *manufacturing industry*, a rather varied group of enterprises which accounted in 1954 for nearly 40 per cent of the total domestic product. Within the manufacturing group as a whole, the main single industries are engineering and shipbuilding, iron and steel, vehicles and textiles. As in all mature economies, the 'tertiary' sectors, concerned with the handling and distribution of goods and the rendering of personal services, are also of considerable significance—together they accounted for nearly a third of the U.K. domestic product in 1954. The primary industries—agriculture and mining—account for about 8 per cent of the current gross output of the British economy.

Each of these general groups contains a number of separate industries, which are often of very different natures. In a brief examination of the main features of the British economy, therefore, it is advisable to confine the discussion to the more important single industries which constitute the main foundation of British economic life. Despite the marked dependence of Britain on imported food supplies, *agriculture* remains an essential element in the economy, employing slightly over a million people in 1955 out of a total civil working population of about 23 million. Unlike most other industries, agriculture is to be found in every part of the United Kingdom. Britain is very much a country of small-scale farming; in 1955, about 80 per cent of the total number of agricultural holdings in Great Britain (other than rough grazing land) were of 100 acres or less in area. British farmers make little attempt to compete with foreign producers in the large-scale production of cereals for human consumption. Of the total value of its gross output in the crop-year 1955–6, less than 19 per cent represented farm crops, and a large part of this comprised cereals for use as feeding-stuffs and potatoes. About 70 per cent, on the other hand, was derived from livestock and livestock products (especially milk and eggs), while a further 10 per cent of the gross output comprised fruit, vegetables and flowers.

Under the provisions of the Agriculture Act of 1947, the government guarantees British farmers a market for their output at prices arrived at in a price-review held in February of each year; the government pays grants to farmers in respect of the difference between these guaranteed minimum prices and the prices at which the products are actually sold in the market. Certain products—notably milk—are purchased by official marketing agencies, and in these cases the products are bought from farmers at the guaranteed prices and then sold to consumers at lower (i.e. subsidised) prices. In 1955 the British government paid out a total of £238 million in respect of subsidies on agriculture and food, either in the form of direct grants to farmers (acreage payments in respect of crops, marginal hill-farming grants etc.) or through the sale to consumers of food purchased by official agencies and re-sold to consumers at a loss.

TABLE VII: *Gross Domestic Product of the U.K. by Industry of Origin*, 1954

	Value of output in £ million	% *of Total*
Agriculture, Forestry and Fishing	755	4·8
Mining and quarrying	556	3·5
Manufacturing industry:		
Chemicals	454	
Metal manufacture	485	
Engineering, shipbuilding etc.	1385	
Vehicles, other metal goods and precision instruments	1096	
Textiles, leather and clothing	860	
Food, drink and tobacco	622	
Other	982	
	5884	37·5
Building and contracting	924	5·9
Gas, electricity and water	370	2·3
Transport, communication and distribution	3130	19·9
Finance, commerce and other services	1884	12·0
Public administration and defence	1015	6·5
Public health and education	484	3·1
Other	704	4·5
	15706	100·0

Source: *National income and expenditure*, 1956 (H.M.S.O., London, 1956), Table 14. The items of stock appreciation and residual error have not been included in the total. The division of the total for manufacturing industry is based on estimates made by the author.

The mining of *coal* is another industry whose strategic importance to the British economy is greater than might be inferred from the value of its contribution to the gross domestic product; a shortage of coal can exercise a crippling effect on the whole economy, and can be relieved through imports only with difficulty and at great cost. Nearly 790,000 people were employed in coal-mining in Great Britain in 1955, producing over 220 million tons of coal. Before the First World War, coal was a major British export. Since that war its importance as an export steadily declined (due to a consistently falling demand, rather than—as in recent years—inadequate output), although in 1938 the value of coal exports still

constituted about 9 per cent of the total value of British goods exported. Since the 1939–45 war, Britain has continued to be a small net exporter of coal; such exports amounted in 1955 to little more than 2 per cent of the total value of exports.

In 1946 the entire coal-mining industry was nationalised and placed under the control of a single National Coal Board. The Board operates through nine Divisional organisations, the Divisions being divided in turn into Areas. In terms of total output, the most important Divisions are the North-East and the East Midlands (especially South Yorkshire, Nottinghamshire and Derbyshire). In these areas are the coal deposits which have been most recently developed in Britain and are the richest to work; they accounted in 1955 for more than 40 per cent of the total physical output of the industry. Next in importance come the coalfields of Durham and South Wales; these are areas which are declining in relative importance—they accounted for only 23 per cent of total output in 1955 compared with over 30 per cent in 1938. Finally come the smaller coalfields of the North and West—Lanarkshire, Northumberland, Cumberland and Lancashire especially—which together accounted in 1955 for about a third of the industry's total output.

Amongst manufacturing industries in the United Kingdom, the *iron and steel industry* is of considerable importance—in 1955 it employed about 460,000 people. It is concerned with the manufacture of not only pig-iron and steel products but also tinplate, sheet-metal, chains, forgings and other steel products. It is an industry characterised by a relatively small number of firms operating establishments on a large scale. In 1956 almost a third of the industry's total manpower was employed in 53 establishments (many of them owned by a single firm) which each had a labour force of 2,000 or more persons. Only about 13 per cent of the people working in the industry were employed in establishments having a labour force of less than 100.

This concentration of the industry into a small number of firms (in 1936, 18 companies accounted for two-thirds of British heavy steel production) has been stimulated by technical developments favouring the highly-integrated, highly-capitalistic plant operating on a large scale. In recent years production of sheet steel and tin-

plate has been dominated by the continuous strip mill, in which the manufacture of the product from raw materials at one end to finished steel at the other (stages of which were previously carried on in completely separate plants) is now carried out under one roof as a single continuous process. In 1949 the industry was nationalised, control being vested in an Iron and Steel Board which proposed to retain the separation of plants and establishments much as it had existed under private ownership. After the return of a Conservative government in 1951, however, the nationalisation Act was revoked, so that the measure never became really effective.

The production of pig-iron is fairly heavily concentrated in the Midlands, Lincolnshire and the North-East Coast; about 56 per cent of the output of pig-iron in Britain in 1955 was produced in this area. In terms of the production of steel ingots and metal for castings, the most important single section of the industry is located in the two Welsh counties of Glamorgan and Monmouthshire; this area accounted for nearly a quarter of total production in 1954. Close behind came the North-East coast, which accounted for about 20 per cent of the total, and the Scottish Lowlands, which accounted for about 12 per cent. So far as the production of tinplate is concerned, British output is concentrated almost wholly in Glamorgan and Monmouthshire, less than 5 per cent of the total being produced in other areas.

The *engineering, shipbuilding and electrical goods* industry is less easily described, since the range of activities included in it is especially great. In 1955 more than 2 million people were engaged in it in Great Britain; nearly a quarter of them were women, unlike the industries already discussed, in which female labour plays a small part. The mechanical engineering section, manufacturing commodities such as steam engines, machine tools, textile and agricultural machinery, is not one which is especially heavily concentrated, although in certain branches, e.g. weighing and sewing machines, single firms or a small number of firms possess something approaching monopoly powers. The manufacture of electrical and wireless equipment, on the other hand, is concentrated fairly highly in the hands of a few giants such as G.E.C., Associated Electrical Industries and B.T.H. The practice of price-fixing under various forms of cartel arrangement is fairly prevalent. Engineering factories

falling into these two sections are to be found in almost every part of the country, since the localisation influences are less clear-cut and compelling than in most other sectors. Nevertheless there can be no doubt concerning the leadership of the Midlands, and especially the Birmingham-Coventry-Rugby area, in these branches of British industry.

The shipbuilding and ship-repairing section, on the other hand, (which accounts for about 11–12 per cent of the total manpower of this group), is heavily concentrated in certain localities as the result of favourable geographical factors. The shipyards of the Clyde in Scotland, Merseyside and Tyneside in England and Belfast in Northern Ireland dominate the British shipbuilding industry, although significant amounts of shipbuilding are also carried on at the mouth of the Humber, on Tees-side and in Newport, Monmouthshire. The industry is largely in the hands of a small number of large firms which are frequently integrated with iron and steel plants and marine-engineering concerns which supply their requirements. In 1935 the three largest firms in the industry provided more than a quarter of the industry's total employment.

The production of *vehicles* has increased greatly in importance in the United Kingdom since the inter-war years, especially in the context of British exports; between 1935 and 1955 the share of road and rail vehicles and aircraft in the total value of goods exported from the United Kingdom rose from 5 to 13 per cent. The three main sections are the manufacture and repair of motor vehicles, railway rolling-stock and aircraft, accounting for about 51, 14 and 21 per cent respectively of the total employment in the industry. The motor manufacturing section is concentrated in the hands of a small number of firms—British Motor Corporation (the fourth largest car manufacturing organisation in the world), Vauxhall Motors and the British branch of the Ford company make between them about 70 per cent of the total output of cars in the United Kingdom. Two other firms—Standard Motors and the motor-manufacturing sections of the Rootes Group—probably account for another 20 per cent of total output between them.

All these manufacturers exhibit a high degree of integration, both forwards and backwards. Most have their own engine factories or interests in engine-manufacturing companies, and some

have interests in nominally independent companies making motor bodies, carburettors and other components. The manufacture of some major components, however—e.g. tyres, electrical equipment and meters—is carried out by separate companies which usually operate on a large scale and possess a monopoly position in their own lines. The Birmingham-Coventry area is traditionally the home of the British car industry, and remains easily the most important single centre. Other important branches exist, however, in London and the Home Counties, and especially Oxford, Dagenham and Luton.

Finally, a word is necessary concerning the *textile* industry, although, like coal and agriculture, this is a section of the British economy which has declined considerably in relative importance during this century. Its most important branches are those of cotton spinning, doubling and weaving, which in 1955 employed about 274,000 people or about 27 per cent of the industry's total labour force, the manufacture of woollen and worsted products, employing about 20 per cent of the total labour, hosiery and knitted goods (12 per cent) and the production and weaving of rayon, nylon and other synthetic textiles (9 per cent). The industry as a whole remains an important contributor to British exports, although its relative importance has declined; in 1955, about 11 per cent of the value of goods exported from Britain arose from the export of textile manufactures of one kind or another, compared with about 24 per cent in 1935.

The production and processing carried out in the cotton, wool and hosiery branches of the industry are featured by small-scale nterprise, although in certain specialised lines, such as the spining of Egyptian cotton, leadership rests with a small number of rge-scale firms. The artificial fibre section (rayon, nylon etc.) is ne having many features in common with the chemical rather than ie textile industry, and a few large organisations, such as Courulds, British Celanese and I.C.I., produce the overwhelming ulk of the total output. Nevertheless, the textile industry as a hole remains one in which the small firm predominates. In 1955, ss than a quarter of the industry's manpower was working in stablishments employing more than 500 people, and about 9 per nt worked in establishments employing 10 people or fewer. It is

an industry, of course, in which a great deal of female labour is employed; in 1955 nearly 60 per cent of the industry's total labour force were women. The cotton industry is highly concentrated in Lancashire, and the woollen industry in Yorkshire, although by no means insignificant branches of the woollen industry are located in the West of England, Scotland and elsewhere. The linen industry is mainly confined to Northern Ireland, and especially the Belfast area.

These half-dozen industries, then, constitute the backbone of the British productive system, together employing nearly 30 per cent of the civil working population of Great Britain. The industrial scene, of course, is constantly changing, and the present trend in the United Kingdom is clearly away from industries concerned with the production of raw materials or commodities having a large raw-material value content and towards industries in which the value of the finished product is high in relation to the raw materials embodied in it. The skill of its labour force remains the biggest single asset possessed by the United Kingdom in a highly competitive and emulative world.

FURTHER READING

MARSHALL, *Principles of Economics*, Bk. IV, Chs. XI–XII.

E. A. G. ROBINSON, *The Structure of Competitive Industry*, Chs. II–V, VII–VIII.

A. BEACHAM, *Economics of Industrial Organisation*, Chs. I, III–IV, VIII.

HENDERSON, *Supply and Demand*, Ch. VII.

G. C. ALLEN, *British Industries and their Organisation* (3rd ed., Longmans, London, 1951).

CHAPTER 10

The Operation of Supply

1. COST CURVES

Having discussed each of the factors which are combined for the production of wealth, it is possible to analyse the relationship between the production, or supply, of commodities and the market price of those commodities. Just as it is convenient to use a simple geometrical diagram for the analysis of demand, so the analysis of supply is made easier through the use of graphical analysis. Instead of demand curves, of course, *cost* or *supply curves* are now involved.

Like demand curves, cost curves trace out a relationship between price and quantity. The demand curve relates the quantities of a commodity *purchased* to various prices of the commodity; the cost curve connects the quantity of a commodity which enterprises are producing with its cost of production at different levels of output. At each level of output, the entrepreneur will need to purchase the services of the various factors of production; this is his total *cost*. By dividing this total cost by the number of units of the commodity which these factors together produce, the *average cost* of the commodity can be obtained. Unless the price the entrepreneur expects to obtain in the market is high enough to repay this cost, naturally he will not undertake production. In the case of demand it was noted that consumers could conceivably purchase commodities by offering other commodities in exchange for them—i.e. through the process of barter. In fact, the modern economic world finds it more convenient to use money as an intermediary. Similarly, although the owners of factors of production could barter their services in return for a share in the goods they help to produce, in practice it is infinitely more convenient for all concerned if they place the services of those factors at the entrepreneur's disposal in return for a

money payment. Hence, labour is paid for by wages (and salaries), land by rent, loan capital by interest and enterprise by profits Later discussion will be concerned with the forces determining these factor-prices.

The *average cost* curve therefore represents the cost to the entrepreneur of the average amount of the factors used up in producing one unit of a commodity at different levels of output—i.e. the total payments of wages, rent, interest and profits divided by the number of units produced. Examples of such curves are shown in Figs. 16 (*a*)–(*c*).

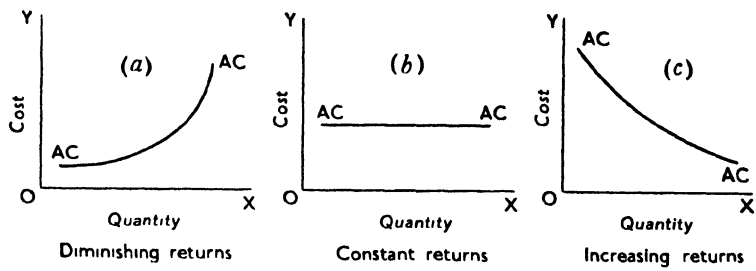


FIG. 16

One important feature of the cost curve used in economic analysis is that it reflects the costs of *all* the factors required to produce a commodity, including the factor of enterprise. Unless an entrepreneur receives a certain rate of profit from his activity, he will go to some other industry or perhaps become a wage- or salary-earner instead. Hence, a certain rate of profit—the 'normal' rate of profit for the economy as a whole—is part of the cost of production of a commodity, and is included as an element of the average cost curve. If price *equals* cost, therefore, this does not mean (as it might in everyday speech) that the entrepreneur is making no profit; it means that he is receiving the 'normal' rate of profit on each unit of the commodity.

In the discussion of demand curves, it was established that such curves would normally be falling from left to right, although their exact slope would vary as between different commodities. In the case of cost curves, no such normal rule can be laid down, since average cost of production may increase, decrease or remain con-

stant as output expands. It has already been noted that because factors are not divisible into infinitely small units, diminishing or increasing returns may be experienced. This means that the average output per unit of resources used will be falling or rising as output increases, and this is the same thing as a rise or fall in the average cost of production. Constant returns (i.e. unchanged average cost of production as output increases) will be experienced where these indivisibilities do not occur, or where a tendency to increasing returns from one of the factors employed is just offset by a tendency towards diminishing returns from another factor.

The three broad types of supply conditions—diminishing, constant and increasing returns—are portrayed by the three cost curves in Figs. 16 (*a*)–(*c*). For the sake of simplicity each curve has been drawn with the same direction (falling, horizontal or rising) throughout its length. In practice, most average cost curves will probably be more like that shown in Fig. 17—average cost falling as output expands and as the optimum combination of factors is approached, reaching a minimum at some output *OM*, where the optimum factor-combination is reached, then beginning to rise again as output expands past this level and the law of diminishing

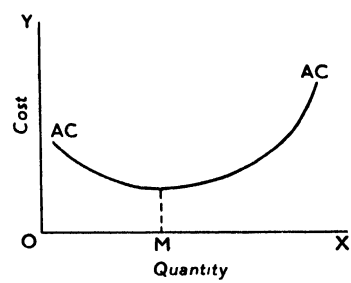


FIG. 17

returns come into effect. For example, most commodities require some minimum capital equipment, such as a factory building, however small output is. As output rises, this capital is used more and more fully until at last its capacity is reached. Thereafter it may not be worthwhile to install more capital (build another factory, say) until output is considerably greater, so that costs rise again until the

output is big enough to justify a second big 'lump' of capital and the cost curve commences to fall once more.

As with demand curves, the effects of output changes on average cost are examined by reading off from a single curve; if the basic conditions of production alter, a completely new curve has to be drawn, showing the cost of production at each level of output under the new conditions—just as a new demand curve has to be drawn whenever a shift in demand (as distinct from a change in price) occurs. For example, let the curve *AC* in Fig. 18 represent the cost curve of a given entrepreneur; when the enterprise is producing *OM* units, the average cost of producing the commodity is *OP*. Now a new method of production is discovered, or a new in-

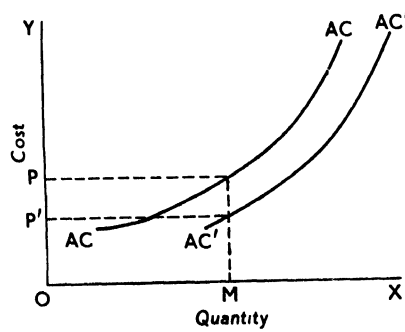


FIG. 18

vention occurs, which enables a greater output to be obtained from any given quantity of resources. The cost curve is therefore shifted downwards throughout its length, say to the position *AC'*, and at the given output *OM*, average cost of production will fall to *OP'*.

Throughout this discussion the cost curve of an individual entrepreneur, or firm, has been under consideration. Frequently the supply of an industry as a whole is of more interest—i.e. the total amount produced at different prices by all firms making the same commodity. To derive this it is necessary to draw up the supply curve of the industry as a whole, and this concept will be discussed in section 3 below.

2. AVERAGE AND MARGINAL COST

In talking of costs of production it has been assumed so far that in increasing or decreasing his output an entrepreneur can increase or decrease his expenditure on the factors of production equally easily and rapidly. In reality this is not so. Certain costs can be increased or reduced without great difficulty whenever output changes, but others cannot be changed for a considerable period, however much output may vary. This distinction, between factor-costs which can be varied whenever output changes and those which cannot, is of great theoretical and practical significance. The former are usually called prime, or variable, costs and the latter overhead, or fixed costs.

By *variable* costs are meant those which vary directly with output, even over a short period of time; the major costs involved are those of labour and raw materials. If a firm making television sets finds that demand and price have fallen, so requiring a reduction of output, it can reduce its labour force more or less immediately, so cutting its total labour costs; similarly, it will need fewer cathode-ray tubes, cabinets etc. and can reduce its purchases of these forthwith. Certain costs cannot be changed so quickly, however, and these are called *fixed* costs. These are commitments which remain at a constant level so long as the firm remains in existence at all, whatever its level of output, and can be changed only after the lapse of a considerable period—for example, the rent of the land on which the factory stands, the interest on the capital locked up in the form of plant and machinery, and the salaries of the essential managerial and administrative staff. The payments under these headings are usually fixed for long periods at a time and have no direct connection with the rate of production at all except in the long run. A striking example of the independence of such costs is provided by the publishing trade. Once it is decided to publish a book, the entrepreneur must incur the cost of having it set up in type, ready for printing; however many copies he subsequently produces, this cost remains fixed. The larger the number of copies which are in fact printed, the greater the volume of production over which this fixed (and major) element in total cost can be spread. Similarly, once the British Electricity Authority has built and

equipped a power-station, it incurs a fixed charge in respect of the cost of the buildings and equipment concerned, whatever the rate at which the station is generating electricity.

As a result, there can be, and usually is, significant difference between the *average* cost of a unit of output and the *additional* cost incurred in increasing production by one unit—i.e. the *marginal* cost. The average cost of a particular unit, of course, is the total cost incurred by the entrepreneur divided by the total number of units produced. The marginal cost of a unit of production, say the xth unit, is the difference between the total cost of producing x units and the total cost of producing $(x - 1)$ units.

Consider an example of the furniture trade; suppose that the data shown in Table VIII represent the weekly costs of producing tables. The only fixed cost is assumed to be that of an initial set of tools sufficient to provide up to 10 tables per week and costing £100. After that, the production of each table requires materials and labour, including that of the entrepreneur himself, these amounting to £50 per table for outputs of up to 3 per week. When output passes the level of 3 per week, however, the entrepreneur has to obtain his wood from more and more distant sources, so that variable costs rise progressively.

TABLE VIII: *Average and Marginal Costs of Producing Tables*

(1) *Weekly output of tables*	(2) *Fixed costs*	(3) *Variable costs*	(4) *Total costs*	(5) *Average cost* [(4) ÷ (1)]	(6) *Marginal cost*
	£	£	£	£	£
1	100	50	150	150	150
2	100	100	200	100	50
3	100	150	250	83	50
4	100	225	325	81	75
5	100	305	405	81	80
6	100	390	490	82	85
7	100	480	580	83	90
8	100	580	680	85	100
9	100	690	790	88	110
10	100	810	910	91	120

For the first table, the total cost of production will therefore be £150; since only one table is involved, this is also its average cost.

The difference between the total costs of producing this table (£150) and the total costs of producing one fewer (i.e. none at all) is £150, so that this is also its marginal cost. Thereafter, however—as Table VIII shows—total, average and marginal costs differ considerably. The marginal cost at first falls rapidly as the fixed cost is spread over an increasing number of tables, and average cost also falls, though not so violently. When output rises above that of 3 a week, however, marginal costs begin to rise. For a time the effect of this is more than offset by the gain arising from spreading the fixed cost over a larger output, and average cost continues to fall. So long as marginal cost is less than average cost, of course, the average *must* fall; a smaller amount than the average is being added to the total and naturally the overall average is reduced. Ultimately, however (in this example when output rises to 6 tables a week), marginal cost becomes greater than the average cost, and the average begins to rise; the increasingly high variable costs become more than sufficient to offset the spreading of the fixed cost. Once again, however, the average will change less violently than the marginal cost.

The relationship betweeen average and marginal costs can be shown more generally with the familiar curve diagram. In Fig. 19, which is based on the data contained in Table VIII, the curve *AC* is the type of cost curve which was discussed earlier; it shows the

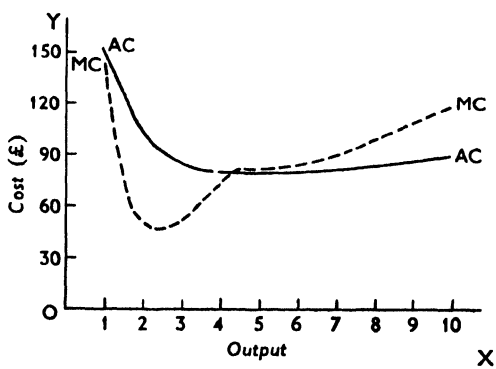


FIG. 19

average cost of production at various levels of output. On the same axes *OX* and *OY* can be drawn the *marginal cost* curve, *MC*. Cer-

tain general rules can be laid down about the behaviour of these two curves. First, so long as the marginal cost curve is below the average cost curve, then average cost must be falling as output rises. This is straight-forward enough; as has already been noted, if marginal cost is below average cost expanding output must be adding a smaller amount to total costs than the previous average, so that the new average will be smaller than before. Similarly, if marginal costs are in excess of the average it means that the average must be rising; a *bigger* amount is added to total cost than the previous average as output rises, and the new average must therefore be bigger than the old. Finally, if marginal cost *equals* average cost, then average cost is neither rising nor falling at that point—in other words, average cost is at its minimum; it has ceased falling but has not yet begun to increase.

3. SUPPLY CURVES

So far the discussion has been confined to the individual firm—the curves which have been considered show the relationship existing between the output of a particular firm and its average or marginal cost of production. For many purposes, however, the total supply of a commodity on the market may be of more interest than the output of one firm producing only a fraction of that total; in considering how the price of milk affects its supply, for example, it is the total output of milk which will be the object of interest, not the production of any individual dairy farmer. For problems such as these, it is necessary to shift the emphasis away from the firm and on to the *industry*—i.e. the group of firms which together produce the entire supply of the commodity concerned to the market. In other words, the discussion must move away from the *cost* curve of the individual enterprise and focus attention on the *supply curve* of the industry as a whole—the curve which relates the supply of a commodity to its market price.

Before considering the way in which such a supply curve can be built up, it must be made clear that the question involves certain rather difficult issues which are better dealt with separately at a later stage and which will have to be left on one side for the moment. For example, it is by no means easy to define what exactly is meant by an 'industry'; if firms are selling products which are similar but

not identical—such as Cadbury's chocolate and Fry's chocolate—are these firms in the same industry or not? How similar do products have to be before they can be regarded as the same commodity? A second, and related, difficulty is that of the relationship which is assumed to hold between the different firms in the industry—how far they are regarded as competing with one another, and how far monopoly elements exist. These issues have a great deal to do with the relationship between price and output, as will be demonstrated at length in Chapters 11–14.

In order to examine the essential nature of the concept of the supply curve, however, these complications can be left on one side. For the moment it can be supposed that there exists a group of firms whose products are more or less identical, so that they can be regarded collectively as forming a single industry. Further, let it be supposed that these firms are in full competition with one another, the output of each individual firm being so small a share of total production in the industry that changes in it have no effect on either market price or the output of the other firms. In such a situation it is clear that a change in the market price for which the commodity is being sold will affect its total supply in one or both of two ways. First, it will stimulate changes in the output of each individual firm in the industry, and secondly it will cause new firms to enter the industry (or existing firms to leave the industry).

For example, suppose the market price of a commodity rises. Firms already in the industry were finding it profitable to produce at the old, lower price. Their new rate of profit per unit of output must therefore be higher than before, and (since they are assumed to be seeking to maximise profit) they will expand production under the spur of these higher profit rates. How far will they expand output? In the previous section the cost of each additional unit of output was described as the marginal cost; so long as the revenue per unit received by a producer (i.e., in this simple case, the price at which the commodity is sold) is greater than, or equal to, marginal cost the producer is obviously increasing his total profit. When the price rises, then, each firm will expand production up to that point at which its marginal cost becomes equal to the market price. Similarly, at the previous price firms outside the industry were unable to earn as much profit in producing the commodity con-

cerned as they could in other industries (or perhaps were unable to make a profit anywhere). If the market price now rises, some of them will find that the price is high enough to cover their marginal costs, and will therefore enter the industry. Once again, they will do this only if, and so far as, price is equal to marginal cost.

An example may help to show how the total supply of a commodity will respond to changes in market price through adjustments to the output of firms in the industry, on the one hand, and through changes in the number of firms on the other. Suppose that the figures shown in Table IX represent the marginal costs at different levels of output of an industry composed, actually or potentially, of five firms. For the sake of simplicity of exposition, it is assumed that each firm has rising marginal costs at all levels of production, although (as was noted in the previous section) this state of affairs is unlikely to occur in the real world.

TABLE IX: *Marginal Costs in a Hypothetical Industry*

Units of output	*Marginal cost per unit in shillings:*				
	Company 1	*Company* 2	*Company* 3	*Company* 4	*Company* 5
1	2	4	8	12	16
2	4	8	12	16	20
3	6	12	16	20	30
4	8	16	20	32	40
5	10	20	30	40	50
6	12	24	36	48	60
7	14	28	42	56	70
8	16	32	48	64	80
9	18	36	54	72	90
10	20	40	60	80	100

Suppose the market price settles initially at 2s.; at this price only Company 1 is able to cover its costs, and even this company cannot profitably produce more than one unit of output. Since even the lowest possible marginal costs of Companies 2–5 are greater than 2s., none of them will enter the industry, and the total supply at the market price of 2s. will therefore be 1 unit.

Now suppose market price rises to 4s. At this price, Company 1 will expand its output to 2 units, since it is only at levels of output higher than this that its marginal cost exceeds the market price of

4s. Further, Company 2 can now compete in the market, since price is high enough to cover its marginal cost of producing one unit. None of the others, however, is yet able to produce the commodity at a cost low enough to sell at this price. Total supply at the market price of 4s., therefore, will rise to 3 units—2 produced by Company 1 and one produced by Company 2. If this process is continued for different market prices, a schedule can be drawn up which will show the total amount of the commodity which will be supplied at each price. The data in Table IX, for example, show that at a market price of 20s. total supply will amount to 24 units—10 produced by Company 1, 5 by Company 2, 4 by Company 3, 3 by Company 4 and 2 by Company 5.

If all these prices and outputs are now plotted in the form of a graph, the result is a supply curve for the industry concerned. An example is shown by the curve *SS'* in Fig. 20; this curve is in fact drawn from the data shown in Table IX. Such a supply curve will

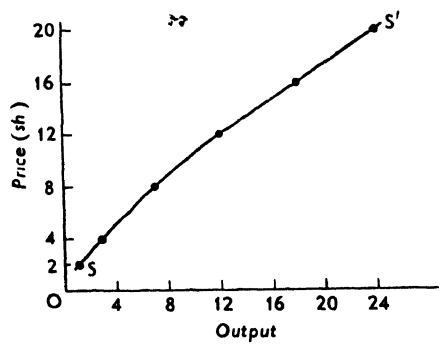


FIG. 20

invariably slope upwards from left to right, reflecting the familiar economic generalisation that a greater quantity of any commodity will be offered at a high price than at a low price. The higher price is, the more profitable production becomes, the greater the output which each entrepreneur will find it worthwhile to produce, and the greater the number of firms which will find it possible, and desirable, to enter the industry concerned.

It was noted in the previous section that, just as changes in tastes, incomes or the prices of related products could cause the

demand curve for a particular commodity to shift bodily to left or right, so a change in the fundamental conditions of production (for example, a new technological discovery) can cause cost curves to shift to right or left. Since supply curves are in effect built up from a group of individual cost curves, it follows that supply curves may shift over a period of time as the result of technological developments. Hence, if the market price ruling for a particular commodity and the quantities of it supplied by an industry were correlated *over a long period of time*, it might well be found that the supply of the commodity was greater when prices were relatively low than when price had been high. This would not mean that the supply curve existing at any given point of time was downward-sloping, however, any more than upward-sloping demand curves really exist in the case of inferior goods (see section 5 of Chapter 3). The situation is simply that a 'normal', upward-sloping supply curve has been shifting through time as the result of developments in productive methods.

Like demand curves, once again, supply curves can seldom, if ever, be constructed in reality. It is doubtful if many individual firms would be in a position to draw up curves showing their cost of production at various outputs, except in the very vaguest and most general terms; it would be statistically impossible to obtain such data on a comparable basis from all the firms in an industry (even assuming that the problems of defining an industry could be solved) so as to construct a supply curve for the industry as a whole. Such curves are nevertheless a valuable analytical device. They enable the principles underlying the relationships between price, cost and output to be embodied in a visual and convenient form, so that they can be set out against a similar representation of the factors at work on the side of demand. In this way, a large number of complex and interdependent influences can be brought together into a single analytical process, and their probable effects can be illustrated in a way which would otherwise be impossible without the use of fairly forbidding mathematics.

One final qualification is necessary. Throughout this discussion, output has been assumed to depend solely on price and cost, and supply has been identified with production. In the real world, neither procedure is really justifiable. The output of firms, or an

industry as a whole, may vary without any change in prices or costs —because of adverse weather, for example, or an epidemic amongst the labour force, or bombing by an enemy during war. Economic analysis of necessity ignores these other influences on production, simply because they are not of an economic character, and the economist has no authority to pronounce on them.

The identity of supply with production is less defensible. If prices change in the market, an important consequence may be, and often is, the accumulation or running-down of stocks of goods by wholesalers and retailers, while such changes in stocks may in turn exert significant pressure on market price. The existence of stocks, in other words, breaks the direct connection between the current production and actual supply of goods offered to consumers. If traders believe that the price of a commodity is going to rise, they may try to accumulate stocks of it in order to sell the commodity after its price has risen, while expectations of a fall in price may result in an unloading of stocks on to the market in an attempt to sell before the price falls. Whether these expectations are well-founded or erroneous, stock changes of this kind can exert a considerable influence on the available supply of a commodity (and hence on its price) without any changes at all in current output.

The complications introduced into the analysis by the existence of stocks can be ignored at this stage, however, for two reasons. First, it is obviously desirable to obtain a clear picture of the fundamental relationships between price, cost and output, and to do this it is advisable to leave subsidiary considerations on one side for the moment, even though they may be of considerable importance in practice. Secondly, most commodities deteriorate in one way or another if they are stored for any length of time, so that over a fairly long period changes in stock positions tend to cancel one another out. The supply of a commodity over the long period is determined primarily by the level of production; stock changes can exert a significant but essentially *temporary* effect on its price.

4. THE SIGNIFICANCE OF TIME IN SUPPLY

In dealing with the effects of a change in price on demand, or of changes in demand on price, it was possible to ignore the possible delay which might elapse between the change in one and the change

in the other. In practice, such delays are unlikely to be long; even if they occurred they would not materially affect the previous analysis of reactions between demand and price. In considering the supply of goods, however, delays, or time-lags, are of considerable significance; this is because in the real world considerable time may elapse before production is fully adjusted to changes in price (factories and shipyards cannot be built overnight) and also because such time-lags have important implications for theoretical analysis.

In distinguishing between fixed and variable costs a rather general reference was made to the degree to which they could be adjusted (upwards or downwards) in a relatively short period of time. To be more precise, economists distinguish between the short and long periods. By the *short period* is meant that space of time in which the productive capacity of an industry cannot be fully adjusted to a change in demand; by the *long period* is meant the space of time in which capacity can be so adjusted. For example, if the demand for transatlantic sea travel increases, shipping firms will be forced to meet the increase by working their existing ships and crews on more frequent trips and by bringing into use ships and crews which for one reason or another had previously been laid up. This will seldom be satisfactory as a permanent solution, however, and if the increase in demand is believed to be permanent new ships will be ordered and new crews recruited and trained. When the new ships and crews are finally put into service, and capacity is fully adjusted to the new level of demand, then the long-period equilibrium will be reached; prior to that time the situation prevailing will be a short-period one.

The actual length of the short and long periods in terms of calendar time will naturally vary from industry to industry. In the cotton spinning trade, for example, it may be only a matter of a few months before new equipment can be installed and additional operatives trained; in the case of air transport, five or six years may elapse before new aircraft can be delivered and additional crews be provided. A second point to note is that the long period may have different lengths in terms of calendar time according as to whether capacity is being expanded or contracted. In general, a factor such as labour can be dismissed more rapidly than it can be obtained, especially if a fairly high degree of skill is required. On the other hand, capital

wears out relatively slowly; it may take several years for a machine to pass out of use through wear and tear, whereas a new one could perhaps be manufactured in a few months. A producer faced with a reduced demand, and therefore with excess capital equipment, can reduce his capital only at the rate at which it wears out. (There will be no point in deliberately destroying it; he would obtain no benefit from that.) The best he can do is to spend no more money on it—i.e. omit to maintain it in a full state of repair and allow it to wear out gradually. This process may take several years.

There are two reasons why this distinction between the short period and the long period is important in economic analysis. The first is that it has implications for the price and output policy adopted by producers in the face of a change in demand. In Fig. 21 the curve S_1S_1' represents the short-period supply of curve of an industry; when price is OP_1, total output is OM_1. Demand now rises, so

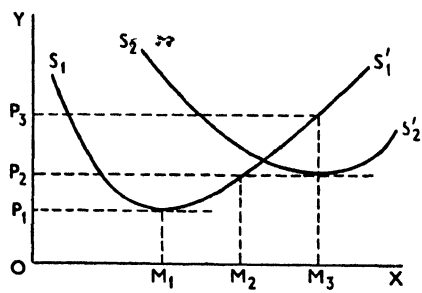


FIG. 21

that price rises to OP_2. The short-period supply curve indicates that the industry will increase its output to OM_2 at that price—that is, with the given supply of factors (labour, capital equipment etc.) it can produce OM_2 before its costs exceed the new price. In general, such a supply curve will rise fairly sharply at outputs greater than the initial level; the reason for this is that since (by definition) the factor supply is fixed in the short period, output can be expanded only through overtime on the part of labour, excessive wear of capital and the use of obsolete machinery previously unused. All this will naturally raise costs of production. In the *long* period, however, the industry will be able to adjust its capacity to

the new level of demand, and an entirely new supply curve, showing output at various prices with the new factor-combination, will have to be drawn. In Fig. 21 this is shown as S_2S_2'; from this it can be seen that long-run output at the constant price OP_2 will be OM_3, which is higher than in the short-run. Conversely, if an expansion of output to OM_3, rather than a given change of price, had been involved, price would have risen to OP_3 in the short period but would have fallen back to OP_2 in the long period. The interrelationship between price and supply will be significantly different according as to whether short-run or long-run effects are considered.

A related cause for distinguishing between the long run and short run is connected with the effects of a fall in price, and here the distinction between average and marginal cost becomes vital. In Fig. 22 is shown the average cost curve (AC) and marginal cost

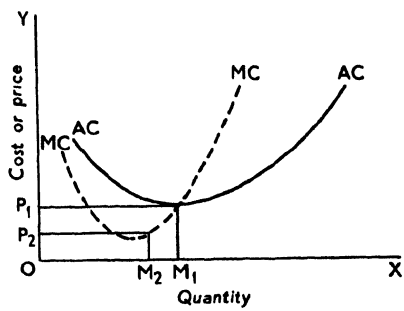


FIG. 22

curve (MC) of a producer who is suddenly faced with a fall in the market price for his product from OP_1 to OP_2. The latter price is below his minimum average cost of production, although at certain levels of output (say OM_2) it exceeds his marginal cost. What effect will this have on his output? In the long run, clearly, he will abandon production, since it is impossible for him to sell at the ruling price of OP_2 and cover all his costs. It by no means follows, however, that he will cease production immediately. In the short run his overhead costs are fixed, and he will have to meet them whether he produces or not. If he can sell his product at a price which covers his marginal (i.e. variable) costs, therefore, he is no worse off than he would be if he stopped production altogether; if price exceeds

marginal cost (even if it is still insufficient to meet average cost) he will make *some* contribution towards his overhead costs, and will be better off than he would be if he shut down completely. This explains why producers often continue to sell at less than their average cost of production, especially in industries of a highly capitalistic nature where overhead costs form a large proportion of total cost. In the long run, of course, when overhead costs *can* be reduced, an entrepreneur will reduce his output until price covers his average cost or will go out of business altogether.

The second reason why time is of especial importance in the analysis of supply is that the concept of *elasticity of supply* is inseparable from the interval of time being considered. In the case of demand, time can reasonably be ignored in analysing elasticity; the response of demand to price can be assumed to be immediate without introducing too much unreality. In the case of supply, however, the time element must be taken into account, because short-run and long-run elasticities of supply may differ considerably. Elasticity of supply can be defined as the responsiveness of output to a change in market price, but it cannot be measured without specifying the period of time over which this responsiveness is examined. In the short run, as has been seen, cost curves tend to become steep fairly quickly, so that quite a large rise in price will call forth a relatively small increase in output. The larger the time allowed for the adjustment of capacity, however, the more elastic supply will tend to become.

As between industries, many factors will enter into the determination of *relative* elasticity of supply. An industry using highly specialised capital or labour, for example, will find it more difficult (and probably more expensive) to increase its capacity than one which uses relatively unspecialised and unskilled factors. Again, an enterprise making more than one product would normally possess a relatively elastic supply potential, since it could switch its resources from one product to another fairly quickly if a change occurred in relative prices. Physical conditions of production may also influence elasticity—the supply of agricultural products is notoriously inelastic, for example, since a farmer, having once sown or prepared his fields, can do little about altering his output until the following season, however much prices fluctuate.

5. JOINT SUPPLY

Reference has just been made to the possibility of a single firm making more than one product. Where the products concerned are not the outcome of a single productive process no special problems arise. Although production may be located in a single factory, the firm can be conceived of as two or more separate enterprises operating in two or more industries, and independent cost curves can be drawn up for each product.

The situation is less straightforward, however, if the process of producing one commodity is technically connected with that of producing another, as frequently happens. In such a situation, the commodities involved would be spoken of as *joint products*, meaning that a single productive process results in the production of two or more products. The production of coal gas, for example, inevitably involves the production of coke; rearing beef cattle necessarily results in the production of hides; supplying planks for a timber-yard also increases the supply of sawdust.

The major problem raised in these situations, where an increase or decrease in the production of one commodity automatically involves an increase or decrease in the output of another, is that of attributing costs of production to either. If gas and coke are produced by the same operation, how can the total cost of production be allocated between the two products? More important, how can the *marginal* cost of either product be calculated?—further analysis will show that it is marginal cost which plays the vital role so far as an individual producer is concerned in determining the equilibrium level of price and output in a free market.

Two types of situation can be distinguished. In one it is technically impossible to vary the relative amounts of the products emerging from the productive process—the cultivation of raw cotton, for example, involves the production of a more or less constant proportion of cotton-seed. Here, the marginal cost of one product can only be taken as the total cost of producing an extra amount of *both* products *minus* the extra revenue brought in from the increased output of the joint product. Where the proportions *can* be varied, however (the amount of straw produced in conjunction with wheat can be altered by changing the type of seed employed, for example),

the marginal cost of one product can be defined as the cost of increasing its output by one unit in such a way as to maintain an unchanged output of the joint product.

The same distinction must be borne in mind in considering the effects of an increase in the supply of one product on the price of the product associated with it. If the proportions of the joint products cannot be varied, an increased output of one, called forth by a rise in demand, will necessarily result in an increase in the output of the other and therefore (assuming the demand for the latter is unchanged) a fall in its price. On the other hand, if the proportions *can* be varied, an increase in the output of one product can occur without any increase in the output of the other (or, indeed, may be associated with a *fall* in the output of the other) so that the price of the joint product may be unchanged or may even rise.

Once again the time element is frequently important. It often happens that the proportions in which joint products are supplied can be altered in the long run but not in the short run. The relative amounts of wool and mutton produced by a sheep-farmer, for example, can be varied by switching from one breed of sheep to another, but obviously this is a process which will take some considerable time. In the short run, therefore, an increased output of one joint product will usually involve an increase in the output of the other, and hence a fall in its price. In the long run, on the other hand, the proportions in which the products are supplied can usually be varied, so that an increased demand for one can be met without depressing the market for the other.

5. OPPORTUNITY COST

Throughout this discussion of supply and cost of production, the expression 'cost' has been taken to mean the money payments which an entrepreneur is forced to make in order to obtain the services of the various factors of production. In reality factors are obtained through money payments, and in economic analysis 'cost' is therefore generally used in this sense.

There is another, and more fundamental, sense in which the word 'cost' can be used, however. Money is merely a device which is used in the everyday world because of its great convenience; underlying monetary transactions, however, are the real sacrifices

inherent in the scarcity of resources in relation to our needs. Hence the concept of 'opportunity' cost is occasionally used to get behind the monetary veil to the fundamental realities which it tends to conceal. By the *opportunity cost* of production is meant the alternative uses of factors of production which are necessarily sacrificed by using those resources in a particular way. The money cost of a table is the total paid out in the course of its production by way of wages, rent, interest and profit; its opportunity cost is the desk, or chairs, or cricket bats which *could* have been made with the same resources but which are sacrificed by using them to make a table.

This concept of the opportunity cost of production is constantly recurring in various contexts. For example, wages in a particular industry are determined partly by what labour could obtain in other industries, and this will be a measure of the opportunity cost of labour. Again, a country may find that to increase a defence programme it is necessary for it to go without motor-cars which would have been supplied to it by an industry now occupied in making tanks and aircraft. For the community as a whole, as opposed to the individual producer or consumer, this concept of opportunity cost is more meaningful than that of money cost. The reason for this is that it brings clearly and unambiguously before our minds the fundamental reality of the economic world—that resources are scarce in relation to the goods and services we should like to consume; that, given full employment, using resources in one way implies a sacrifice of their alternative uses; that we cannot eat our cake *and* have it.

FURTHER READING

MARSHALL, *Principles of Economics*, Bk. V, Chs. VI–VII.
HENDERSON, *Supply and Demand*, Chs. IV–V.
* D. H. MACGREGOR, *Economic Thought and Policy* (Oxford University Press, London, 1949), Chs. 1–2.

PART IV

THE THEORY OF PRICE

CHAPTER II

The Nature of Market Equilibrium

I. PRICES, COMMODITIES AND MARKETS

Having discussed at some length the nature of the forces influencing, on the one hand, the demand for goods and services, and, on the other, the supply of goods and services, it is now necessary to turn to an examination of the way in which demand and supply interact, and thus to find an explanation of the determination of the price of any commodity in the market.

Before going further, the meaning of three important words in that last sentence must be made quite clear: price, commodity and market. The word 'price' will not give much trouble. In a modern economy goods are exchanged not directly but through the medium of money, and by price is meant the sum of money which a person hands over to a seller in order to obtain ownership of the commodity concerned. Nevertheless price is a little more complicated than this; in treating it an economist may be concerned with either *absolute* or *relative* prices—that is, he can be seeking to answer either the question 'Why are pencils 4d.?' or the question 'Why is the price of pencils only one-tenth the price of a packet of cigarettes?' It is with the second kind of problem that the analysis is concerned for the moment—i.e. the problem of the causes underlying the value of one commodity relatively to that of other commodities. The problem of the absolute level of prices will be discussed later on.

An unambiguous definition of the word 'commodity' is rather more difficult. In a general, and rather loose, way a commodity can be spoken of as being a good, or service, for which there is no close substitute and which satisfies a particular, distinct want felt by consumers. Unfortunately, the identification of a particular commodity is very difficult in practice, since neither wants nor commodities are sharply distinct in the real world. Is the demand

for tea a demand for tea-drinking as such, or is it simply one part of the demand for beverages, and hence met by coffee, cocoa and a dozen other fluids as well as tea? And is *tea* the commodity, or is Indian tea a different commodity from China tea?; is Lipton's tea a different commodity from Lyons' tea? How close does a substitute have to become before it ceases to be a separate commodity? This kind of difficulty arises with almost every product in the economic system, and presents virtually insoluble problems. In speaking of 'a commodity' in the sense mentioned above, then, a fairly drastic over-simplification is being made in order to render the analysis at all manageable, and the real world is considerably more complex and less clear-cut than the analytical picture of it.

Finally, the word 'market' has to be defined. In everyday speech the word is frequently used to mean a building or area of ground in which trading takes place on certain occasions. In economics the word is used much more widely, and refers to all the persons and institutions concerned in the exchange of any commodity. In this sense it can include an entire country, or even the whole world, if people buying or selling a commodity are in contact with one another over those areas. On the other hand, a market may mean only a handful of people, if the persons in contact with one another for the purpose of exchanging some commodity are few in number. In any case, a market in this sense will have no single physical location; the person encouraged by an effective advertisement to buy a commodity in his local shop will be unknown to its manufacturer, who is perhaps thousands of miles away, but they are nevertheless both part of the market for that commodity.

2. INDUSTRY AND FIRM: THE DEMAND CURVE

Assume that an individual commodity can be defined closely enough for the present analysis. By a firm is meant any group of factors of production operating under the direction of an entrepreneur: by the industry is meant the collection of firms engaged in the production of the commodity with which the analysis is concerned.

Now this distinction between firm and industry is an important one, and it will underly all the analysis of the various possible types of market situation with which this book will have to deal. In parti-

cular, the distinction has important implications for two things: the *demand curve* involved in the analysis and the *equilibrium conditions* which have to be satisfied. Consider the implications for the demand curve first of all.

For the industry as a whole, the relevant demand curve is that for the commodity (since it is *total* demand and supply which is involved when the industry is considered) and it will therefore be of the type discussed in Chapter 3. Now, in the earlier discussion of those curves the analysis was concerned with the relationship between the quantity purchased and the commodity's price, and such a curve can therefore be called an *average revenue* curve. Every unit of the commodity has to be sold on the market at the same price: if the producer's total receipts were divided by his total sales, his average revenue per unit of output, which will obviously be the same thing as its price, would be arrived at.

But just as economists distinguish between average and marginal cost, so a distinction can be drawn between average and marginal revenue. If average revenue is equal to the total receipts from the sale of a commodity divided by the total number of units of the commodity sold, then *marginal revenue* is equal to the *change* in total receipts brought about by a change of one unit in the total number of units of the commodity which are sold.

At first sight it may seem that, as all units of a commodity are sold at the same price, marginal revenue (as so defined) must be the same thing as the price at which the marginal unit is sold, and therefore the same thing as average revenue. But this is not necessarily so. The essential point about the falling demand curve (i.e. one sloping downwards from left to right) is that in order to sell more of a commodity the price (*of all units*) has to be reduced; the marginal revenue from one unit of output is therefore the price at which it is sold *minus* the fall in revenue from all the other units of output which the sellers have to offer at a lower price than before in order to persuade the market to absorb this extra, marginal unit. Given a stable demand curve of the usual, downward-sloping type, marginal revenue must always be *less* than price; it may even be negative.

Consider the example shown numerically in Table X and diagrammatically in Fig 23. In order to increase sales, the price of the commodity has to be progressively reduced. As a result, the total

revenue rises as sales increase up the sixth unit, although it increases less and less rapidly as the market price is reduced. After the sixth unit total revenue actually declines, since the expansion of sales is insufficient to offset the decline in price necessary to secure it (i.e. elasticity of demand falls below unity); as a result, marginal revenue becomes negative.

Wherever the demand curve is falling then, as it usually is for the output of an industry as a whole, a distinction must be drawn between *average* revenue, or price (the *average* receipts per unit at

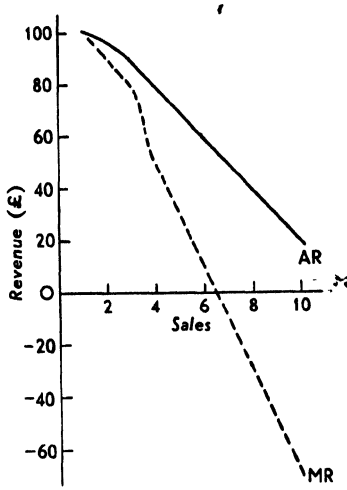


FIG. 23

TABLE X: *Price and Marginal Revenue*

Number sold	*Price (= Average Revenue)*	*Total Revenue [(1) × (2)]*	*Marginal Revenue*
(1)	(2)	(3)	(4)
	£	£	£
1 -	- 100	100	100
2 -	- 95	190	90
3 -	- 90	270	80
4 -	- 80	320	50
5 -	- 70	350	30
6 -	- 60	360	10
7 -	- 50	350	- 10
8 -	- 40	320	- 30
9 -	- 30	270	- 50
10 -	- 20	200	- 70

each level of output) and *marginal* revenue (the addition to total receipts resulting from the last unit produced).

3. INDUSTRY AND FIRM: EQUILIBRIUM CONDITIONS

The second context in which the distinction between firm and industry is important is that of the conditions necessary for equilibrium. It will be remembered that in economics by equilibrium is meant a situation in which there exists no tendency towards change. For an industry, then, equilibrium means a situation in which there is no tendency for new firms to enter it or for existing firms to leave it; for a firm, it means that the entrepreneur has no incentive either to increase or to reduce his output. Under what conditions will such equilibrium situations be achieved?

Consider first the industry, which on this occasion presents the easier problem. In an earlier chapter the concept of 'normal' profit was introduced, by which was meant the average rate of reward per unit of output prevailing for the average entrepreneur over the economy as a whole. Now, entrepreneurs will be attracted *into* an industry if the average rate of profit being earned in it is above this normal level, and they will tend to move *out* of an industry if they are earning in it less than this normal level. Equilibrium, for the industry, means that firms are neither moving in nor moving out; it means, therefore, that the level of profits in it is neither above nor below the normal level, and hence is equal to it.

When will this normal level of profit be attained? The average amount received by entrepreneurs per unit of output is given by the average revenue or price curve—the demand curve discussed at some length in Chapter 3. The amount paid out by entrepreneurs per unit of output *including the normal profit due to them in respect of their own contribution* is given by the average cost curve discussed in Chapter 10. It follows, therefore, that entrepreneurs are receiving normal profits, no more and no less, when average cost equals average revenue, or price. If average cost is *above* price, then they are earning something less than normal profit per unit; if average cost is *below* price, then they are earning excess (i.e. above normal) profit. For the industry to be in equilibrium, average cost must equal average revenue.

Turn now to the firm. The basic assumption in all economic analysis is that each person is seeking to maximise his economic well-being and acts rationally towards that end. It follows that the entrepreneur must be assumed to seek the maximum profit from his firm. So long as a firm is in existence at all, the entrepreneur will carry on that level of output which brings in to him the greatest profit. When is this situation reached? Here the two concepts of marginal revenue and marginal cost must be brought together. Marginal revenue is the addition to total receipts arising from the last unit of output; marginal cost is the addition to total cost which it incurs for the entrepreneur. Now, if an entrepreneur by increasing his output can add more to his total receipts then he adds to his total cost, obviously he is increasing his total profit. If he adds more to cost than to receipts, he is reducing his total profit. He will

maximise his profits, then, when marginal cost *equals* marginal revenue.

The example shown in Fig. 24 illustrates this graphically. Let the curve *MR* represent the marginal revenue brought in by each unit

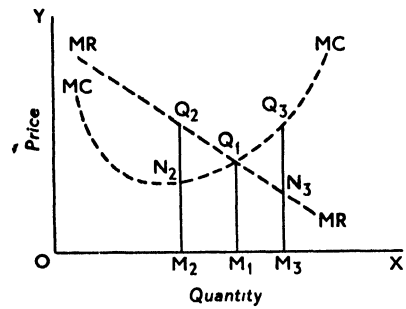


FIG. 24

of output and *MC* the addition to total cost made by each unit. The entrepreneur's profits will be maximised at output OM_1, where the curves intersect and marginal revenue equals marginal cost. At any smaller output, say OM_2, it will be possible to increase total profit, since the extra cost involved in expanding output (M_2N_2) is smaller than the extra receipts which the extra output would earn (M_2Q_2). At any larger output, say OM_3, the extra receipts (M_3N_3) are smaller than the extra costs (M_3Q_3). At output OM_1, where the curves intersect, the entrepreneur adds just enough to his revenue to cover the marginal cost of production and his own normal profit.

For full equilibrium in an industry, then, it is necessary to have (*a*) the number of firms in the industry constant and (*b*) the output of each firm constant. For (*a*) the necessary condition is that the profits made by entrepreneurs should be at the normal level, which means that average cost must equal average revenue, or price, at the ruling level of output. For (*b*) the necessary condition is that each entrepreneur is making the maximum profit possible, which means that marginal cost must equal marginal revenue at the ruling level of output. In full equilibrium, therefore, each firm will be producing that output at which average cost equals average revenue ***and*** marginal cost equals marginal revenue.

One last word of warning should perhaps be added. The fact that a producer is maximising his profits does not mean that his output is at an optimum from an economic point of view. The optimum output is that at which the factors are combined in the most efficient way—i.e. the output at which average cost of production is at its lowest point. This output *may* be that which yields the producer maximum profit (i.e. marginal cost may equal marginal revenue at the point of minimum average cost); as the succeeding chapter shows, however, this occurs only in the special and limiting case of perfect competition.

FURTHER READING

* JOAN ROBINSON, *The Economics of Imperfect Competition* (Macmillan, London, 1948), Introduction, Chs. 1 and 2.

CHAPTER 12

Price under Perfect Competition

1. WHAT IS PERFECT COMPETITION?

As a starting-point for the explanation of the process whereby the exchange value of a commodity is determined, it is convenient to postulate a hypothetical situation called perfect competition. No-one suggests that such a situation has ever existed for other than a few exceptional commodities; the hypothesis is used merely as a simple framework in which the essential factors are analysed before moving closer towards the infinitely more complex situation which exists in the real world. Many subjects learnt at school (algebra, for example) will have only limited direct usefulness in later life, but that by no means implies that the study of them is useless. The important thing is that they help people to think in a careful and systematic way when faced with a particular type of problem, and in this way may contribute a great deal to their work in later years. So it is with the concept of perfect competition. It is most unlikely that perfect competition will ever be seen in the real world, but, nevertheless, through its study it is possible to acquire a system of approach which in due course will help a great deal in tackling the more complex situations which are in fact encountered in the real world.

This state of perfect competition, then, is a purely hypothetical one. To the economist, it implies the absence of monopoly power—i.e. the absence of any power on the part of any individual firm or consumer to influence market prices. In it certain things are assumed about each of the two groups of persons involved in economic exchange—producers and consumers. So far as the supply side is concerned, it is assumed, first, that there is a large number of sellers competing with one another—sufficiently large a number, that is, for each seller to be contributing an infinitely small part of

the total supply, and hence to be unable to affect the market price significantly by changes in his output. Secondly, it is assumed that each producer (being a very small element in total production) has a perfectly elastic supply of the factors of production; he can increase his output, in other words, without driving up the level of wages, rent or interest currently prevailing in the industry. Thirdly, it is assumed that there are no restrictions on the free entry of firms into the industry or their movement out of it. Finally, each producer is assumed to be aware of the rate of profit being made by every other producer.

So far as the demand side is concerned, it is assumed again that the number of consumers is so large that no single buyer is able to perceptibly influence the market price by altering the scale of his purchases. Secondly, all buyers are assumed to have perfect knowledge of the price being asked for the commodity concerned in every part of the market, and it is supposed that neither transport costs nor inertia prevent them from taking advantage of any divergence in price which might develop in different parts of the market. Finally, buyers are assumed to have no preference as between the various producers of the commodity—that is, that there are no differences, actual or imaginary, between the product of one firm in the industry and that of any other firm in the industry.

This is a formidable collection of assumptions—so much so that it is worth repeating that their purpose is solely to put on one side, for the moment, the large number of subsidiary influences which usually operate, and to expose the basic forces common to all markets. To attempt an analysis of the highly complex markets of the real world, where these simplifying assumptions do not hold, without first clarifying the essential features of a simplified situation would be attempting the impossible; it would be setting out on Scarlatti without first mastering the scale.

2. THE DEMAND CURVE UNDER PERFECT COMPETITION

Consider the demand curve facing the individual firm in this hypothetical situation of perfect competition; it will be convenient to refer to this as the *sales curve* so as to prevent confusion with the demand curve for the commodity facing the industry as a whole. Since the firm's product will be identical (by definition) with that

of the rest of the industry, it might seem that its sales curve will be the same as that facing the industry as a whole—i.e. downward-sloping, as in Fig. 23. This is not the case, however, as will become clear by recalling the assumptions of the situation of perfect competition. By hypothesis, each firm is producing so small a part of the industry's total sales that no change in its output can perceptibly influence price. Whatever its output, therefore, the individual firm sells at the (given) market price, so that its sales curve is a horizontal line drawn through the market price, *OP*, as in Fig. 25, and extending over the whole range of its possible output.

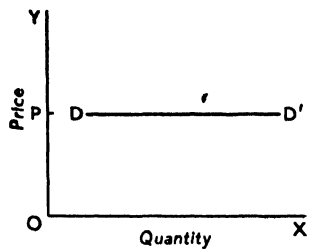


FIG. 25

Can a separate marginal revenue curve be drawn in this case? Obviously it cannot; if each unit is sold at a constant price, each one must add exactly the same amount to total receipts as every other unit. Here there is only price to consider; there is no adjustment necessary because of the price-reduction enforced by increased sales, because no price-reduction is involved. For the firm in perfect competition, then, average revenue, or price, and marginal revenue coincide.

It may seem rather odd that each firm in an industry should face a horizontal (i.e. perfectly elastic) sales curve while the firms collectively face a falling demand curve. How can the addition of a large number of horizontal curves result in the industry's falling curve? The answer is that the individual sales curves facing each firm cannot legitimately be added up. Each of them is drawn up on the assumption that the rest of the industry maintains a constant output, so that any expansion by an individual firm will not result in any significant movement along the output axis of the industry's demand curve. If *all* firms expand, of course, there is naturally a perceptible movement along the industry's output axis and the market price will fall. But for any firm in isolation, market price can be taken as determined by forces entirely outside its control.

3. MARKET EQUILIBRIUM UNDER PERFECT COMPETITION

The various strands can now be pulled together in order to see

the full pattern of perfectly competitive equilibrium. It has been established that each firm faces a perfectly elastic (i.e. horizontal) sales curve, such as that shown as *DD'* in Fig. 25, and that with such a sales curve average revenue and marginal revenue coincide. The condition for equilibrium in the industry is that average cost must equal average revenue, while for equilibrium of the firm marginal cost must equal marginal revenue. But average revenue and marginal revenue coincide under perfect competition. The equilibrium conditions therefore require that, at the ruling output, average cost and marginal cost should *both* equal average revenue, or price, and

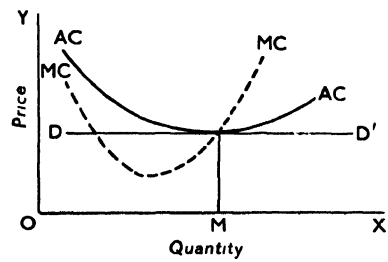


FIG. 26

must therefore be equal to one another. It follows that the equilibrium situation must be of the kind shown in Fig. 26—output settling at that point where the sales curve is tangential to the average cost curve, and where the average cost curve is at its lowest point.

Why should this rather special combination of features necessarily be involved? Consider it in two parts: first, *why* should output settle at the point of minimum average cost?; second, *why* should the sales curve be a tangent to the average cost curve at that particular level of output? The answer to the first question is that, as has been established, equilibrium requires that both average and marginal cost should equal price and therefore one another; but with a normal U-shaped cost curve, marginal and average cost can be equal only at the lowest point of the average cost curve. So long as marginal cost is *below* average cost each extra unit of output is costing less than the average for output as a whole, and overall average cost must be falling. So long as marginal cost is *above*

average cost, the average (on the same kind of reasoning) must be rising. When marginal cost equals average cost, the average must be neither rising nor falling—i.e. it must have reached its lowest point and not yet started to rise again.

Turn to the second question: why should the sales curve just touch the average cost curve at this minimum point, thereby forming a tangent to it? The reason is that the equilibrium conditions require price (average revenue, equal in this case to marginal revenue) to be equal to both marginal and average cost. It is only at this point, where average and marginal cost are themselves equal, that the single price can be equal to both average and marginal cost. Hence it is only if the sales curve touches the average cost curve at this particular point that full equilibrium, of both firm and industry, can be attained.

This can be seen if other cases are considered in which this special combination of requirements is not achieved. In Fig. 27 the horizontal sales curve is above the point of minimum cost. If the firm produces output OM_1 it will itself be in equilibrium, because

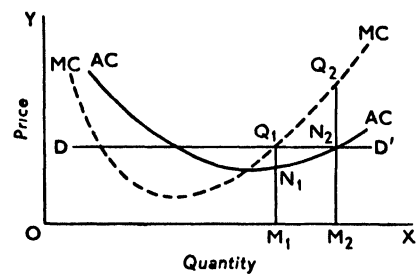


FIG. 27

marginal cost equals marginal revenue at that point and its profits are maximised. But at that output its average revenue (M_1Q_1) will be in excess of its average cost (M_1N_1) and it will therefore be earning more than normal profits; new firms will be attracted into the industry by these excess profits, so that the industry will not be in equilibrium. (As new firms come in, total supply will expand and the market price will fall, so that the individual firm's sales curve will fall until it touches the average cost curve at its lowest point and profits are back to normal.) Conversely, if it produced OM_2

average cost and average revenue would be equal, profits would be normal and the industry would be in equilibrium. But the firm would not be maximising its profits; at output OM_2 the marginal cost of production (M_2Q_2) would be in excess of the marginal revenue brought in (M_2N_2) and the firm would be reducing its total profit with every unit it produced after OM_1.

Whatever other situation is postulated, then, either the industry or the firm will be in a state of disequilibrium, and price will continue to change. Only when price is equal to both marginal and (minimum) average cost will both the individual firm *and* the industry as a whole be in a state of complete balance. In all this, of course, it is assumed that the forces of adjustment—changes in output or the movement of firms—are given time to work themselves out, although the analysis could easily be conducted in terms of short-run and long-run cost curves if it was desired to allow for the differences introduced by the existence of time-lags.

So a fundamental conclusion of the theory of value is arrived at: *given conditions of perfect competition, the price of a commodity tends to equality with its marginal cost of production.* As has been seen, price will also be equal to the minimum average cost of production. If for any reason the equality of price and marginal cost is disturbed, forces will automatically be brought into play to restore it. Neither demand nor supply can be said to have a dominant role in the determination of price, therefore; price is the outcome of the interaction of demand and supply, each making a contribution of equal importance.

FURTHER READING

MARSHALL, *Principles of Economics*, Book V, Chs. I–III.
HENDERSON, *Supply and Demand*, Ch. IV.
E. H. CHAMBERLIN, *The Theory of Monopolistic Competition* (5th ed., Harvard University Press, Cambridge, 1946), Ch. II.
* ROBINSON, *Economics of Imperfect Competition*, Ch. 7.

CHAPTER 13

Price under Monopoly

1. THE NATURE OF MONOPOLY

In the previous chapter a hypothetical situation was discussed in which the number of competing producers of a commodity was so large that no single producer was able to exert a perceptible influence on the market price through variations in his output. Logically the opposite extreme of *monopoly* can be conceived of—a situation in which a single producer controls the entire output of a certain commodity. Like the concept of perfect competition, monopoly is easier to define theoretically than to identify in practice. A great deal turns, as before, on the definition of a commodity, or of the extent of the break in the chain of substitutes which is required for a product to become a separate commodity rather than one of a number of types, or brands, of a commodity. In a sense, every producer has a monopoly of his own product provided there is even the smallest difference, actual or imaginary, between the goods he sells and those sold by his rivals; on the other hand, the existence of monopoly could be denied altogether on the grounds that a substitute of some kind exists for every product under the sun, as indeed it does.

In economic analysis the use of the word monopoly lies between these two extremes. Each brand of chocolate, or motor-cars, or cigarettes, is not treated as a separate commodity, and therefore their producers are not treated as monopolists. On the other hand, a product such as coal is treated as a separate commodity even though gas and electricity and oil are to some degree substitutes; hence the National Coal Board, having the sole right to produce and supply coal in Britain, is treated as a monopoly.

Just as a flexible, common-sense attitude is adopted in deciding when a product is or is not a separate commodity, so a reasonable

attitude is adopted towards the extent of control over total output required before a producer is regarded as a monopolist. To say that a firm does not have a monopoly because it produces only 95 per cent, and not 100 per cent, of the total supply of a commodity would clearly be unrealistic. Rather, a producer would be said to have a monopoly of a commodity if he controls such a large share of total output as to be able to exercise decisive control over the supply and price of the commodity through variations in his output.

The essential point about monopoly, then, is that *there ceases to be a distinction between the firm and the industry*. A firm which has a monopoly position in any market *is* the industry so far as equilibrium analysis is concerned. By definition, no other producers can exert any perceptible influence on total supply or on price, whatever the level of profits enjoyed by the monopolist. So far as equilibrium is concerned, it is only the firm which requires analysis; so long as the firm is in equilibrium, then of necessity the industry is in equilibrium.

2. SOURCES OF MONOPOLY POWER

A firm—or group of firms operating under some common direction—may acquire and retain a dominant position in the market for a commodity by various means. A common source of monopoly power in history has been the control of one or more factors of production. In particular, the ownership of deposits of *raw materials* may give an unchallengeable position to a firm in producing a commodity for which those raw materials are essential. In the long run, a position built up on this advantage is likely to be undermined by either the exhaustion of the deposits involved or the development of alternative natural or synthetic materials.

A second important type of factor promoting a monopoly position is that of *legal restriction* on the entry of potential competitors into an industry. In Britain, for example, many fields of activity are closed to private firms, and the sole right of operation is conferred by statute on one enterprise—the Central Electricity Board, National Coal Board and the General Post Office are familiar examples. More commonly, the law of *patents* can be invoked by a firm to secure the sole right of manufacture of a new product or use of a new process. Finally, *tariffs*, or taxes on imports, may be used

to prevent foreign producers from competing with a domestic firm, or the same result may be attained by import restrictions of a quantitative kind.

Thirdly, competition may be prevented by *financial* factors. Highly capitalistic industries, such as the manufacture of iron and steel, require enormous outlays on capital equipment before production can be carried out on the level necessary to compete with an established firm; as a result, potential competitors may find it virtually impossible to enter the industry. In other industries, a firm may have acquired such goodwill in a market that new entrants could break into the market only through an expensive advertising campaign.

Finally, a monopoly position may be built up through various devices specifically designed to prevent competition. An existing producer may deliberately sell a commodity at a *loss* in order to drive competitors with smaller reserves into bankruptcy. A trade union may pursue a *closed shop* policy—i.e. threaten strike action unless employment in a factory is confined to its members—with exactly the same end in view. Another common practice is for producers to form themselves into a *trade association* of one form or another, and enforce restrictions on full competition through collective measures. For example, a mutinous entrepreneur may be brought to heel by a closing up of his channels of supply or of sales, since the association as a whole can threaten suppliers or retailers with a complete stoppage of orders or deliveries unless their measures are enforced.

3. MONOPOLY EQUILIBRIUM

Having seen how monopoly power may originate, it is necessary to consider what the exercise of that power may mean in terms of the equilibrium price and output in the market for a monopolist's product. As was mentioned above, the analysis is relatively simple, since by definition the condition for equilibrium in the market as a whole becomes identical with the equilibrium condition for the (single) firm. For the firm, equilibrium is attained when profits are maximised, and the possession of monopoly power makes no difference whatever to this rule. The peculiarity introduced into the analysis by monopoly power is not that the requirements for equili-

brium are affected; it is that there are no additional requirements which must be met for full equilibrium of the *industry* to be attained, as there are under perfect competition. The rule is simply that the monopolist will maximise his profit, and the previous reasoning, by which it was established that profits are maximised when marginal cost and marginal revenue are equal, remains perfectly valid.

Monopoly equilibrium, therefore, implies a situation such as that shown in Fig. 28. As the monopolist supplies the whole market, the demand (or average revenue) curve facing him will be that for the commodity as a whole, and hence will usually be downward-sloping; the distinction between sales curve and demand curve disappears. The monopolist will maximise his profit at that output which brings marginal revenue into equality with marginal cost (*OM* in the diagram). Over and above the 'normal' profit which is an element in the average cost curve, he will receive on each unit of output excess, or monopoly, profits equal to the difference between average cost and price (the area *RSPQ* in the diagram therefore represents the monopolist's total excess profit).

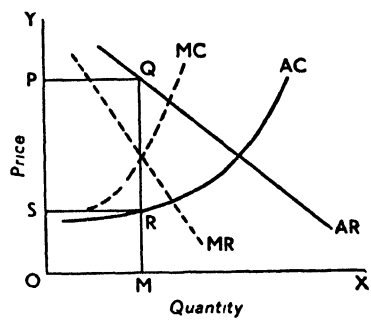


FIG. 28

So far the argument has been conducted with reference to a single unified market throughout which the monopolist charges a single price. Precisely similar considerations apply to what is called a *discriminating* monopoly—that is, a situation in which a monopolist is able to break his market into two or more separate and isolated sections in which separate prices can be charged. A Harley Street specialist may charge two different fees for rendering an identical service to two different men, if he thinks one can afford to pay more that the other; British coal is currently sold at one price in the home market and a different price in the export market; the British Electricity Boards supply a single product—electricity—at different prices to industrial users on the one hand and domestic consumers on the other. In such cases a monopolist seeking to maximise pro-

fits will sell in each section of the market that output which brings the marginal revenue from each section into equality with the marginal cost of his total output. If the elasticity of demand for his product differs considerably between the different sections of his market, then the prices he will charge will also differ considerably, since the greater the elasticity of a demand curve the closer does marginal revenue come to price at any output. As was noted earlier, an elasticity of demand greater than unity means that total expenditure on the commodity increases as price falls, so that marginal revenue (the *change* in total expenditure) will be positive and may approach the level of the price itself. Elasticity *below* unity means that total expenditure falls as price decreases; hence marginal revenue, being negative, is much below price. The monopolist will be producing only that output which makes marginal cost equal to marginal revenue, so that the more rapidly marginal revenue falls (i.e. the less elastic the demand) the smaller the monopolist's output, and therefore the higher the market price.

This last consideration helps to bring out one of the two main features of the price and output policy of a monopolist. This is that a monopolist will never expand output beyond a point at which elasticity of demand falls close to unity. An elasticity of unity implies that total outlay is unaffected by a reduction in price, and this means that marginal revenue (the change in total outlay) is zero. Since the producer must incur *some* marginal cost by expanding output, his profits will be reduced if he has no additional revenue to offset marginal cost. Hence a monopolistic agricultural organisation, for example, might prefer to destroy crops rather than sell them in a market. If the demand for the crop is inelastic, the sale in the market might result in a *negative* marginal revenue and thus profits would be bound to fall, even if the crops destroyed had no marginal cost whatever.

The second main feature of monopoly equilibrium is that, in comparison with *perfect* competition, price will usually be higher and output will usually be lower. Under perfect competition, as the last chapter showed, output will settle at that point where price is equal to marginal cost. The monopolist produces that output at which *marginal revenue* is equal to marginal cost. As the demand curve for any normal commodity is downward-sloping, margina

revenue will always be to the left of average revenue, or price (i.e. at any output marginal revenue will always be lower than price). It follows, then, that the marginal revenue curve will cut the marginal cost curve to the left of the point at which the price curve cuts it—i.e. at a point where output is lower. The monopolist makes his excess profits, in other words, by holding down output and holding up price, and he is not prevented from doing this (as he would be under perfect competition) by a movement of firms into the industry under the attraction of the abnormally high level of profits prevailing there.

Two qualifications must immediately be made to this conclusion, however. First, it rests on the assumption that the level of costs is the same under monopoly and under competition. This may not be so. A monopolist, by concentrating production under single control, may gain considerable economies of large-scale production. As a result, the monopolist's marginal cost curve may be considerably lower than that prevailing under competition, so that price may be lower and output higher under monopoly, despite the presence of excess monopoly profits. This is not so improbable as might be thought at first sight. In the supply of gas or electricity, for example, the market is usually a limited geographical area; frequently one modern generating plant can easily supply the entire market. It would obviously be wasteful, and therefore costly, if the market were served by rival generators, with rival pipes and cables in the same streets, all operating at much below their full capacity. The same is true of railway systems; the idea of duplicating expensive permanent ways and operating staff has only to be considered to be seen to be absurdly wasteful. Monopoly and excessively high prices, therefore, are by no means synonymous.

The second qualification is that when monopoly is compared unfavourably with competition, it is *perfect* competition with which the comparison is being made, and perfect competition seldom, if ever, exists in reality. The more appropriate comparison in reality is between monopoly and the imperfect form of competition which is usually the practical alternative. As will be found in the next chapter, imperfect competition frequently involves a good deal of waste and inefficiency (especially in matters such as product differentiation and competitive advertisement) which would not be

experienced with a monopoly. This strengthens the possibility, then, that monopoly costs (including selling as well as production costs) may be lower than those prevailing under the kind of competition actually experienced.

Nevertheless, the fact remains that the existence of monopoly power (whether in the form of pure monopoly or as an element in the half-competitive, half-monopolistic world examined in the next chapter) has a restrictive effect on output, tending to result in higher prices to the consumer and higher profits for the producer than would otherwise prevail. This is particularly the case when demand is relatively inelastic, so that marginal revenue falls rapidly; the steeper the demand curve, the further to the left will be the point at which the marginal cost and revenue curves intersect; this means that a monopolist's output would be smaller, and his price higher, than would be the case under competition despite a substantial reduction of cost which might be secured through the monopoly organisation. Hence the problem of monopoly policy cannot be ignored in an economy where monopolistic elements exist to any significant degree.

4. MONOPOLY AND PUBLIC POLICY

There has been a growing realisation during the present century of the dangers of monopoly, and especially of the possible exploitation of the consumer through restricted output and high prices, and the undue concentration of economic power in a small number of private individuals, which may result from a monopoly position. There can be little doubt that these dangers are real, although it must be remembered that the analysis of monopolistic prices rests (as always in economic theory) on the assumption that entrepreneurs invariably act so as to maximise their profits. If this assumption does not hold—and this may frequently be the case in reality—then the conclusion that price will usually be higher under monopoly than under competition will have to be qualified.

Assuming, however, that other things being equal a monopolist does usually seek to maximise his profits, how can the exercise of monopolistic power, harmful as it is to the consumer, be restricted? In the first place, the monopolist's power is always limited by the existence of substitutes, to which consumers may turn if prices are

set too high. In other words, substitutes will increase the elasticity of the demand curve facing the monopolist, and it has already been seen that the gap between monopoly price and competitive price will always be smaller the greater is the elasticity of demand. Even if no close substitutes are produced in the same country, a monopolist must always be careful to keep his price below the point at which it is possible to import similar products from abroad and market them at a competitive price.

Secondly, a monopolist may not charge the high price which maximises his profit for fear of stimulating rival producers to enter his market in order to share in his high profit. A monopoly position resting, for example, on the high costs of entry into the industry may be only a qualified one; competitors may not be prepared to incur the high costs of setting up in the industry so long as the rate of profit in it is only a little above the normal level, but they may reconsider their judgement if the profits being earned by the monopolist become very high indeed and present a strong temptation.

Thirdly, a monopolist may be content to earn something less than the maximum rate of profit open to him in the hope of extending the market and so increasing his long-run profits. He may deliberately hold down the price of his product in the belief that, by accepting lower profits in the short run, the growth of the demand for his product will be encouraged; a relatively low rate of profit on a large output in the long run may attract him more than a high rate of profit on the small output indicated by existing demand conditions.

For all these reasons, then, the *existence* of monopoly powers cannot necessarily be taken to imply the full *exercise* of these powers. In many cases, the loss suffered by the consumer as the result of a moderate use of monopoly power may be more than offset by the gain he enjoys because of the lower costs—production and selling costs—brought about by the concentration of output in a single enterprise, or a small group of enterprises operating in association with one another.

Nevertheless, in many—and probably most—cases the consumer suffers to a greater or lesser degree from the existence of monopoly, despite all these qualifications and limitations of the monopolist's power. When the abuses of monopoly become substantial, it may

be thought that only intervention on the part of the government can ensure adequate protection for the consumer. Such intervention may take the form of a complete transfer of the control of an industry out of private hands, or at least a direct control over the industry's price and output policy. For example, the relatively high capital costs incurred in the production and supply of gas and electricity made it inevitable, from an early stage in the history of these industries in Britain, that a single producer would operate on a monopolistic basis in each area. The wastes involved in duplicating or triplicating gas works and pipe systems in an area which could be adequately served by a single producer (and which would *have* to be served by a single producer if the potential economies of large-scale output were to be realised), would be obvious to even the bitterest opponent of monopoly. Hence, the supply of gas and water was either taken over by local government authorities themselves or was conducted by private corporations having the sole right of operation in each area but restricted fairly severely in their powers to raise prices or to increase their profits. Subsequently, of course, the economies gained from a consolidation of the industries into single units operating on a nation-wide basis have been thought sufficient to justify their nationalisation. For similar reasons, in most countries postal services are provided by an agency of the government itself.

The operation of an industry by a government or some semi-governmental body, however, is not always considered a desirable development; this may be because of political considerations, or because of technical reasons which suggest that a particular industry is not conducive to efficient management by an official body. In such cases, other means of restraining or eliminating monopoly powers have to be adopted. Practice varies widely from country to country in this respect. In the United States, for example, a vigorous anti-monopoly campaign has been conducted for many years through the processes of the law. Specific monopolistic practices, and particular forms of company organisation which are considered to be means of securing monopoly power, have been declared by statute to be illegal, and an energetic branch of the government has been charged with the discovery and prosecution of breaches of the law. Although it is impossible to measure the degree of success enjoyed

by measures of this kind (after all, even the extent of monopoly itself cannot be measured, let along the influence of public policy upon it) it seems likely that it has exercised a salutary effect in discouraging producers from trying to build up monopoly positions.

In Great Britain the attitude towards monopoly has been rather different until recent years, the tendency being to avoid general condemnation of particular practices or forms of organisation. The reason for this has been partly the belief that there are few legal restrictions which a competent company lawyer cannot invent some means of surmounting, but mainly the conviction that practices which may be undesirable in some cases may be perfectly defensible in others. It had been thought better to consider each industry separately, therefore, and to legislate when necessary on an *ad hoc* basis.

A significant departure from this policy occurred in 1948 with the establishment of a Monopolies Commission charged with the consideration of practices and policies being pursued in industries referred to it for examination by the Board of Trade. In its reports after each investigation, the Commission was required to indicate practices which it considered to be contrary to the public interest, and the government were given powers to declare such practices to be illegal. A far-reaching development of this policy occurred in 1956 with the passage of the Restrictive Trade Practices Act, under which various practices of a monopolistic nature were declared to be illegal unless firms adopting them could satisfy a Restrictive Practices Court that they are not contrary to the public interest in their particular case. Each industry has to register these designated practices within a specified time with a Registrar appointed for the purpose and to either obtain approval of them from the Court or abandon them within that specified period.

At the moment of writing it is too soon to arrive at any judgement as to the efficacy of this policy. It is fairly certain, however, that its degree of success will depend heavily on the extent to which the existence of this legal machinery and examination induces the general body of entrepreneurs to abandon monopolistic practices (or abstain from their adoption), rather than on the number of successful prosecutions conducted with its aid. All statutory law depends, in the last resort, on the general acceptance of it by the over-

whelming majority; a law which the bulk of the population regards as a bad law cannot succeed or survive in a democratic society. This is true of the law relating to monopoly; it will attain its end only if it expresses and reflects a general belief amongst entrepreneurs themselves that monopoly is in general undesirable, and is something from which in the long run they, like society as a whole, can only suffer.

FURTHER READING

MARSHALL, *Principles of Economics*, Bk. V, Ch. XIV.
ROBINSON, *Monopoly* (Cambridge University Press, London, 1941), Chs. I, III, VIII, XI and XII.
BEACHAM, *Economics of Industrial Organisation*, Chs. V–VI.

CHAPTER 14

Price under Imperfect Competition

1. THE NATURE OF IMPERFECT COMPETITION

The two preceding chapters have been concerned with the analysis of the forces determining price under certain extreme and hypothetical circumstances, that of perfect competition on the one hand and of pure monopoly on the other. It is almost impossible to discover a single commodity in the actual world which is exchanged under conditions of perfect competition, and equally difficult to discover instances of pure monopoly. As frequently happens in other connections, the reality is to be found somewhere between the two extremes; the vast majority of markets display characteristics of both monopoly and competition, the monopoly element tending to dominate in some, that of competition in other. Such situations, where neither pure monopoly nor perfect competition prevails, economists call *imperfect competition.*

How can this mixture of competition and monopoly arise? Usually it is caused by factors operating on both the supply side and the demand side of the market mechanism. On the supply side, perfect competition postulates an infinitely large number of producers, each unable to influence the market price perceptibly, while pure monopoly requires a single seller free from the potential competition of rival producers. In reality there are usually several producers of the same commodity but not an infinite number of them, with the result that each producer *can* exert a significant influence on total supply through changes in his own output, and can therefore influence price perceptibly; at the same time, some degree of competition exists, so that the equilibrium of the industry, as well as of the individual firm, has to be taken into account.

On the demand side, elements of both competition and monopoly

are found because of the existence of different brands, or types, of any given commodity; these possess individual characteristics but are nevertheless close substitutes for one another. Perfect competition rests on the assumption that consumers regard the products of the various firms in an industry as identical in all respects; in the real world, through advertisement and branding, the consumers of most commodities are led to believe (rightly or wrongly) that Company A's cigarettes are a different product from Company B's cigarettes, even though the two brands may actually be physically identical. Pure monopoly, on the other hand, postulates that there is no close substitute for Company C's product, whereas the products of Companies D and E are usually fairly close substitutes. The usual position is therefore one of part monopoly and part competition; each producer has, in a sense, a monopoly of his own product (Company A, obviously, is the sole supplier of 'A Brand' cigarettes) but each is facing the competition of other brands of the same product which consumers regard as fairly close substitutes.

The prevalence of imperfect competition in real life is indicated by the widespread use of competitive advertisement (or, more generally, 'selling devices'—free gift vouchers, competitions, golliwogs in jars of marmalade and jam, and all the rest of it) in order to attract customers. Such phenomena could not occur under either perfect competition or pure monopoly. The seller under perfect competition could gain nothing from competitive selling campaigns, since by definition his product is identical with that of the large number of competing producers; any benefits resulting from the campaign would be felt no more by the firm undertaking it than by any of his rivals, so that the procedure would be pointless. The pure monopolist, on the other hand, has no need of advertisement; he is the sole supplier of his product and there are no rivals to be beaten through these various selling techniques. Of necessity, then, the existence of competitive advertisement and product differentiation in an industry connotes imperfect competition; since these things are absent from singularly few industries, it is clear that in considering imperfect competition the analysis is coming fairly close to the real world at last.

2. MARKET EQUILIBRIUM

As imperfect competition combines features of both monopoly and perfect competition, it is not surprising that the analysis of market equilibrium under imperfect competition indicates that it involves elements of both monopolistic and competitive equilibrium. From the monopoly analysis is taken the falling sales curve facing the individual firm; unlike the producer under perfect competition, the entrepreneur operating under imperfect competition faces a range of market prices rather than a single price determined for him by forces outside his control. If he raises his price, some of his customers will turn to rival brands which they consider to be fairly close substitutes and which they will prefer now that his product has become relatively expensive. On the other hand, he will not lose *all* his customers, since some will still retain a preference for his particular brand of commodity despite its relatively high price.

One important factor in this context may be the existence of transport costs. Customers may prefer the product of a nearby supplier to otherwise identical products being sold some distance away; they may prefer to pay a price for it which is somewhat higher than that charged for rival products rather than incur the expense, or effort, of removing their custom elsewhere. They may be undeterred by an increase in the relative price of a product, also, by the intangible but powerful force of goodwill; customers become attached to a particular seller through habit, or tradition, or through personal and social relationships with the producer.

For any of these reasons, then, a producer under imperfect competition would not lose all of his customers by raising price in comparison with his competitors, as he would under perfect competition. Conversely, by reducing price he would attract some, but not all, buyers who were previously customers of his rivals. The firm in imperfect competition, then, shares with the monopolist the phenomenon of a falling sales curve for its product, and does not face the horizontal sales curve characteristic of the firm operating under conditions of perfect competition.

Unlike the situation under monopoly, however, market equilibrium under imperfect competition involves a consideration of the

conditions for equilibrium of the *industry* as well as of the individual firm. The monopolist is untroubled by the concept of equilibrium of the industry; the monopolistic firm *is* the industry, and so long as it is itself in equilibrium that is an end of the matter. In this respect, however, the situation of imperfect competition is similar to that of perfect competition. For full equilibrium in the market, it is not enough that each individual producer should be in equilibrium; firms are free to enter (or leave) an industry if the profits being made in it are higher (or lower) than the normal level for the economy as a whole, and the industry will be in full equilibrium only when the number of firms in it is showing no tendency to change. (If there is *not* complete freedom of entry into the industry, of course, firms may persistently enjoy super-normal profits. However, there is no need at this stage to attempt to deal with every possible type of situation which could arise in practice.)

Equilibrium under imperfect competition, then, requires the same double set of conditions as under perfect competition—i.e. equilibrium of the firm *and* equilibrium of the industry. As has been seen for both perfect competition and for monopoly, the firm can be in equilibrium only when its profits are maximised, and, as before, this occurs when marginal cost and marginal revenue are equal. The condition for equilibrium of the industry is also exactly the same as under perfect competition—that is, that the prevailing level of profit is equal to the normal level over the economy as a whole, a situation which is established only when average cost and average revenue are equal. Unlike the perfectly competitive situation, however, marginal revenue and average revenue (price) are not identical under imperfect competition, because the monopoly elements in the situation give the individual firm a falling sales curve rather than a horizontal one.

The situation in which the two sets of equilibrium conditions will be realised is illustrated in Fig. 29. At output OM, the equilibrium of the firm is secured by the equality of marginal cost and marginal revenue, while the equilibrium of the industry is assured by the equality of average cost and average revenue. But why should *both* equalities occur so conveniently at the same level of output in this way? The answer to this question is simply that both equalities can occur only when the two average curves—AC and

AR—are tangential to one another as they are at the point *N*. The fact that *AR* (the sales curve) is a tangent of *AC* at *N* *means* that the slopes of the two curves are equal at that point. At any point to

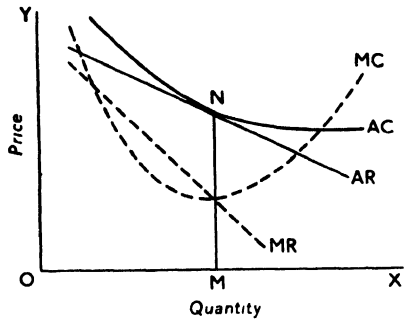


FIG. 29

the left of *N* the curve *AC* is steeper than *AR*; this means that marginal cost must be smaller than marginal revenue—it is pulling the *average* cost down more rapidly than marginal revenue is pulling down average revenue. At any point to the right of *N* the curve *AC* is less steep than *AR*; this means that marginal cost must be greater than marginal revenue—it is pulling the *average* cost down less rapidly than marginal revenue is pulling down average revenue. Hence, at the point *N* where the slopes of the average cost and average revenue curves are equal, marginal cost and marginal revenue must be equal. It is at the point where the average revenue, or sales, curve makes a tangent with the average cost curve—and *only* at this point—that the necessary dual equilibrium conditions are fulfilled.

Two points of contrast between this situation and that prevailing under perfect competition are worth noting. First, price under imperfect competition will be equal to average cost and will therefore be *greater* than marginal cost; under perfect competition, price tends to equality with marginal cost. Second, and more important, the output at which a firm will be in equilibrium under imperfect competition will be smaller than that at which average costs are at a minimum. The output of the firm settles at the point where marginal cost equals marginal revenue and where average cost equals average revenue. Since the firm's sales curve is downward-

sloping, marginal revenue must be less than average revenue at every point. It follows that when the dual equality, shown at output *OM* in Fig. 29, is achieved, marginal cost (being equal to marginal revenue, which is less than average revenue) must be less than average cost, so that average cost is still falling at the equilibrium level of output. (This point is most easily appreciated, perhaps, by trying to imagine the tangency shown at *N* in Fig. 29 occurring with a falling sales curve if the curve *AR* was *not* falling to the right of *N*.) The firm under imperfect competition therefore operates permanently at less than optimum output, since the optimum level of output is that at which goods are produced at the lowest average cost of production.

There is one point of contrast between the equilibrium position under imperfect competition and that under monopoly. The monopolist maximises the profits at a level above the normal for the whole economy, and the more inelastic the demand facing him the greater will be the excess of monopoly profit over normal profits. Under imperfect competition profits are not abnormal; as long as new firms can move into the industry, they will prevent abnormal profits from continuing, while firms will leave the industry if profits fall below the normal level.

3. THE SIGNIFICANCE OF IMPERFECTLY COMPETITIVE EQUILIBRIUM

This analysis brings out clearly the sort of situation which will tend to prevail in an industry characterised by imperfect competition. Each firm will be earning only normal profit, but each will be operating at a level of output which is too small to permit the full exploitation of the economies of large-scale production. In the industry as a whole, there will be too many separate firms producing individual brands of the product, and as a result each will have some degree of unused capacity.

It would follow from this that imperfect competition has substantial wastes necessarily associated with it. On the one hand there are the direct and obvious wastes of competitive advertisement and other 'selling' measures. Firms expend resources—labour, raw materials etc.—in an attempt to influence consumers away from rival products and towards their own, although all the products may

be identical in all important respects. As all firms will be indulging in this selling process, the competing efforts are bound to cancel themselves out to a large extent, so that no-one benefits ultimately except the persons making a living from sales promotion. Even if it is not entirely wasted through this mutual cancelling-out, the expenditure of resources in this way can benefit one producer only by damaging another, so that no net benefit can be gained by producers as a whole. The consumer suffers in all this, of course, because the price of the product concerned will have to be high enough to cover these selling costs; society as a whole suffers through the frittering-away in this way of productive resources which could be used to more useful purpose in the creation of wealth.

Not all advertisement can be condemned as being wasteful and socially undesirable, however. Selling techniques which are educative have a necessary role in a dynamic and developing world; they may inform the consumer of the existence of substitutes, new uses for commodities and prices in different parts of the market, thus helping him to gain the maximum satisfaction from his limited income. Indeed, it is difficult to conceive of new products ever being successfully marketed if producers were unable to bring them to the notice of consumers, and stimulate their taste for them, through all the channels of modern salesmanship. In such circumstances, selling costs are just as necessary an element in total production costs as are those of labour or capital. But only a limited proportion of present-day advertisement is of this strictly informative kind; the greater part is clearly competitive—designed to capture business from rival producers through exploitation of the psychological susceptibilities of the consuming public, rather than through lower prices or higher quality.

The second type of waste and inefficiency incurred as a result of imperfect competition is less obvious to the public eye but probably of far greater absolute magnitude. This is the waste implicit in the existence of too many firms in each industry, all operating at less than their optimum output; the analysis has shown that this is an inevitable consequence of imperfect competition. Society as a whole is bound to suffer if production is carried on at an average cost higher than the minimum which is technically necessary. If the number of competing producers were reduced and the same output

supplied by firms operating at optimum capacity, average production costs would be reduced, which means that resources would be set free to add further to the wealth of society. Wherever monopolistic elements exist, in the form of pure monopoly or of imperfect competition, costs of production will be above the minimum possible and society as a whole will be poorer than it need be.

Having said all this, two important qualifications must be made. The first point is that the important thing for consumer satisfaction is that consumers should have what they want, not what someone else (however benevolent and knowledgeable) thinks they *should* want. The ideal towards which society should try to work, from a purely technical point of view, may be that of a perfectly competitive world where producers, all operating at minimum average cost, supply identical products; it may be, however, that the majority of consumers prefer the sort of world in which any given product is supplied in a large number of brands, each having (or believed to have) its own special characteristics. It has to be remembered that the practice of branding products has a good deal to be said for it from a consumer's point of view. It is much easier to ask for a packet of X soap or a bar of Y's chocolate, and to be certain from experience of the exact nature of the product which will be obtained, than to choose from a large array of un-named varieties of an infinite range of quality and composition differing from shop to shop, which would be the closest approximation to perfect competition which society could ever hope to attain. It is convenient to be able to rely on paying the same price for any given brand wherever it is bought. Above all, the human race derives satisfaction from variety as such, and many people would sooner have a selection of Ford cars and Morris cars and Austin cars and so on—and pay rather more for them—than be reduced to the drab uniformity of a cheaper People's Car for all. In other words, the costs of advertisement, and even of excess capacity, may be costs which consumers as a whole are prepared to pay in order to obtain the choice and variety associated with a world of imperfect competition. The words 'perfect' and 'imperfect' are used by economists in a purely technical sense to describe the degree of competition, not the world which emerges from it. It is by no means certain that most people would *like* a world of perfect competition even if they could have it.

But how much variety do people want, and how much are they willing to pay for it? Are they deliberately *made* to want variety through the slick techniques of modern salesmanship? Is the gain from branding really sufficient to justify the costs which it involves? These are questions which are hardly open to objective analysis; in the last resort the consumer must choose for himself, and the task of the economist is to indicate what is involved in the choice rather than to presume to make it.

There is a second qualification to any general condemnation of the wastefulness of imperfect competition. In the actual world it may be that, if economies of large-scale production are substantial, the entire needs of the market for a commodity can best be met by a relatively small number of firms each producing a large output. If an industry consisted of a great many small firms, so that something approaching perfect competition existed, none of the firms would be producing enough to obtain full economies of scale. The gains to the consumer from a higher degree of competition might thus be more than offset by his loss in having to pay a price sufficient to cover unnecessarily high costs of production. In a complicated world, therefore, it is unwise to conclude that a situation in which competition is imperfect is *ipso facto* undesirable from a social point of view.

FURTHER READING

* ROBINSON, *The Economics of Imperfect Competition*, Chs. 11–14.
E. H. CHAMBERLIN, *The Theory of Monopolistic Competition*, Chs. IV, V & IX.
R. TRIFFIN, *Monopolistic Competition and General Equilibrium Theory* (Harvard University Press, Cambridge, 1941).

CHAPTER 15

The Price Mechanism

1. THE ALLOCATION OF RESOURCES

Occasionally men are encountered who have organised and managed enormous factories, employing perhaps ten thousand people, in the consistently successful production of some particular commodity. Such people are regarded as being of outstanding administrative ability, as indeed they are; it is an immensely difficult task to maintain an efficient co-ordination of thousands of different people making the various parts of a commodity at different speeds, different times and often different places. Every refrigerator, or motor-car, or television set that comes off the assembly lines of a modern factory is in itself a miracle of administrative as well as technical ability.

It is an extraordinary thing, then, that the vastly more complex economic system as a whole continues year after year to meet an infinite number of demands for an infinite number of commodities with the incredibly high degree of efficiency which has come to be accepted as commonplace. If the daily paper fails to appear through the letter-box one morning people are very surprised, whereas they are not in the least concerned when it *does* arrive as usual, although this is a much more astonishing event. Consider the extraordinary co-ordination of effort which lies behind the simple fact of the delivery of a morning newspaper. Trees have been grown and felled in Canada or Scandinavia; they have been processed into woodpulp and transported across thousands of miles of sea; the pulp has been turned into newsprint in paper-mills; the news has been collected via an intricate network of national and international communications; type has been set up and the news printed with machinery manufactured in a hundred different engineering factories; road and rail services of great complexity have been used to bring the news-

paper to the door of the consumer. And this is a highly simplified account of a process no more complicated than a hundred thousand others which continue, day after day, in order that when a person asks for marbles or beans or hairpins or razor blades or anything else in his local shop the goods he wants will in fact be there.

If organising the manufacture of a single commodity in a single factory is a respectable administrative feat, therefore, the intelligent co-ordination of an economy comprising, in the case of Britain, a working force of some 22 million people must involve a fantastic degree of organisational ability. Yet, in an economy such as that of the United Kingdom no single person acts as a co-ordinator and controller—which is just as well, since the task would be far beyond the range of human ability. But how can the economic system function without the perpetual chaos that would result, for example, if a single railway system tried to operate with no-one in the signal box? The answer is that *price* does for the economic system exactly what the signalman does for the railway system. It is through the *price mechanism*, as it is called, that the co-ordination of activity in the whole economic system is achieved. The analytical system built up in the preceding chapters can throw a good deal of light on how this is done. It will be easiest if the process is first outlined within the framework of something approaching perfect competition; in the following section the picture will be modified to allow for some of the imperfections found in the real world.

The theory of demand indicates how the consumption of wealth is guided and influenced through the medium of price. The total consumption possible for each individual is determined, in the first place, by the operation of price. Each consumer has a certain income in the form of a specific sum of money becoming available to him every week, or every month, or every year. The price of a commodity is the amount of money which he has to hand over in order to obtain possession of it. The total volume of goods and services which he can consume is therefore determined for him; he can go on consuming until all his income has been spent. More important, the price mechanism allows him to distribute his expenditure amongst those goods and services which he believes will give him the maximum possible satisfaction. The earlier analysis has shown that by arranging his expenditure so that the marginal utility he

derives from each commodity is everywhere in the same ratio to price, the consumer will obtain the greatest possible amount of satisfaction. From the consumption side, therefore, the price system limits the total amount which any individual can consume, but at the same time leaves him complete freedom of choice so far as the pattern of his expenditure is concerned. (Although, as was noted in the previous chapter, that choice may be influenced by advertisement and other techniques of sales promotion.)

The theory of production helps to explain how the preferences of consumers as between different commodities are transmitted to producers and thus influence the flow of the factors of production to each industry. Consumers will be willing to pay different prices for different quantities of a commodity; entrepreneurs can therefore estimate the marginal revenue to be derived from various levels of output. Given the technological conditions governing production and the prices of the various factors of production, entrepreneurs will be able to calculate marginal cost at various levels of output. Marginal revenue and marginal cost will together determine how much of each commodity will be produced by each entrepreneur. The analysis of Part V will show that the prices of the various factors of production will determine the proportions in which the various factors are combined in each enterprise; this will in turn determine the distribution of total income.

So the complete picture takes shape; total income, through the price system, determines total consumption; the structure of consumption and production is reflected in the pattern of relative prices, while total income is determined by the numbers of units of the various factors at work and by their respective prices. At every stage, price enters into the process as the expression of the outcome of the pressures of demand and supply in each sector. Every consumer and every producer exercises an individual choice in his use of resources, although in cases such as those discussed in the following section the government also intervenes. Through the price system the decisions of both consumers and producers are brought together in such a way that the factors of production are attracted into the uses in which they can best meet the needs of consumers.

2. OPTIMUM ALLOCATION

In the hypothetical world of perfect competition, which, fortunately or unfortunately, is far removed from the world of reality, the free operation of the price mechanism would not only co-ordinate all the infinite number of economic activities going on at any moment, but would do so in the ideal way. It will ultimately secure the *optimum* allocation of all economic resources—that is, guide all of the factors towards their most productive uses, and hence ensure that the maximum satisfaction is being obtained by society from the resources at its disposal.

It will be remembered that under conditions of perfect competition price will everywhere tend to equality with marginal cost and every producer will operate at his point of minimum average cost. The fact that average costs are everywhere at their lowest point means that every firm is operating at peak efficiency; there is no loss anywhere because of firms operating at too low an output (and leaving unexploited some potential economies of large-scale output) or at too high an output (and incurring diminishing returns).

Not only will each firm operate at its optimum output. For the economy as a whole, the equality of marginal cost and price will ensure that productive resources are distributed between commodities in exactly the right pattern. Price, it has been seen, reflects the extra satisfaction derived by consumers from the last unit of a commodity they consume; marginal cost measures the extent to which resources are used up by the last unit of a commodity produced. If the marginal cost of a commodity exceeded its price, society would be losing more in the form of resources used up than it would be gaining in the form of extra satisfaction, and the output of the commodity would therefore be too high; if price exceeded marginal cost, society would be receiving more in the way of satisfaction than it would be losing through the loss of resources in the process of production, and output of the commodity would be too low. Total satisfaction will always be increased if resources are switched away from industries where the extra satisfaction to consumers by their output was less than their cost, and towards industries where the extra satisfaction provided by their output exceeded their cost. Such gains would continue until marginal cost

was everywhere equal to price. Under perfect competition, then, all resources would be used in the best possible way and in the best possible use.

Does this mean that in the 'competitive', free-enterprise economies existing in the real world this ideal state of affairs is actually achieved? Unfortunately it does not, so that it is impossible to draw the dogmatic conclusion that if all economic activity were determined solely through the operation of free market forces—i.e. the price mechanism—all would be for the best in the best of all possible worlds. There are three main reasons for this.

First, and most obviously, perfect competition is seldom, if ever, attained in reality, since it postulates a great many conditions which it is impossible to attain in the world as it exists. Consequently, some degree of monopoly is to be found in almost all industries. The implication of this, of course, is that price and marginal revenue cease to be equal, and since entrepreneurs operate at that output where marginal cost equals marginal revenue (which is always less than price) the result will be that marginal cost will be lower than price in an imperfectly competitive world. *How much* marginal cost will be below price will depend on the slope of the firm's sales curve (i.e. on the degree of monopoly power it enjoys), but as long as it is below price at all, resources cannot be in their optimum use. The economic welfare of society will be smaller than it would be if output were expanded until price fell to equality with marginal cost. And, as was noted in the last chapter, other direct wastages—advertisement etc.—will occur when competition is imperfect.

The second factor is the existence of gains or losses to society as a whole which individual consumers cannot, or will not, appreciate and which therefore fail to be expressed through price. For example, the maintenance of the Armed Forces could not be left to private enterprise because no individual can measure the value of such services to society as a whole; not all parents would be willing to pay for the education of their children, because not all of them would appreciate the value of education for their own children or for future generations of society; the provision of roads in isolated areas could not be left to private initiative, because they would seldom prove to be profitable in the monetary sense.

Over a great many fields, then, social or political considerations may be important or even dominant, and it is generally agreed that in such cases the State has the right, and even the duty, to interfere in some way with the operation of free market forces. The State may work *through* the price mechanism, if it is believed that the true benefit accruing to society as a whole (including future generations) from the consumption of a commodity is not accurately measured by the price established by ordinary market forces. For example, it may impose a tax on a commodity whose consumption is considered undesirable for some reason (e.g. a tax on motor-cars to discourage their consumption at home and encourage their export); conversely, it may subsidise commodities which it wishes to see consumed on a bigger scale (e.g. milk). Occasionally it may prohibit altogether private production and exchange of a commodity, as in the case of drugs or poisons.

Finally, the allocation of resources achieved by perfect competition would be an optimum in the true sense only if the distribution of income prevailing in the economy was the optimum distribution. It will be remembered that even if every consumer of a commodity is willing to pay the same price for it, it still cannot be said that all consumers derive equal marginal utility from it. A rich man may be satisfied with a low marginal utility from a commodity because the sacrifice implied by its price is a small one to him; a poor man may derive a high marginal utility from the same commodity at the same price. The price mechanism will allocate resources in the way that is indicated by demand; it is conceivable that, in an economy where a few are rich and the majority are poor, resources could be drawn by the pressure of demand into the production of luxury yachts and private swimming-pools when (by any reasonable standards) they would have been better employed in clearing slums and equipping hospitals.

For all these reasons, then, there exists in most economies some degree of State intervention designed to correct the distortions introduced into economic activity by imperfections of competition or the maldistribution of income, or to make provision for needs of a social, strategic or political nature. How far such intervention should proceed is largely a question of personal judgement (expressed through the appropriate political machinery). Society as a

whole must arrive at its own conclusions concerning, first, the relative efficiencies of private and public enterprise in the achievement of purely economic ends, and, secondly, the relative importance in the use of resources of economic and non-economic considerations. In almost every society it is widely accepted that only the State can effectively direct resources to the achievement of ends such as defence and law and order; in a few countries the State has assumed control of all economic activity to the exclusion of private enterprise. In most countries, the position is somewhere in between—i.e. they have 'mixed' economies, as they are called. The relative strength of private and public control in the mixture, of course, is a question for political decision, because ends rather than means are involved, and in such decisions the economist has the right to be heard as an ordinary citizen but no more.

3. CHANGES IN SUPPLY AND DEMAND

In view of the fact that most present-day economies are 'mixed', in the sense that resources are allocated partly by the decisions of private individuals and enterprises and partly by government intervention of one form or another, it is useful to consider the way in which adjustments occur in both the private and public sectors. This section will be concerned with the private sector—that is, with the adjustments carried out through the operation of a free market.

Assuming that the private sector is characterised by some degree of competition, suppose that an increase occurs in the supply of a commodity, as the result of an expansion in an industry or of some technical development enabling a greater output from existing capacity, without any corresponding change in the demand for the commodity. If the extra output is to be sold in the market, normally price will need to be reduced; the law of diminishing utility suggests that consumers derive a smaller marginal utility from a commodity as its total supply is increased so that they will be willing to buy extra units only if its price falls. If the increased output is due to expanded capacity rather than a reduction in costs, the fall in price will drive profits in the industry below the normal level; firms will move out of the industry, and both supply and price will gradually revert to their original level. If costs have been reduced

through some technological improvement, however, the fall in price will continue until it becomes equal to the new (lower) level of average cost. When this point is reached, output will cease to expand and the market will rest in its new equilibrium position.

If supply contracts for any reason, the opposite sequence will occur. The commodity will have become scarcer in relation to others and consumers, warned of this by the rise in its market price, will be encouraged to reduce their consumption of it. If the scarcity is due to some fundamental technical change (i.e. a rise in the cost of production) price will continue to rise until it equals the increased average cost, and output will contract until this point is reached. If costs have not risen, producers will be encouraged to enter the industry by the rise of price above average cost and the consequent emergence of excess profits. As new firms come in, price will be driven down again to its original level and the previous equilibrium will be restored.

In this process of adjustment to changed conditions of supply, two points are worthy of note. First, the analysis assumed some degree of competition—perfect or otherwise—so that it was taken that firms could enter or leave an industry in response to changed profit-levels. If no competition existed, however, adjustment could only take the form of changes in a monopolist's price and output; in general, the movement of price and output in response to any given change in supply conditions under monopoly can be expected to be smaller than if some degree of competition exists. Secondly, the process of adjustment, if allowed to operate freely, will ensure that if variations of efficiency exist amongst producers, the relatively inefficient entrepreneur will be driven out first in any movement of contraction and brought in last in an expansion. If prices fall relatively to costs, profits will turn into losses first of all amongst producers whose costs are relatively high, so that the pressure on such producers to leave the industry will be especially great. Conversely, if prices rise in relation to costs, the super-normal profits will be smallest for relatively inefficient entrepreneurs, and the incentive to expand output or to enter an industry will be comparatively weak.

With changes in demand, the sequence is similar. A rise in demand will force up prices, causing profits to increase in the industry concerned. This will indicate to producers that the commodity

involved is now wanted with a greater intensity than before; existing producers will expand output under the spur of higher rates of profit, while new firms will be attracted into the industry by the same cause. If consumers' total expenditure is unchanged throughout, this increase in demand for one commodity will be paralleled by a decline in the demand for other commodities; if consumers spend more out of a fixed total on one thing, then they must spend less on others. In some other industry or industries, therefore, a precisely opposite process will be occurring; demand, prices and profits will have fallen, and producers will be using up less resources for the production of their commodities, setting those resources free for the expansion in the industry faced with an increased demand.

Through shifts in demand and supply, then, the economic system is constantly adjusting itself to changes in the relative scarcity of different commodities, changes which can occur through alterations in either the supply conditions of a commodity or the strength of the public's desire for it. It will have been noticed that a crucial role in the adjustment process is played by profits. It is because of a rise or fall in the rate of profit that plans are put into operation for the increase or reduction of output. The rewards of other factors of production (labour, land etc.) are also liable to fluctuate in response to changes in demand, encouraging those factors to move out of uses for which they are wanted less urgently than before and into uses where supply is unduly small and more factors of production are required. But since it is the entrepreneur who performs the function of engaging and organising the other factors, it is through *profits* that the impulses arising from changed demand or supply conditions are first transmitted. In the modern world where wages, rent and interest commitments are not easily varied (for reasons discussed at a later stage), a change in price affects profits first and in the greatest degree. This is emphasised by the characteristic observed when the factor of enterprise was first discussed—enterprise is the only factor whose reward can become not merely zero but negative. If a man's wages fall to zero in a job, he will soon abandon that job, but if he actually has to pay for the privilege of working in someone else's factory he will abandon it even more rapidly. The entrepreneur may be in just that position; his penalty may be not merely an absence of reward but a positive loss, and he

is therefore the sensitive point through which the forces of adjustment work most quickly and strongly.

4. TAXATION AND RATIONING

The methods by which the State can influence the use of economic resources are difficult to describe briefly precisely because they are numerous and may be based on considerations outside the range of purely economic reasoning. In its use of resources which it employs directly, for example—in the Civil Service or Armed Forces —the State is not impelled by the motive of maximising economic gain, which is the motive force economic analysis has to assume at work in privately-owned enterprises. It is motivated by concern for the social welfare of the people, or for the safety of the country, or even in some cases by the aim of maintaining or increasing its physical power over its citizens. Motives of this sort cannot be handled by economic analysis; the economist can simply say that resources used for such purposes are lost to the direct creation of wealth, and that the responsibility of the government is to try to balance out (by some criteria) the gain accruing in social and political welfare and any loss suffered in economic welfare.

Much the same is true of other means by which the State intervenes in normal economic activity in the pursuit of political ends. Tariffs may be imposed on imports, for example, in order to protect an industry (such as agriculture) the prosperity of which is considered essential on strategic or social grounds. The state may control the use of raw materials or labour in order to favour defence production, exports, or the development of particular regions. Here again the problem is one of having to balance a political or social gain against an economic loss, and such questions are ultimately matters for individual human judgement rather than scientific analysis of the kind which economists seek to employ.

A more extensive form of government intervention, however (at least in peace-time), is the manipulation of market prices through internal taxation or subsidies, and, to a much lesser extent, the control of demand through rationing. In the former case the effects (although not, of course, the motives) are amenable to economic analysis because the adjustment occurs through the medium of the market price. The imposition of a tax can be treated as an increase

in costs of production, while a subsidy can be treated as a reduction in the cost of production. As was established at an early stage, the total revenue brought in by a tax on a commodity (and, conversely, the total amount expended in the form of subsidy payments) will vary with the elasticity of demand for the product. With taxation, the consumption of a commodity will be much affected by the imposition of a tax if its demand over the relevant price range is comparatively elastic, and the revenue brought in will be small; the output and consumption of a commodity with an inelastic demand will be little affected by an indirect tax, and the total tax revenue will be relatively large. The effects of a subsidy are similar so far as consumption and output are concerned, being greater the more elastic the demand; it follows that total subsidy payments will be larger the more elastic is demand.

Will the price of a commodity change to the same extent as a change in the tax or subsidy associated with it? This depends on two sets of forces. On the demand side, the elasticity of demand, and hence the slope of the marginal revenue curve, will determine the extent to which output will fall (and price rise) as the result of the imposition of a tax; the price must rise until marginal revenue and the new level of marginal cost are again brought into equality. In general, the more elastic the demand for a product, the less will price increase (and the more output will fall) as the result of an indirect tax. There are two limiting cases; where elasticity of demand is zero, the full tax will be added on to price and output will be unaffected, while, if demand is infinitely elastic, price will be completely unaffected and the tax will be borne wholly by the producer, or output will be reduced to zero (although such a combination of events is obviously inconceivable in the real world).

Secondly, the effects of a tax on price will depend on the conditions of supply. If a commodity is produced under conditions of diminishing returns, the rise in price caused by the tax will be offset to some extent by the lower costs of production as output falls. Conversely, if increasing returns prevail, price will tend to rise by more than the tax, since average costs of production will increase with falling output. The effects of a subsidy can be analysed in a similar way, because a subsidy is merely a negative tax.

Taxation, positive or negative, is a means whereby production

and consumption are influenced through price. Rationing is an attempt to achieve a similar end by regulating the demand at any given price. If a commodity is in short supply (as it usually will be if rationing is considered necessary) the supply curve will have shifted to the left, and price would normally be driven up. The effect of rationing is to cause a similar shift to the left of the demand curve, preventing or moderating the rise in prices. In this way the shortage does not result in the relatively poor members of the community being pushed out of the market by the relatively rich, as might happen if the forces of demand and supply were allowed to work unchecked. In wartime conditions, when the commodities in short supply may be essential to the health and strength of the community, it may be expedient to share out the shortage, so to speak, through rationing in order to ensure that everyone, and not just the fortunate members of society who can afford high prices, obtains an amount necessary for his or her health.

In normal times, however, rationing has several disadvantages. One is that a substantial amount of labour and other resources may be required for its administration and enforcement, and such resources would probably add more to the wealth of society if they were employed in productive uses. Secondly, by arbitrarily limiting the consumption of everyone, the principle of maximising satisfaction through free choice is contravened. I might prefer more bacon and less sugar than the rationing system allows me, while another person might prefer less bacon and more sugar. Given a free choice each consumer would spend his money appropriately and get the most satisfaction out of his limited income. If rationing is imposed, consumers are prevented from achieving their optimum patterns of consumption, unless they indulge in the (usually) illegal and inefficient process of bartering or selling ration coupons. Thirdly, by preventing a shortage from having its normal effect of a rise in price, rationing impedes the normal forces of adjustment—high profits, stimulating higher output and new entrants to the industry. In special situations (such as wartime or other crisis periods) this process of adjustment may be impossible, or able to work only very slowly, so that a rise in price would not serve this useful purpose. In normal times, however, this consideration is an important one which is not lightly disregarded. Unless it is unavoidable, measures

to deal with a shortage should not have the effect of preventing its disappearance.

In the light of the analysis of consumer behaviour, then, it can be concluded that if a government seeks to protect or favour a particular section of the community, it is generally better to do so through an adjustment to the *incomes* of the section concerned; interference with the *price* of a commodity has the effect of dispersing the benefit or burden indiscriminately amongst all consumers, and of compromising the exercise of free choice. This is not to say that in particular cases the influencing of market prices cannot be justified; for example, the goverment may wish to increase the consumption of milk by children, or discourage the consumption of harmful commodities, and direct intervention in the market for those particular commodities may be necessary to attain this end. Each case has to be considered on its own particular merits. But it will usually be found that the use of indirect taxation—i.e. the taxation of particular commodities—has to be justified on the grounds that if the government raised *all* the revenue required through direct taxation (i.e. taxes on persons or enterprises, rather than commodities) there would be a possibility of serious ill-effects on incentives.

FURTHER READING

MARSHALL, *Principles of Economics*, Bk. V, Ch. 13.

SIR D. H. ROBERTSON, *The Control of Industry* (Cambridge University Press, London, 1928), Chs. VII–VIII.

L. ROBBINS, *The Economic Problem in Peace and War* (Macmillan, London, 1947), Chs. I and III.

J. E. MEADE, *Planning and the Price Mechanism* (Allen and Unwin, London, 1948), Ch. I.

PART V

THE THEORY OF INCOME DISTRIBUTION

CHAPTER 16

Wages

1. THE DEMAND FOR LABOUR

So far the analysis has been concerned with the factors which determine the price or value of different commodities. Now it is necessary to turn to the question of the way in which this value is shared out between the various factors of production which are involved in the output of those commodities. It has already been noted that the total proceeds from the sale of commodities are used to pay the wages (using this term to include salaries) of labour, the rent of land, and interest on capital; whatever is left after these costs have been met constitutes the reward for the factor of enterprise—i.e. profits. In what way are these various shares determined? It will be convenient to leave the discussion of interest on one side for the moment, since it is a purely monetary phenomenon and can be fully understood only after the examination of the monetary system in Part VI.

The analysis has shown that supply and demand together determine the prices of goods; similarly, they determine the prices of factors—i.e. wages, rent and profits—and the same type of analysis can be employed. It is convenient to begin with wages, as this is the most important element in total costs; at the present time more than two-thirds of the annual national product of the United Kingdom is paid out in wages and salaries. It is important to remember, however, that wages are more than simply an element in cost, or the price of a factor; they are also the main form of income for the great majority of the population. As a result, many non-economic influences enter into the determination of wages. The purely economic factors bearing on the wages paid to different kinds of labour can be set out, but the conclusions established by doing so will have to be modified subsequently in order to take account of these other

factors which play an important part in the actual world. Again, the analysis can begin by discussing wages in general terms, but it will also need to examine the forces which govern the relative wages paid to different kinds of labour.

What constitutes the demand curve for labour? In the case of commodities it was possible to relate the quantity purchased to the utility derived by consumers from the commodity concerned. The people who purchase labour, however—i.e. the entrepreneurs who pay for the services of workers in their enterprises—derive no direct utility from the labour with which they are supplied. Labour is wanted by them not for itself but for its contribution to the production of the goods they are supplying. The utility of a unit of labour to the entrepreneur is represented by the value of the output which that labour can produce. If a certain man can produce in, say, a week's work goods from the sale of which the entrepreneur will receive an amount greater than the wages he has to pay out (after meeting any other costs involved) then that man's labour will be purchased; on the other hand, if the man cannot add as much to the value of output as he has to be paid in wages, then the entrepreneur will not find it worth his while to employ him.

Along the vertical axis for the demand curve, then, can be measured the value of the marginal product of labour, as in Fig. 30—that is, the net revenue derived by the entrepreneur from the sale of the output resulting from the employment of each unit of labour. Along the horizontal axis, of course, is measured the quantity of labour employed—for the sake of simplicity the unit of labour can be taken as one man-week.

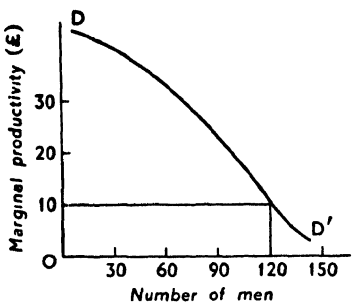


FIG. 30

Now this marginal productivity curve—the demand curve for labour by an industry—is shown as sloping downwards from left to right. Why is this? The answer lies partly in the law of diminishing returns, which has already been discussed. The curve (like all demand curves) is drawn on the assumption that other things remain equal; in this case it is assumed that the

amounts of the other factors employed in the industry are unchanged, as well as the state of demand for the product, and of technological knowledge. It has been seen that if increasing amounts of a factor (of constant quality) are applied to a constant amount of another factor, or group of factors, the extra physical product resulting must progressively diminish. Hence the curve of the marginal productivity of labour will be falling as the amount of labour employed increases (although it may first of all rise to a maximum point in the early stages, of course).

A second reason for the downward slope of the labour productivity curve is that for the industry as a whole (and for the individual entrepreneur if competition is imperfect) the addition to total revenue derived from each unit of output will decline as output rises, because the price of the commodity will need to be reduced if an increased output is to be sold. The *marginal revenue* derived from each unit of output after the first, as was shown in Part IV, is bound to be less than price and is also bound to fall as output expands. What matters to entrepreneurs, of course, is not the *price* at which labour's product is sold but the *additional revenue* which it brings in. Labour will be employed only if its wage does not exceed this latter amount, which progressively diminishes as more labour is employed even if the physical productivity of labour is constant.

If the assumption of other things being equal is removed, then the same procedure as that adopted for the curve of the demand for a commodity must be followed: the entire curve must be re-drawn. If the amount of capital employed in the firm concerned is doubled, for example, the physical productivity of labour will almost certainly be increased, and the whole curve will therefore shift to the right. On the other hand, if there should be a fall in the price of the commodity being produced, the whole curve will shift to the left, since the *value* of productivity in physical terms (i.e. tons of coal or pairs of roller skates produced per week) will have diminished, and it is the value of labour's productivity which matters to the entrepreneur.

If a certain wage-rate is given, then the demand curve will indicate how much labour the entrepreneur will be willing to engage. In Fig. 30 it is assumed that the ruling wage-rate for the type of labour concerned is £10 per week; the entrepreneur, of course, will

have to pay *all* units of labour at the same rate, whatever their marginal productivity, just as a consumer obtains all units of a commodity at the current market price, whatever the level of marginal utility he derives from them. The curve shows that at this wage-level, 120 men will be employed. If entrepreneurs employed fewer than this, they would not be maximising their profit, because each man engaged is still adding more to the value of output than entrepreneurs would need to pay him in wages; if entrepreneurs employed more than 120 they would again fail to maximise profits, since they would be paying every man after the 120th more in wages than he was adding to the value of total output.

2. THE SUPPLY OF LABOUR

The total supply of labour in an economy is more or less directly related to the size and structure of the total population. The forces which determine it are very complex and by no means entirely economic in nature, ranging from the prevailing social conceptions of the meaning and value of family life, the costs of rearing and educating children, and welfare provisions for the sick and aged, to general attitudes towards the employment of women and children, legal regulations governing hours of work and the nature of retirement practices. The supply of labour to a particular industry, or firm, however, is more easily analysed. Bearing in mind the basic assumption that in his economic activity man acts so as to maximise his own well-being, the analysis can start from the approximation that the higher the wage obtainable in an industry the greater the amount of labour which will offer itself for employment in it. In other words, the supply curve of labour will be rising from left to right like the curve SS' in Fig. 31. An increase in the wage might encourage the existing labour force to work longer hours, and will also draw into the firm or industry men who were previously employed elsewhere and who are attracted by the higher wages. For the economy as a whole, of course, the total labour supply cannot be increased in this second way, but higher wages might attract into employment people who were previously not seeking employment—married women, for example, or retired people.

The 'normal' supply curve of labour will therefore be of this upward-sloping kind. But just as 'perverse' demand curves arose in

cases where an income effect was superimposed on the price, or substitution, effect (see Chapter 3, section 5), so an income effect may operate in the supply of labour. The normal substitution effect in

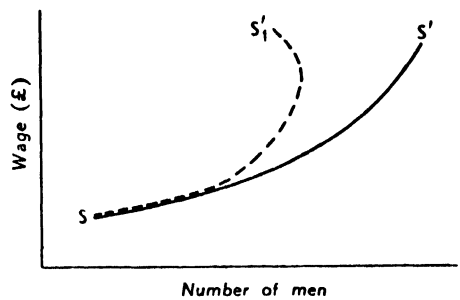


FIG. 31

the case of labour is that a rise in wages induces a greater supply of labour, and a fall in wages induces a smaller supply. It becomes more, or less, worthwhile for a man to sacrifice his leisure and work longer hours. But an income effect can also operate in this way. If wages rise, a man can now earn any given income by working fewer hours. If he aims at maintaining a certain standard of living (requiring a given income per week or month) rather than obtaining the *maximum* possible income, the rise in wages will mean that he will be able to enjoy his 'target' standard of living with fewer hours of work, so that the supply of labour will be reduced rather than increased. In effect, the wage increase has made the worker better off and he proceeds to use his extra wealth to buy more leisure—e.g. Saturday mornings or a mid-week afternoon—just as the income effect of a fall in the price of bread is seen in a reduction of the consumption of bread and an increase in the consumption of higher-quality foods. Conversely, a fall in wage-rates may mean that *more* labour is supplied than before, as workers attempt to maintain their incomes.

If this income effect is pronounced the supply curve may become backward-sloping at least over part of its length—like SS'_1 in Fig. 31—reflecting the fact that over a certain range of wage-rates the supply of labour will decrease as wages rise. For any single firm or industry, the income effect is unlikely to be dominant; it would

probably be more than offset by the inflow of labour from other industries stimulated by higher wages (or vice versa). For the economy as a whole, however, or for industries which do not easily attract outside recruits (the coal-mining industry may be an example), the income effect may be important. Indeed, it is a feature of advanced economies that the working hours of labour tend to fall; high productivity brings high wages and living-standards—and also a demand for increased leisure in which to enjoy them.

3. THE DETERMINATION OF WAGES UNDER COMPETITION

Bringing together the demand and supply curves, the way in which the level of wages is determined in any industry can be seen. In Fig. 32 is shown a demand curve DD', relating the marginal productivity of labour to the numbers employed; this indicates how much the employers are willing to pay for different amounts of labour. The supply curve, SS', shows the amounts of labour which will be forthcoming at various wage-rates. If wages are competitively determined—that is, assuming away for a moment the pos-

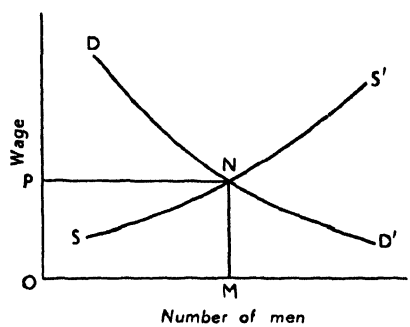


FIG. 32

sible influence of Trade Unions, government wage legislation etc. —the wage-rate will tend to settle at OP, because only at this wage is the demand for labour in the industry equal to the amount of labour being offered (i.e. OM units). If the wage was above this level, the demand would be insufficient to absorb the total amount of labour being offered at that wage; unemployment would result, until the competition for jobs forced the wage down again. If the

wage were below this level the supply of labour would fall below the demand at that wage; the pressure of competition amongst employers to obtain labour would eventually force wages up to the equilibrium point. All this assumes that wages are determined solely by the free play of competitive forces; in the real world, of course, this is by no means always true, and the process of wage determination is consequently a good deal more complex.

In a situation of perfect competition the individual firm (being an infinitely small component of the industry) would be faced with the current market price for the labour used in the industry, and would be unable to influence the level of wages by increasing or diminishing its demand for labour. The supply curve of labour to the firm would therefore be a horizontal line through the market wage rate. For the industry as a whole, however, a shift in the demand curve for labour will significantly influence the level of wages, the exact degree of movement in wages in response to a change in demand depending on the elasticity of the labour supply curve.

Two other points are of some importance. First, an individual firm, or even an industry, can usually treat its wage level as independent of the demand for its product. If wages in that industry are increased, the income, and therefore the expenditure of its workers, will rise. The workers in a particular industry, however, are unlikely to form a considerable element in the market for the commodity it produces, and the impact of a wage increase on the market demand curve facing the industry concerned will probably be insignificant. For the economy as a whole, however, the position is different. A *general* wage increase is almost bound to be followed by an increase in demand generally, so that demand as a whole cannot be considered to be independent of costs of production. As will be seen in Part VII, this fact has important implications for policy measures aimed at securing stability of the general price level and of employment.

The second point relates to the relationship between wages and the productivity of labour. So far these have been treated as if they were quite separate, with productivity as one of the independent variables determining wages. In many cases, however, the position may be reversed, and wages may influence productivity. If a low

level of wages means a low standard of living, with inadequate housing, education and nutrition, it may well result in a low standard of efficiency amongst the labour force. Many employers would propound the principle of the *economy of high wages*, meaning by this that if a man is well paid, so that he enjoys a high standard of living, his working efficiency will be raised and with it his productivity. It may sometimes be necessary, therefore, to break this vicious circle of low productivity resulting in low wages which in turn cause low productivity. It seems that a realisation of this fact underlay the attempts at government intervention in wage determination in Britain early in the present century. And it is important to remember that it is no mere coincidence that the labour force of the United States is at once the highest-paid and the most productive in the world; American wages are high because productivity is high, but the high productivity in turn owes something to the high wages.

Summing up, then, it is obviously of considerable importance that a clear distinction should be maintained between wages and labour-costs per unit of output; while the worker is primarily concerned with the total wages or salary he receives, the employer is interested primarily in the cost of the labour involved in making each unit of output. The two things are not necessarily correlated. For example, if in one factory men are paid £10 a week and each produces an average of 250 units of output a week, while in another producing the same commodity men are paid £8 a week and each produces an average of only 160 units, the labour costs per unit are obviously lower in the first factory than in the second, despite the higher level of wages prevailing in it. The fact that the level of wages in a country are very high by no means implies that the commodities produced in that country are more expensive than those produced elsewhere; if this were so, indeed, it would be impossible to explain the success of United States exporters in world markets.

4. RELATIVE WAGE-LEVELS

If wage-rates are determined by the interaction of the supply of labour and its marginal productivity, it might reasonably be expected that they would tend towards equality, labour moving into highly-paid industries and moving out of low-paid industries until

relatively little variation of reward remained. Admittedly, some degree of inequality would always be expected, since some people have greater innate mental or physical ability than others, and therefore have a higher productivity. Most of the variations *within* any given occupation could certainly be explained on some such grounds.

This factor could not explain the wide differences which exist between different occupations however—there is no special reason why all the men of innate ability should cram into one or two occupations and desert all the rest. Yet such differences do exist; in the Britain of 1956, for example, an agricultural labourer would consider himself lucky to earn £350 a year, an experienced worker in the engineering trades would expect at least £750 a year, while the average doctor would be disappointed to receive much less than £2,000 a year. To some extent such differences in annual earnings may be due to differences in hours of work—i.e. in the number of labour-units supplied by an individual—but this would not go far in explaining the inequalities of earnings which exist. What are the causes of these differences?

First, the wage-rate may not be synonymous with the actual value of payments for labour. On the one hand, labour may receive certain rewards not included in the wage-rate; the coal miner receives concessionary coal, the farm worker obtains milk, meat or eggs, the transport worker is allowed free travel, the civil servant has valuable pension rights, and so on. On the other hand, the worker may have *deductions* to make from his wages—to buy tools or uniform necessary in his work, or to compensate for expenses of travel or training and education which someone, if not he himself, has incurred on his behalf.

Secondly, the nature of the employment has to be taken into account. A high wage may be necessary not so much because of high productivity but in order to compensate a worker for bad working conditions. The work may be dangerous or unhealthy (like coal mining) or dirty and unpleasant (like sewerage disposal), or the employment may be insecure because of seasonal or other fluctuations in demand and output. Another factor is that a job may involve a high degree of responsibility meriting special reward, or may involve a long and expensive period of training. This latter

factor may be especially important in maintaining a scarcity (relative to the prevailing demand) of people possessing 'professional' skills—lawyers, engineers, architects etc. Conversely, a man may willingly accept a relatively low wage because of the attractiveness of the work; it may have high social esteem, which he values; or it may provide an opportunity for the able few (amongst whom almost everyone automatically includes himself) to earn an extremely high income—barristers as a whole are not well paid, but many enter the profession despite this fact because of the fabulous earnings of a small minority of the profession.

Thirdly, a group of advantages or disadvantages may be associated with the general environment within which the employment is situated. Certain geographic areas are widely regarded as being more pleasant to work and live in than others; the opportunity for wives or children to obtain employment varies from place to place and may be an important consideration for a family; rents are generally lower in the country than in large towns; housing accommodation may be easier to obtain in one area than in another. All these factors can enter into the mind of a worker when balancing up the attractiveness of different occupations.

It is closer to the truth, then, to say that labour looks at the *net advantages* of different occupations, rather than to imagine that money wage-rates are the sole attraction for labour. Analytically, all these other factors enter into the shape and position of the supply curve. For two industries, offering identical wages but possessing different degrees of non-monetary advantages, there would be separate supply curves, the industry possessing the greater net advantage from the point of view of labour having a supply curve further to the right than the other.

Having introduced all these complications, it still does not follow that in the real world labour moves from industry to industry until the net advantage from employment has been brought to the same level in all. There are the influences of Trade Union and government intervention which may prevent the forces of competition from fully working themselves out; more will be said of these in succeeding sections. There is also the pressure of convention; certain wage differentials, as they are called, have existed for many years and have become established in the minds of both workers

and employers as necessary and normal. Thirdly, there is the influence of public opinion; certain occupations, such as teaching or medicine, are regarded as being of special responsibility and in some rather vague sense *entitled* to a relatively high reward. Most important of all, there are the hindrances to mobility, both between areas and between occupations, which were described in Chapter 6. The effect of all these influences is to break the total labour force into 'non-competing groups', and by restrictions on the free movement of labour to prevent the equalisation of the net advantages accruing from different occupations. It is mainly because of this that there is still substantial truth in the generalisation that the occupations which are most attractive, in all senses, are also the most lucrative.

5. TIME RATES AND PIECE RATES

The word 'wages' covers a good many different forms of earnings of labour. It is usual to distinguish, for example, between *wage-rates* and *earnings*, the former being the basic minimum rate specified for a certain occupation (generally expressed as a rate per hour), and the latter the total amount actually paid to labour, including overtime pay and bonuses of all kinds. Another important distinction is that between time rates of wages and piece rates. *Time rates*, as the term implies, refers to wages which are paid in respect of a specified period of time—an hour, or a week containing an agreed number of working hours—irrespective of the amount produced during that time. *Piece rates*, on the other hand, are paid on the basis of the number of units produced by the worker, without regard to the time taken to produce them. It is important to note, however, that when time rates are fixed in practice the employer (and possibly the worker) usually has in mind a certain output to be received for the wage; conversely, in fixing piece rates some account is usually taken of the time required to produce each unit of output. In other words, the principles underlying the two methods of payment are by no means mutually exclusive.

The relative advantages of these two types of wage payment are open to some dispute. Clearly, piece rates are applicable only where the nature of the work is such that the output of an individual worker can be unambiguously measured, and where differences in out-

put arising from causes outside the worker's control—variations in the quality of the raw material, in the state of the semi-finished product when it reaches the worker, or the machinery in use—can be reduced to a minimum, or where some agreed allowance can be made for them. In such circumstances the employer has a clear incentive to adopt piece rates; he is paying only for work actually done, and by stimulating high production he ensures that his capital equipment is being used intensively, so that his overhead costs are spread over the largest possible number of units of output.

The diligent and efficient worker also stands to gain by a properly administered piece-rate system. He is conscious of being paid for the work he performs, no more and no less. At the same time, although such systems frequently operate to the eminent satisfaction of both employee and worker (they are especially prevalent in the United States) there is sometimes hostility towards them on the part of Trade Unions. Partly from experience, and partly from instinct, the Unions sometimes feared in the past that piece rates were liable to have ill-effects on the workers themselves, believing that the system tended to induce undue physical and mental strain and injure the older and slower (though not necessarily the less efficient) worker. It is doubtful if considerations of this kind are regarded by Unions nowadays as of any great significance. Secondly, systems of piece rates often give rise to complicated problems of output measurement and rate-setting which frequently do great damage to management-labour relations. Finally, piece-rate payment has in the past often been exploited by employers in various ways; the rate per unit, for example, may be initially established by reference to an unusually rapid worker, so that the average worker is forced to operate at an unduly high pressure in order to maintain his usual level of earnings.

Time rates, on the other hand, have the disadvantage that unless expensive supervision is maintained the employer may find himself paying men for time which they have spent in idleness. The efficient worker, also, may suffer in that he receives no more for his work than a lazy and incompetent colleague. Nevertheless, many employers prefer time rates because the worker is not attempting to speed up output at the expense of the quality of the work or the life of the machinery he is using. Where the work is of a highly

skilled nature, or where expensive machinery or materials are involved, these may be important considerations. Again, where the nature of the work to be done varies considerably from one unit to another (as happens with repairs, or production to individual specifications, for example), the payment of piece rates is obviously impracticable. There is no single answer to the question of which system of payment is superior, therefore. Piece rates favour high-speed production of a repetitive nature; time rates encourage careful work and individual attention. Which is more applicable depends on the nature of the work, the employer and the labour force.

6. REAL WAGES AND MONEY WAGES:

People do not offer their labour to employers in order to obtain money as such, but in order to obtain the goods and services of all kinds which can be bought with the money wages they receive. There is an important distinction therefore, between money wages and *real* wages—meaning by the latter the purchasing power of a money wage in terms of goods and services. If a man's money wages rise 5 per cent at a time when the price-level of the goods he consumes has risen by 10 per cent, his real wages have fallen even though his earnings have risen in money terms. In such circumstances he would quite rightly consider himself worse off than he was before, and will probably press for a further increase in his money wages so as to restore their previous real value.

This is a factor of tremendous importance in the determination of wages at the present time. A rise in the general price-level will stimulate wage-earners to press for higher money wages to offset it; the costs of production in the industries concerned will naturally rise if the wage increase is granted, and this will tend to cause a further rise of prices and renewed pressure for wage increases. This subject of the 'inflationary spiral' will be examined in more detail at a later stage; for the moment the important point to note is that, if they act rationally, workers will require a certain real wage in return for their labour, rather than aim at some fixed money wage.

The importance of this distinction between real wages and money wages has led many Trade Unions to seek a system of wage settlement which embodies some mechanism for automatically adjusting

wages to changes in prices—either by linking wages to the price of the finished product concerned (a system widely used in the British coal industry in the past) or, more generally, to the index of retail prices discussed in Chapter 28 below. Such arrangements are usually known as *sliding scales*. They have the obvious advantage, so far as the worker is concerned, of protecting him from the erosion of the real value of his wages through price rises, and rendering unnecessary a resumption of wage negotiations every time such price changes occur.

From the point of view of the economy as a whole, however, the advantages of sliding scale arrangements are more dubious. Insofar as they reduce the extent of wage disagreements between employers and workers, and the frictions which often arise from them, they are certainly advantageous. On the other hand, they involve certain serious dangers. First, the automatic increase in wages as prices rise naturally involves an increase in costs, which will usually result in a further rise of prices and a further increase of wages. Without anticipating the discussion of the forces underlying the general level of incomes and prices which is left until Part VII, it can easily be seen that a sequence of this kind could rapidly degenerate into a cumulative process which might well get out of hand altogether. A second, and related, danger is that an automatic increase in some wage-rates as the result of price increases will inevitably stimulate demands for corresponding increases on the parts of other workers who are seeking to maintain a position of equality with other trades or to maintain the traditional wage-differentials mentioned earlier. Once again this could degenerate into a spiral of wages chasing prices and prices chasing wages, in which none of the wage-earners achieved any real increase in their incomes for more than a brief period and society as a whole suffered from the ill-effects of price instability. Finally, sliding-scale systems seldom work easily both ways; wages rise with rising prices easily enough, but Unions seldom accept wage-cuts without resistance, whether wage-rates are tied to prices or not. Hence there is a danger of the development of what has been well called a 'ratchet economy', with costs rising at the slightest provocation but seldom or never coming down. The dangers of such a situation will be appreciated more clearly after the discussion of foreign trade problems in Part VIII.

7. THE ROLE OF TRADE UNIONS

So far the determination of wages has been discussed solely in terms of the forces which would operate in a competitive world. In real life, however, the forces of competition are not allowed to function with complete freedom (although in the long run they are bound to determine the broad trend of wages), and a major factor in this context is the power and influence of Trade Unions. The basic aim of a Trade Union is that of strengthening the bargaining power of workers in relation to that of employers, so as to ensure that the worker obtains his fair share of the product to which he contributes and that he is given decent conditions of work. Without some kind of association with his fellow-workers, the individual employee is almost invariably in a weak bargaining position in comparison with the employer. To the average entrepreneur, the question of whether he employs one more worker or not is seldom of any great importance; to the worker concerned, however, it is frequently a matter of infinite importance, since (unlike his employer) he seldom has reserves on which he can live in idleness rather than accept what he considers to be an unfair bargain.

Theoretically, the forces of competition (as the analysis has shown) will ensure that labour is paid a wage equal to the value of its marginal product; the question of unfair bargains does not appear to arise. In practice there is no guarantee that labour will receive the full value of its marginal product. First, it is seldom possible to calculate what the value of labour's marginal product actually is, except within fairly wide limits; when output rises with an expanded labour force it is generally extremely difficult to say how much of the extra output is attributable to labour, how much to the capital used, or how much to the entrepreneur's own ability. Secondly, it has already been observed that the flow of labour from industry to industry is restricted by a great many factors, so that great inequalities of marginal product and earnings may persist. (In any case, if *all* employers paid their workers less than their marginal product, even complete mobility would be insufficient to rectify the situation.) The precise level of wages in any industry, therefore, is a matter for bargaining between employer and worker, and the fundamental aim of Trade Unionism is to ensure that workers bar-

gain collectively and on terms of at least equal strength with the employer. In many cases it has succeeded so well in this task that the bargaining strength of a large Trade Union is considerably greater than that of an individual employer, so that employers also resort to conducting wage negotiations through a collective organisation.

How can a Union succeed in obtaining higher wages for its members than could be secured by an individual worker? Essentially it does this through an actual or threatened restriction of the supply of labour—by shifting the supply curve to the left. Occasionally it may do this by imposing limitations on the numbers entering the industry (laying down apprenticeship requirements, or restricting new entrants to sons or daughters of its own members) or by restricting the amount of labour supplied by workers in the industry (by limiting their hours of work, or their rate of physical output). The power of the Trade Union lies fundamentally, however, in its ability to reduce the supply of labour to zero (so far as its own members are concerned, at least) by calling a strike. In all these measures, of course, the extent of the Union's power will depend on the proportion of an industry's labour force which is included in its membership, on the degree of determination and discipline of its members, on the extent of its financial reserves and on the ability of its leaders.

However powerful a Union may be, there are nevertheless definite limits to the extent to which it can increase the wages of its members. In the first place, although the precise value of labour's marginal product may not be known, a point will be reached when it will be quite clear to an employer that at the current wage it is simply not worth his while to employ labour if wages rise further. If the Union persists in pursuing wage demands past that point, unemployment is bound to occur and the Union will be forced to moderate its demands. (At the same time, the possible stimulus to productivity of high wages, noted at an earlier stage, should be borne in mind in this context.)

A second, and closely related, factor is the *elasticity of substitution* existing in the industry—i.e. the ease with which other factors can be substituted for labour in the production of the commodity concerned. As wages increase, the entrepreneur may find it has be-

come profitable to employ less labour and more land or capital, because the latter factors have become relatively cheaper. Much will depend, in turn, on the elasticity of supply of these alternative factors to the industry. If the capital employed is of a highly specialised nature, for example, it may not be possible to obtain more of it without considerable delay or without causing a substantial rise in its price. The greater the elasticity of substitution in the industry, of course, the more limited are the powers of the Union to secure higher wages.

Finally, much will depend on the elasticity of demand for the product in question. If the demand which an entrepreneur faces is highly inelastic, he will be able to pass on the higher wages to the consuming public in the form of higher prices without seriously damaging his sales; his resistance to Union pressure will therefore be relatively slight. (The same applies when the demand curve is steadily shifting to the right, as it will be when the general level of incomes is rising.) Conversely, an entrepreneur facing a highly elastic demand knows that an increase in price will seriously reduce his sales, so that higher costs can only be sustained at the expense of correspondingly reduced profits; hence he will not easily accede to wage demands. However strong a Union may be, therefore, a point may be reached after which higher wages will be secured only at the price of significant unemployment amongst its members; no responsible Union leader would willingly incur the risk of that.

8. STATE INTERVENTION IN WAGE-DETERMINATION IN BRITAIN

The general policy of modern British governments has been that the process of wage bargaining is best left to the individual Unions and employers concerned, and legislation has been restricted to defining the powers and legal position of Trade Unions. Two main types of exception to this broad statement have to be made, however, because governments have intervened in wage determination in particular industries where (*a*) Trade Unions have been unable to develop and attain reasonable strength, or (*b*) a dispute has proved insoluble through the existing conciliation machinery.

Under the first heading, government action in Britain dates back to the *Trade Boards Act* of 1909. This Act empowered the Board of

Trade to create machinery in four 'sweated' trades to decide upon and maintain a minimum wage. In these trades a great deal of female and juvenile labour was employed at extremely low wages and frequently under appalling conditions; partly because of the character of the labour force, and partly because of the small-scale and scattered nature of the trades concerned, virtually no Trade Union organisation had developed which could correct these features. These trades constituted classic examples of the vicious 'low productivity—hence low wages—hence low productivity' spiral mentioned previously.

The effect of the institution of minimum wages was not primarily to cause a reduction in employment, but to increase the efficiency of labour; this resulted partly from the improved diet and living conditions made possible by the higher wages, and partly from the fact that employers were forced to use the more expensive labour supply more effectively. The number of industries covered by the Trade Board system was gradually increased, and the Trade Boards Act of 1918 empowered the Minister of Labour to create Boards without requiring Parliamentary approval for any industry where, in his opinion, inadequate machinery for wage negotiations existed. Further extensions of the Trade Board system thereupon occurred, especially in other industries employing a large proportion of juvenile and female labour. In 1945 the Wages Councils Act continued, and somewhat extended, the provision of the 1909 and 1918 Acts, the Boards being given the new title of *Wages Councils*. Together with Wages Councils created by other legislation for agriculture and the catering industry—again industries in which physical conditions make it difficult for a strong Trade Union to develop—these statutory wages councils now establish minimum wages for about one-third of the employed population of the United Kingdom.

Under the second heading, the government in Britain has a long tradition of providing means for the settlement of industrial disputes which still remain after the normal processes of negotiation have been exhausted. If both parties to a dispute so desire, the Minister of Labour is empowered to appoint independent arbitrators or arbitration tribunals; these hear both sides of the case and announce a final decision. Although there is at present no com-

pulsion on either party to accept this decision, few Unions or employers would persist in refusing to accept it. Apart from positive arbitration, the Minister of Labour provides facilities for carrying out further negotiation and discussion between the disputing parties which must be tried before a strike or lock-out can be declared.

The powers of the government, then, are still confined to fixing minimum wages, or intervening in disputes, in particular industries only. There is no national wages policy in Britain; the overall level of wages is still the outcome of a large number of individual and unrelated wage agreements in various industries. In recent years there has been an increasing realisation that, given full employment and the enormous Trade Union power which that implies, there is a need for some machinery whereby the total increase in wages as a whole in any given year should be agreed, so as to avoid the wage-price spiral mentioned earlier. So far this has not progressed further in Britain than the stage of a mere interchange of views and aspirations between the government and the leaders of the Trade Union movement.

FURTHER READING

MARSHALL, *Principles of Economics*, Book VI, Chs. III–V.

M. DOBB, *Wages* (Cambridge University Press, London, 1948), Chs. II–IV, VII–VIII.

P. SARGANT FLORENCE, *Labour*, Chs. VII–VIII.

J. R. HICKS, *The Theory of Wages* (Macmillan, London, 1932), Chs. I, V & VII.

B. WOOTTON, *Social Foundations of Wage Policy* (Allen & Unwin, London, 1955).

CHAPTER 17

Rent

1. RENT AS A PAYMENT FOR LAND

In its usual sense, the word 'rent' means the payment made by a person for the privilege of using for a period of time an asset belonging to someone else. A person pays rent for the use of someone else's house or farm; equally, people 'rent' a car or cooker or television set. In economics the word originally related only to the income received from the ownership of land; in recent years the word has also come to have a special technical meaning in economics, and this will be discussed in the next section.

Considering the first, and simpler, definition of rent as the income received from the ownership of land, it is obvious that it is more limited in extent than when the word is used conversationally. When people speak of the rent of a house they certainly refer to a payment for the use of the land on which the house is built; the greater part of the payment, however, represents a fee for the use of the resources locked up in the building itself—i.e. a payment for the use of *capital.* Similarly, the rent of a television set has nothing to do with land; it is in fact a payment for the use of the capital which the owner has invested in the set. Even in the case of a farm, the rent that a tenant farmer pays to his landlord is only partly a payment for the 'natural and indestructible powers' of the soil; part of it is a return on the capital embodied in the farm in such things as farm buildings, fences and the drainage system. In economic analysis care must be taken to use the word rent *only* for the income received from the ownership of land, even though in practice it is frequently impossible to determine how much of a payment usually called rent is in fact rent in this technical sense of the word and how much is interest on capital.

How is the rent of land determined? The familiar demand-and-

supply analysis can answer this question exactly as it did in the problem of wages. Considering the supply side first, it is obvious that the concept of a supply curve is hardly appropriate for land as a whole, as opposed to particular types of land. A supply curve connects the different amounts of a commodity or factor forthcoming and the prices which have to be offered to obtain those amounts. But the total supply of land does *not* vary, except within extremely narrow limits, and another feature of land, of course, is that it has *no* cost of production. (The implications of these peculiarities of the overall supply of land are examined at some length in the next section.) If a supply curve is to be drawn up, then, it can only refer to the amounts of land used in particular activities—such as wheat-growing or sheep-rearing, or building. If the rent offered for, say, pasture-land is increased, landlords will withdraw their land from arable farmers and lease it to sheep-rearers. Conversely, if only a low price is offered for pasture-land, landlords will prefer to lease their land to builders or wheat-growers who can offer a higher price for it. A supply curve for land in specified uses is therefore a perfectly valid and helpful concept.

On the demand side no special difficulty arises. Each unit of land will add a certain sum to the total sales receipts of the entrepreneur employing it; this sum is the value of its marginal productivity. The amount which an entrepreneur will be willing to pay for any given unit of land will be determined by its marginal productivity, so that the demand curve for land will be a productivity curve similar to the labour demand curve. Like the latter, the demand curve for land will be downward-sloping, and for two reasons—diminishing returns (i.e. falling *physical* marginal productivity) and the smaller *value* of any given physical product as total supply increases and market price is reduced.

By constructing demand and supply curves on the same axes, the determination of the rent paid for land in any particular use can be analysed in the familiar way. The level of rent may be affected by a shift in the demand curve or in the supply curve. The former could occur as the result of, for example, a change in market demand for the commodity concerned, while the latter might be caused by a change in the earnings of land in other uses, which

would naturally affect the amount of land offered to the given industry in response to any specified rent.

2. RENT AS A SURPLUS

So far the determination of rent has been treated, like that of wages, as a special case of the general theory of competitive price developed in Part IV. If it is attempted to extend the argument to cover the determination of the rent of land as a whole (as opposed to land employed in a particular use) a serious difficulty is encountered. In the case of factors other than land a supply curve for the factor as a whole can be conceived of, a curve which embodies the relationship between the average price offered for the factor and the total quantity of it supplied. If the general level of wages falls so low that it is below that necessary to maintain a bare level of subsistence, the total amount of labour available will fall as the population is reduced through starvation; if the rate of interest falls, the amount of capital offered will diminish, and so on. All the factors except land have a cost of production, in some sense, and the supply will contract if the price of the factor falls below this cost. The cost of production of labour, or enterprise, is the effort and training required if a day-old baby is to be developed into a potential worker; the cost of production of capital is the sacrifice of the immediate command over wealth which must be incurred if it is to be used to bring in an income over the future. But land has no cost of production; it is a free gift of nature which no-one has ever sacrificed anything to produce. Further, as was noted in Chapter 7, the total supply of land is fixed within narrow limits; no matter how high or how low its price may be, its quantity is unchanged, precisely because there is no *necessary supply price* which must be paid in order to bring forth a supply of land.

Because of these unique characteristics it follows that for all land taken together rent is a *surplus*, in the sense that it corresponds to no effort or sacrifice incurred in supply and requiring a reward. If wages or profits or interest fall to zero the supply of the corresponding factors would dry up, so that these payments obviously perform a necessary function. But rent has no such function; it is entirely a surplus corresponding to, and compensating, no cost of production. The realisation of this fact has led to the development

of the concept of *economic rent*; that is, the word 'rent' in this special economic sense is used to describe a payment over and above the necessary supply price of a factor.

If rent is defined in this way, it will be apparent that rent can be received by any factor, not only land. For the total factor supply, only the payment for land is wholly economic rent and nothing else, because land is the only factor having no cost of production at all. But in the case of factors supplied to any particular industry, any or all of them may receive some rent. It was pointed out in Chapter 10 that the *real* cost of a commodity is the sacrifice of the alternative uses of the factors used up in its production. A parallel concept can be applied to factors of production themselves. Before an entrepreneur can obtain the use of any factor he must offer it at least as much as it could earn in the most suitable alternative employment open to it. This sum of money, its 'best earnings elsewhere', is the necessary supply price of that factor to the industry; it is usually referred to as its *transfer cost*. The reward paid to a factor must be at least equal to its transfer cost, otherwise (by definition) it will move into some other occupation; any amount paid to a factor in excess of this transfer cost is a surplus—i.e. it is economic rent. The surplus, or rent, is the difference between the actual earnings of a factor and its necessary supply price, or transfer cost.

In any particular industry, then, economic rent may be received by labour, enterprise or capital as well as by land. For example, suppose that a successful barrister earns £2,500 a year, and that if he ceased to practise law the most lucrative alternative employment he could obtain would be that of, say, a civil servant at £1,500 a year. So long as he earns more than £1,500 at the Bar, therefore, he will be content to remain a barrister; he will abandon the profession only if his earnings fall below £1,500. His transfer cost, in other words, is £1,500. But everything he earns in excess of this is rent, since he would remain in that occupation even without this extra reward. Similarly, an entrepreneur in the washing-machine industry, Mr. X, may earn £5,000 a year because he has a special flair for washing-machines; if he moved into any other industry he might earn no more than, say, £800 a year. The difference between £800 and £5,000 a year is economic rent, some-

times referred to, in this context, as a *rent of ability*. This expression is used because earnings far in excess of transfer cost can arise only if the factors concerned possess some special qualities which are not reproducible. For example, if Mr. X had no special ability setting him apart from all other entrepreneurs, two possibilities would immediately arise. Either *all* entrepreneurs would be earning £5,000 a year in other industries, or they would move into the washing-machine industry, since, being as able as X, they would want to share in the high profits X was making. In the former case X's transfer cost would be £5,000, since he presumably could also earn this sum elsewhere, being of the same ability as other entrepreneurs; in the latter case, X's earnings would be driven down by the competition of the entrepreneurs attracted into the industry. In either event the gap between actual earnings and transfer cost would be destroyed, and the economic rent would disappear.

It is difficult to conceive of economic rent arising from the ownership of capital, because (unlike innate human ability) any special features of capital assets resulting in unusually high earnings can in general be reproduced in other capital assets. If a sudden increase in the demand for houses to rent results in a high yield on the capital invested in such houses, other houses (in normal times) would be built with capital diverted from other uses; this switch of capital would continue until *either* the high rents disappeared *or* the interest paid to capital in other uses rose into equality with the level prevailing in the house-renting industry. However, it is possible to conceive of capital receiving rent in exceptional cases. For example, the revenue derived over a period of years from exhibiting the original Stephenson 'Rocket' to the general public would certainly be far in excess of either its original cost of production or the cost of building an exact replica. This excess would be in the nature of economic rent.

Economic rent arises, then, when a factor receives a reward in excess of its necessary supply price, or transfer cost. It can be enjoyed only by a factor which possesses some quality which cannot be increased in supply in response to its high earnings; land enjoys this distinction of fixity of supply in respect of all its innate qualities, but other factors may also possess special attributes of economic value and may therefore receive economic rent. The

concept has important implications for taxation policy. If wages, profits or interest are taxed beyond a certain degree it is possible that serious effects may be felt on the supply of factors and hence on production: workers may be unwilling to work long hours, entrepreneurs may be less inclined to assume risks for the sake of large profits, investors may be reluctant to lend money. But the taxation of economic rent, by definition, could have no influence on incentives or effort, since it is an unnecessary payment so far as factor-supply is concerned. Unfortunately it is extremely difficult to distinguish in practice between economic rent and other forms of income.

3. QUASI-RENT

Considering the elasticity of supply of factors of production, there are at one extreme factors which are completely fixed in supply, and at the other extreme factors, such as ordinary unskilled labour, which are in fairly elastic supply so far as any individual industry is concerned. Between these extremes can be found factors the supply of which is fixed in the short run but which is responsive to demand changes after the lapse of time. In the short run such factors may receive rewards considerably in excess of their supply price, and if this happens they can be said to earn economic rent. The difference between them and the resources discussed in the preceding section—land and other factors possessing some peculiar, innate qualities of economic value—is that the rent they earn is essentially temporary; the passage of time is bound to remove the causes leading to their excess earnings. For this reason the surplus payment received by these factors is known as *quasi-rent.*

An example will indicate how and why quasi-rent can arise. When commercial television was first introduced, the demand for people trained in the business of producing and transmitting television programmes was greatly increased. As there was in existence only a limited number of people with the requisite qualifications, the result was a substantial increase in the wages paid for this specialised form of labour, and the people concerned found themselves earning considerably more than they could have obtained in any alternative occupation open to them—i.e. their earnings

exceeded their transfer cost. After the lapse of a time sufficient to train the new entrants attracted by these high earnings, however, the supply of labour to the industry will increase and the earnings of that labour will return to a more normal level. The surplus earnings enjoyed by the existing labour force between the initial rise in demand and the full adjustment of the labour supply to the new conditions can be called quasi-rent.

Similarly, an increase in the demand for a product will generally cause a rise in its price, and hence in the profits of enterprises producing it. If the capital equipment involved in its production is to some degree specialised, as it frequently is, it will be some time before extra equipment can be manufactured and installed in order to meet the increased demand, and until the capacity of the industry has been fully adjusted the owners of existing enterprises will be earning unusually high profits. The excess of these profits over the necessary supply price of the equipment concerned is a quasi-rent.

Like pure economic rent, therefore, quasi-rent is a surplus over and above the necessary supply price of a factor which arises out of some special quality or ability of the factor concerned. Unlike pure economic rent, quasi-rent is essentially temporary; the qualities or abilities which give rise to it *can* be increased in supply, but only after the lapse of a period of time during which quasi-rent is earned.

From the concept of a payment for the use of land or other natural agents, then, rent has developed an extended significance in economics. The word is now used *either* (in its original sense) to mean the income derived from the ownership of land, *or* to mean any income paid to a factor which exceeds its necessary supply price, or transfer cost. In the latter sense, the modified form of 'quasi-rent' is employed, in reference to rents of a temporary nature arising only from a time-lag in the supply of factors to changed conditions of demand. It is most important, of course, that the sense in which the word 'rent' is being used should always be made clear.

FURTHER READING

MARSHALL, *Principles of Economics*, Book VI, Ch. IX and Appx. K.
* ROBINSON, *Economics of Imperfect Competition*, Ch. 8.

CHAPTER 18

Profits

1. PROFITS AS AN ACCOUNTING SURPLUS

Like the word 'rent', the expression 'profits' is one which is used in ordinary conversation a good deal more widely and loosely than in economic analysis. Many people would say that profit is the difference between the price at which a commodity is sold and the cost of the labour and raw materials used up in its production. A moment's consideration will show that the sum arrived at in this way would be considerably more than the rewards of the factor of enterprise, which is the meaning of the word profit in economics. It would include a certain amount which entrepreneurs would need to set aside for the *depreciation* of the capital employed; as was noted in Chapter 8, a certain amount of capital consumption occurs in almost all production, and a deduction must be made from gross receipts merely to maintain capital intact.

A second deduction which has to be made is that of *interest* on the capital employed. Apart from the need to maintain capital intact, a payment must be made to the lenders of the funds with which the capital was purchased to compensate them for sacrificing the control of their wealth over the period during which it is used in production. It may be that an entrepreneur has employed his own capital in the business; if this is so, part of the apparent profit received by him is in fact interest on his capital, even though a separate deduction under this heading is not actually made.

Thirdly, some of the accounting profit may really be wages for the labour performed in management. In a public joint-stock company, managers are usually paid officials whose salaries are shown as an element of cost in exactly the same way as the wages of operatives. If an entrepreneur carries out management duties himself,

however, a separate payment may not be made to him on this account; nevertheless, an element in gross profits in the accounting sense will need to be deducted as wages of management. A farmer's net income, for example, may constitute not only profits but also sums received by him as wages in respect of the labour he (and probably his wife) has supplied during the period. If he owns his own farm, another part of his net income may represent the rent of the land which he owns and, so to speak, puts at his own disposal for the purpose of agricultural production.

True profit, then, is the sum remaining from the value of output after payments (actual or imputed) have been made in respect of the land, labour and capital employed in production—including any of these factors supplied by the entrepreneur himself—and after a deduction has been made for capital depreciation. It is this which is the reward paid to the entrepreneur as such; it is completely distinct from any reward he may obtain in respect of functions he may discharge as a supplier of labour, land or capital.

2. PROFITS AS A COST

Defined in this way, profits could still be regarded as a surplus, since it is the sum left over from the sale of output after the factors of land, labour and capital have received their share of the product. It is a residual in the sense that the rewards of the other factors are determined in advance, whereas profits fluctuate according to the proceeds of current production. Unlike the other factor-prices, it may even be negative; the entrepreneur may find that at the end of an accounting period his economic activity has actually left him poorer than when he began.

Yet profit is still a cost of production from the economist's point of view. It is a cost because it is necessary to call forth a supply of the factor of enterprise, which is as necessary to the production of wealth as any other factor. To be exact, an entrepreneur is induced to undertake production because of the profit he *expects* to receive; the profit he actually receives cannot influence his decision since it emerges only after the event. In practice, however, the expectations of profit which entrepreneurs hold are bound to be strongly influenced by the profits they have actually received in the past from similar production efforts. Hence, profit can be treated as the

cost of enterprise, even though it is the *expectation* of profit which is the operative element.

True profit is a cost of production because a price has to be paid to ensure a supply of entrepreneurs. As was noted in the previous chapter, profit, like all other factor incomes, may include an element of rent or quasi-rent; nevertheless, there will always be a minimum price necessary if the factor of enterprise is to be supplied. If profits fall to an unduly low level, men who would otherwise have fulfilled the functions of the entrepreneur will instead seek employment as workers, managerial or otherwise, or be content to live on any income they can obtain from the ownership of land or capital. But profit fluctuates much more than the payment to any other factor, both from one industry to another and from one time to another; since it can be negative in some cases there may be no limit to its fluctuation in either direction. Why is this?

3. THE BEARING OF UNCERTAINTY

In Chapter 9 it was stated that the special role of the entrepreneur in an economy was that of undertaking the responsibility for bringing the other factors together to form a productive unit, and of bearing the uncertainty inevitable in such a process in a changing and unstable world. The supplier of labour, or land, or capital, at least knows in advance the reward which will be obtained during the term of the contract under which these factors are engaged. This is not to say that the owners of these factors run no risk whatever. The owner of capital may discover that a borrower is a fraud who proceeds to dishonour his contract to repay the capital; the landowner may be ruined by flood or earthquake; the worker may lose his job and suffer prolonged unemployment. But these risks, of dishonesty, natural calamity or economic depression (like those of war, or personal violence, or disease), are risks which are shared to some degree or other by all members of society. The entrepreneur, whether a single person or a group of ten thousand shareholders is involved, cannot rely on some guaranteed level of earnings; he has to sustain the special uncertainties of changes in fashion, or technological development, or misjudgement of the market—all the thousand and one factors which may falsify expec-

tations regarding the volume of sales of a product, the price it will obtain, or the costs it will involve.

In a free and dynamic economy these factors are bound to operate. Demand is certain to jump unexpectedly from one product to another, unless consumers are forcibly deprived of their right to exercise a free choice; new inventions will render machinery and methods obsolete and relatively wasteful, unless technological advance is somehow brought to a standstill; errors of judgement will surely occur unless men are suddenly endowed with a capacity for predicting the future accurately. Production in anticipation of demand must always involve uncertainties, and if production is to be carried on at all someone must be willing to shoulder them. Someone must undertake the task of anticipating the way in which demand, innovation and consumer taste are going to change in the future and take the chance involved in developing a new product or process. Not only must a 'best guess' be made in these matters; it has also to be backed with resources which may be lost if the guess turns out to be a bad one. The suppliers of labour, land and capital have to be paid their reward whatever the outcome; if the price eventually obtained for a product is insufficient to meet these rewards, then the entrepreneur must make up the deficiency from his own resources. The placing of one's own wealth in peril of loss is not a prospect which would appeal to many people unless there was some compensating hope of gain. Profit is the reward which induces people to assume the special and vital role of the entrepreneur.

It is the existence of uncertainty, then, which lies at the root of profit. If the world relapsed into a static state in which neither demand nor costs of production could conceivably change, uncertainty would vanish from the economic scene, and with it would vanish the necessity for profit. In the real world, of course, uncertainty is always present. Certain undesirable eventualities (losses through fire, or theft, or tempest, for example) occur in society as a whole with such regularity that insurance companies are prepared, for an appropriate fee, to relieve individuals of the possibility of loss implicit in them; to the extent that these eventualities can be insured against, of course, the entrepreneur can treat the insurance premiums as a cost of production, like any other cost, and profit

will not be required to induce entrepreneurs to face such risks. But there will always remain the possibility of eventualities—fluctuations in demand or technological developments, for example—which do not occur with sufficient regularity to enable insurance companies to indemnify entrepreneurs against them. And the insurance companies themselves will always be bearing the residual uncertainty, so to speak, involved in the contingencies for which they *do* provide cover; their calculations of the probability of death or fire may be completely upset by a plague or wave of arson, so that they in turn will require profit to compensate *them* for taking this chance.

It follows that the rate of profit necessary to induce an adequate supply of entrepreneurship—the 'normal profit' included as an element in the cost curve—will vary from industry to industry according as the degree of uncertainty varies. Some commodities (most foodstuffs, for example) have a stable demand curve over long periods of time, while the possibility of sudden and drastic changes in the method of their production is not especially great. In such industries a relatively low rate of profit could be expected to prevail. By contrast, a brand-new product has an unknown future; it may be a resounding success but it may also be a dismal failure. A particularly high rate of expected profit may be necessary to induce people to risk resources in such a venture. Wherever the uncertainties attached to producing goods are especially great, profits will have to be high if entrepreneurs are to come forward, whether the uncertainty arises from the nature of the product, the method of production, the time-lag between committing resources and recovering them through the sale of output, or from the prevailing environment.

4. PROFITS AND THE ALLOCATION OF RESOURCES

The way in which the price mechanism allocates resources to different uses in a competitive economy was described in Chapter 15. The rather special role of profits in this process justifies particular stress, however. After all, it is the entrepreneur who takes the initiative in a competitive society. It is he who has to decide what shall be produced and in what quantity, how much labour, land and capital shall be used and in what proportions.

Since profit is the element which determines the action of entrepreneurs, it follows that it has a vitally important function in the allocation of resources.

It is through changes in the rate of profit that the competitive impulse is first felt. If the output of a certain commodity is greater than the consumer thinks desirable (i.e. the utility derived from it is low in relation to the scarce resources absorbed in its production), the price at which it will sell in the market will fall. Since the entrepreneur has to pay the agreed rewards to land, labour and capital whatever the price at which the product sells, profits will inevitably be small; if price falls far enough, profits may be replaced by losses. In this way the entrepreneur is warned that he is not employing resources in a manner which will render the maximum possible satisfaction to the community, and he will be forced to use the resources at his command in a different way or go out of production altogether.

Conversely, an entrepreneur who uses resources so as to cater for a novel and relatively strong want felt by the community will obtain a high price for his product, and will therefore enjoy high profits. This will encourage him to increase his production and add even further to his profits; entrepreneurs will also be encouraged to move resources from less profitable industries into this high-profit industry. Through fluctuations in the level of profits, then, the community expresses its judgement that more resources should be used in some directions and fewer resources in others, and the entrepreneurs concerned are stimulated or discouraged accordingly.

Profit not only acts as a guide to the industries in which entrepreneurs should expand or contract output. It also provides a means whereby the more efficient entrepreneurs are enabled to draw resources away from the less efficient. At any given market price, a highly successful entrepreneur, who will have, by definition, relatively low costs of production, will be enjoying a high rate of profit. He will therefore have an incentive to obtain more factors of production in order to expand output, and will be prepared to offer slightly higher wages, rent and interest so as to obtain them. The inefficient entrepreneur, by contrast, will have been making only a minimum rate of profit, and even this will disappear as factor-prices are bid up by his competitors; eventually his profit

disappears completely and he is forced out of business, setting free the resources he was previously using for employment by more efficient producers.

In this way, then, profit does more than merely ensure that an adequate supply of the factor of enterprise is forthcoming. In a perfectly competitive world it provides to entrepreneurs a compelling indicator as to the kind and quantity of commodities which the community considers they should produce with the available factors of production. Bearing in mind the basic assumption that consumers set out to distribute their expenditure so as to maximise the satisfaction they obtain, it follows that by means of relative profit-levels resources are distributed amongst different industries in such a manner as to maximise the satisfaction which a society derives from the scarce resources at its disposal. Furthermore, the function of profit is to weed out the inefficient entrepreneurs and ensure that production is carried on under the direction of the entrepreneurs whose relative efficiency is highest.

All this, however, presupposes the existence of a perfectly competitive world, and, as was established at an earlier stage, the real world is far from perfectly competitive. The picture must be modified, therefore, to allow for the existence of monopoly.

5. MONOPOLY PROFITS

It was concluded earlier that under imperfect competition price would be higher than it would be if the commodity was produced under conditions of perfect competition. This was not because the producers were making relatively high profits, however; when equilibrium prevails in an imperfectly competitive industry, all entrepreneurs are earning only 'normal' (or average) profits. Price is unduly high under such conditions because there are too many firms producing at a rather low level of output, and at some point on the average cost curve which is above the possible minimum. Under imperfect competition, prices are high because costs are high; the evil of the situation is not especially high profit-margins but the waste of unused capacity. (It will be remembered that the question of whether the advantages of imperfect competition are sufficient to outweigh this evil is one of personal opinion rather than one of objective analysis.)

In monopoly situations, however, different considerations apply. Because of his controlling position in total production, a monopolist is able to hold price above average cost; price is high not only because production is carried on at a relatively high-cost level (which it will be, of course) but also because the producer is making exceptional profit out of his monopoly power.

How can such excessive profits be reconciled with the considerations determining the nature and functions of profit which have just been discussed? The answer can only be that the excess, or supra-normal, profits received by a monopolist are a form of economic rent, in its sense of a surplus over and above transfer cost. It has already been established that economic rent can be enjoyed by any factor having some quality which cannot be increased in supply. The monopolist obviously has such a quality—he has the ability, derived from natural or artificial factors, to restrict the supply of a commodity and hence to obtain a high price for it. Previously economic rent was conceived of as arising from innate and non-producible qualities, or, in the case of quasi-rent, from the physical impossibility of increasing the supply of a certain factor in the short run. In principle there is no difference between, on the one hand, rent arising from a scarcity due to the physical impossibility of increasing the supply of a rent-earning quality, and, on the other hand, rent arising from the natural or artificial limitations which constitute the foundation of monopoly power. If the monopoly position is a permanent one, the entrepreneur's profits in excess of the normal level are a form of pure economic rent; if the monopoly power is temporary (such as one arising from patent rights during a specific period of time) then the excess is a quasi-rent.

What are the implications of monopoly profit for the allocation of resources? Obviously it can no longer be concluded from the existence of a high level of profits in the actual world that the industry involved is one in which an unusually large element of uncertainty prevails, or that the entrepreneur concerned has exceptional ability. The high profits may be a reflection of some degree of monopoly power which permits an entrepreneur to reap an economic rent at the expense of the consumer. Nor can profits always be relied upon to stimulate entrepreneurs to make the

optimum use of society's scarce resources. A monopolist, as the analysis has shown, does not carry production to that point at which average costs are at a minimum. High profit will not lead to higher production because the monopolist himself is already operating at the output which gives him maximum profit, while the presence of monopoly power means that other entrepreneurs are not able to move into the industry in order to share in the high profits prevailing in it. If monopolies exist in the economy (as they do in the real world) then the profit mechanism may not lead to the optimum distribution of resources; output may be too low, profits may be unnecessarily high, and the consumer may be deprived of the full benefit of the wealth-producing powers of his society.

FURTHER READING

MARSHALL, *Principles of Economics*, Book VI, Chs. VII–VIII.

F. H. KNIGHT, *Risk, Uncertainty and Profit* (London School of Economics Reprints, No. 16, London, 1940), Chs. I–II, IX–X.

* B. S. KEIRSTEAD, *An Essay in the Theory of Profits and Income Distribution* (Blackwell, Oxford, 1953), Ch. II.

CHAPTER 19

The National Income

1. MEASURING THE NATIONAL INCOME

The journey through the elementary analysis of the functioning of the economic system is now well advanced. It has been seen that economic activity is aimed at the satisfaction of as many of society's infinite wants as possible from the limited resources at its disposal; that consumers distribute their expenditure so as to derive the greatest benefit from their income; that resources are combined so as to meet the demands of consumers; that the value of a product in relation to other products is determined by the interaction of demand and supply; and finally the analysis has shown how the value of every product is shared between wages, rent, interest (to which the discussion will have to return later), and profits.

If all this is brought together the concept of the *national income* is arrived at—that is, the total flow of wealth produced, distributed and consumed by the economy as a whole. Just as an individual can arrive at a figure for his total income by adding up everything he has received from work or property during a given period, so a nation as a whole can add up all the incomes of its citizens over a period and arrive at its *national* income.

How can this be done in fact? It is known that every penny of every price is someone's income, and that income can arise ultimately only from the production of something; hence it follows that the total income received by society as a whole during, say, a year, can be calculated *either* by adding up the value of all sales of finished goods and services, *or* by adding up the values of all types of production, *or* by adding up the incomes received by all individuals or organisations. These three things—total expenditure, total production and total income—are really the same thing looked

at from different angles. In practice, all three methods of measuring the national income are employed; since there are gaps of varying degrees of importance in the statistics available under each heading, the three methods check and supplement each other.

It may be helpful to indicate broadly the way in which the official statisticians set about the task of compiling their national income estimates, which in Britain are published every year. The first method is that of estimating total expenditure on finished products. The total amounts spent on *consumer goods* are estimated from statistics relating to the annual sales of wholesalers and retailers. A basic source of such data is the Census of Distribution which is carried out periodically; for this census, every shop makes a return showing (amongst other things) the total value of its sales. For the intervening years the statisticians rely on data supplied by a smaller number of shops which are regarded as a representative sample. To these estimates of purchases of goods has to be added the estimated expenditure on personal services of all kinds, such as entertainment, repairs, professional services and domestic help. Then the amount spent on *capital goods* has to be added. As such goods are purchased mainly by companies, the major sources of data here are company balance sheets, tax returns, and the Census of Production (briefly described below); from these can be calculated the amount spent on fixed assets (plant, buildings, vehicles etc.) and on stocks of raw materials and finished or semi-finished goods.

The calculation of total expenditure, however, is not entirely a straightforward affair even in principle (the statistical problems of estimation, of course, are very considerable). Certain pitfalls have to be avoided, and in the process of avoiding them the published national income estimates become somewhat complicated. It is obvious that in calculating a country's national income *via* estimates of total expenditure, only purchases which provide income for someone else in the community should be included, while care should be taken to see that no such purchases are excluded. This is a little trickier than might be supposed at first sight because of two complications. The first is *foreign trade*. An effect of the existence of such trade is that some of the purchases in a country do *not* provide income for someone else in the same

country, because they involve goods which have been imported. The income corresponding to my purchase in Britain of New Zealand butter is in New Zealand, where its producer lives, not in Britain itself. Conversely, the man in Birmingham who helps to make a car which is sold in America is paid wages which have to be included in an estimate of the British national income; there is no sale of goods *in Britain*, however, which corresponds to that income. To meet this difficulty, then, the British statisticians have to arrive at a total for all expenditure by persons and enterprises in Britain, *subtract* from it the value of goods imported into Britain during the year, and *add* to it the value of goods exported during the year.

The second complication involved in calculating a total for the national income from expenditure data arises from the existence of *indirect taxes*—(i.e. those levied on items of expenditure rather than on income or capital). If I pay 4s. 0d. for a packet of cigarettes, there will not be a corresponding income of 4s. 0d. somewhere else in the economy, since a good deal of the price paid disappears into the government's purse in the form of taxation receipts. Conversely, when my wife pays 4d. for a pint of milk for my small daughter the income received for that milk is considerably more than 4d. because the government pays part of the farmer's price for milk in the form of a *subsidy*, which is a kind of negative indirect tax. If the estimate of total expenditure is to be used to measure the incomes actually received as a result of current production, then, statisticians have to take the basic estimate of total expenditure, *subtract* the value of indirect taxes and *add* the value of subsidies. When they do this it is said that they are measuring the national income at *factor cost* (i.e. at what the various factors of production were actually paid for it); if they do not make this adjustment they are measuring at *market prices*.

The second method of measurement of the national income is the addition of the values of the goods and services produced by all the enterprises in the economy—i.e. the estimate of the value of total *production*. For almost all of industry, such estimates are based on the Census of Production; returns for this are made by all industrial establishments in the United Kingdom every five years and by a fairly large sample in every year. These returns pro-

vide a great deal of information about the productive activities of the enterprises; in particular, they state the value of the goods produced by them during the year concerned. For industries not covered by the Census of Production, estimates are based primarily on statistical material covering the amounts paid in wages, salaries, rent and interest in such industries, together with balance-sheet and tax data bearing on the profits made by the firms concerned. As has been established, the payments made to factors of production plus the residual sum going to profits (or minus any losses) must equal the values of the goods produced.

As in the expenditure method of calculating the national income, care has to be taken to avoid possible errors of principle in totalling the value of current production. Once again there are two major considerations. The first is that *double-counting* must be avoided—that is, the total must not include the value of both goods *and* the value of raw materials etc. used up in their production. For example, assume that a refrigerator sold for £100 embodies steel worth £20. If the statisticians added up the gross value of the production of steel and refrigerators, the £20 worth of steel would be counted twice—once as part of the output of the steel industry and again when it is included in the value of the output of refrigerators. To avoid this error, the output measured for each industry is not its gross output—i.e. the full sales-value of the goods produced—but its net value, or the 'value added' by each industry—i.e. the market value of the output *minus* any raw materials or semi-finished goods purchased from other producers.

The second point about production totals is similar to this. Just as the statisticians must avoid including in the value of output goods or materials already counted in the output of some other producer, so they should avoid including any *capital* consumed. It was noted in Chapter 8 that when capital depreciates, or wears out, in the course of production (as it almost invariably does), the extent of that depreciation has to be subtracted from the value of output to arrive at a true figure for current production; unless this is done, a certain amount of capital consumption is included in what purports to be a figure for current creation of wealth. In measuring the national *income*, then, the consumption of *capital* involved in its production during the period involved should be subtracted. As

was explained in Chapter 8, this is an extremely difficult task statistically, but an attempt is nevertheless made nowadays to measure capital consumption in the United Kingdom. If this deduction is not made from the value of production, the resulting national total is called the *gross national product*: only when a deduction *is* made for capital depreciation can the term national income be correctly applied to the estimate.

The third and final method by which the national income may be estimated is that of totalling all *incomes* received by persons and organisations in the country. Such a total can be derived from data collected for the purposes of income-tax assessment in respect of both persons and companies. To this must be added the incomes received by official agencies of one kind or another which carry on productive activities—public corporations, such as the B.B.C. and the British Overseas Airways Corporation, nationalised industries and trading branches of the central and local government.

Once again, however, a complication has to be dealt with. As was emphasised earlier, the national income represents the value of the wealth produced by a community over a certain period. The total of incomes should therefore include only those receipts which are directly related to the production of goods—i.e. only factor-rewards. In a modern economy, however, the state frequently attempts to make the distribution of incomes more even by taking income from the relatively rich in the form of taxation and handing it over to the relatively needy in the form of what are called *transfer payments*—for example, children's allowances, old-age pensions and national assistance grants. These payments enter into the incomes of a good many people in the community. But if the statisticians included them in the income total for national income purposes they would be falling into the old trap of double-counting. If a man earns £500 a year as a coal-miner and pays £100 a year in taxation, which is transferred to someone else as a pension, the total amount of wealth *produced* by both the miner and the pensioner is only £500, since the pensioner is producing nothing. The government has decided that of the £500-worth of wealth produced by the coal-miner, the miner should dispose of £400-worth and the pensioner £100-worth. But if the statisticians simply added both incomes together they would arrive at a total of £600,

which overstates the true value of current production; the transfer income has been included in both the original income from which it was taxed *and* the income of the person to whom it was transferred. To correct the income total, then, all transfer incomes must be subtracted; only incomes directly related to current production of goods and services must be counted in.

These are the three methods, then, by which an estimate of the total national income may be attempted, and some (by no means all) of the complications involved in them. In theory, the three approaches should lead to identical totals; if they do not it can only be as a result of imperfections and gaps in the statistical data. In fact, the art of compiling such estimates, in Britain at least, has now reached a sufficiently high degree of perfection to enable a reasonably accurate total to be arrived at each year. It will be instructive to consider briefly the estimates of national income and expenditure in the United Kingdom for the year 1955.

2. NATIONAL INCOME ACCOUNTS

Each year the official estimates of the national income and gross national product of the United Kingdom are published in a Blue Book called *National Income and Expenditure*. The Blue Book contains a large number of tables giving various items in considerable detail; the present discussion need only be concerned with the basic table; this will be sufficient to indicate the kind of information which can be obtained from the annual estimates.

The most comprehensive table is that of the gross national product and national income, since this brings every item together and gives a picture of the economy as a whole. For 1955 the estimates were those shown in Table XI; the preceding discussion will have explained the nature of most items in the table. Five features of the income side of the table call for brief comment. The first is the item 'income from self-employment', which represents the incomes of people such as farmers, shopkeepers and small-business proprietors who receive a 'mixed' income from both work and property. As it is never possible to separate out the various elements of their income—how much is wages, how much a reward for entrepreneurship etc.—only the total is shown. Secondly, the item 'interest' does not appear; the reason for this is

that the profits shown are those *before* payments of interest and dividends, so that to include a separate item 'interest' would involve double-counting. Thirdly, the word 'rent' is used in

TABLE XI. *Gross National Product of the United Kingdom 1955.** (*£ million*)

Expenditure	£	*Income*	£
Consumers' expenditure	12,783	Income from employment	11,150
Public authorities' current expenditure on goods and services	3,192	Income from self-employment	1,718
Gross capital formation	3,165	Profits	3,272
		Rent	731
		Unidentified errors	−32
Total domestic expenditure	19,140	Total domestic income	16,839
Plus exports	4,483	*Less* stock appreciation	−200
Less imports	−4,565	*Plus* income from abroad	145
Plus subsidies	346		
Less indirect taxes	−2,620		
Gross national product	16,784	National income and depreciation	16,784
Less capital consumption	−1,558	*Less* provision for capital depreciation	−1,558
NATIONAL INCOME	15,226	NATIONAL INCOME	15,226

* Based on *National Income and Expenditure 1956* (H.M.S.O. London, 1956), Tables 1 and 56.

national income accounting in the usual sense of income from real property. It is impossible, of course, to calculate receipts of pure economic rent or quasi-rent. Fourthly, there appears a mysterious item 'stock appreciation', which refers to the change in the value of stocks of goods arising not from a change in their quantity, or volume, but from movements in the prices of the goods involved. Any profit made from a rise in the value of a given stock of commodities will be included in the item 'profits' (just as will be sums required to make good the depreciation of capital), but since it corresponds to no increase in the current production of wealth, and to no item of expenditure, such value-changes have to be excluded from the total of current national income (just as capital depreciation must be). Finally, any income received from the ownership of property in other countries is part of a nation's disposable income and will have been used for its current or capital expenditure; hence it is added in by the item 'income from abroad'

From a table of this kind a great deal more than simply the total of the gross national product can be learnt. On the expenditure side is shown the broad pattern of the way in which the community is *disposing* of its income—how much is being spent on personal expenditure, how much the government is spending on goods and services and how much is being set aside for addition to the stock of capital goods. From the income side is seen the relative importance of the main factor-rewards; what proportion of the total product is paid to labour, what proportion to profits etc. The relative importance in year-to-year changes in total incomes of wages, rent and profits can be discovered, thus throwing a good deal of light on the underlying forces at work in a period of rising or falling incomes and demand.

Other tables published annually proceed to break down these aggregates into more detail. For this purpose the economy is divided into six sectors—persons, companies, public corporations, local government authorities, the central government, and a 'rest-of-the-world' sector showing transactions with other countries. Separate tables then set out the different sources of the incomes of each sector and the various ways in which those incomes are disposed of. The personal income and expenditure table, for example, shows the relative importance of wages, salaries, transfer payments and dividends in total personal income, and changes in these components; on the expenditure side it shows how much of this total is spent, how much paid in direct taxation and how much is saved.

A full appreciation of the wealth of information to be gleaned from a set of national income accounts can be obtained *only by looking carefully through an actual set of accounts*; there is no fuller or clearer guide to the characteristics and development of an economy. They will provide an answer to a great many questions. How rapidly is the national income expanding? How does the rate of increase compare with that of other countries? How much of the total income is being set aside to add to capital? How does the size of this share compare with previous years or other economies? How important a role does the government play in the consumption of total income? How important are undistributed profits and personal savings as sources of finance for investment? Are any significant changes occurring in the pattern of consumer expendi-

ture? The list is almost endless. A later section will use the national income accounts to throw some light on the important question of the distribution of the national income.

3. NATIONAL INCOME AND WELFARE

The national income estimates produced annually should never be accepted too slavishly, precisely because they *are* estimates; of necessity they are frequently based on inadequate or untrustworthy data. Apart from the statistical difficulties of estimation, however, there are important theoretical reasons why changes from year to year in the estimated national income cannot always be taken as implying corresponding changes in the economic welfare of a society. The first and most important point is that, as was emphasised at an early stage, economic analysis (which includes national income accounting) can take account in practice only of transactions which are normally carried out through the medium of money. A change may occur in the nominal value of the national income, therefore, simply because some activity which was previously carried on without the use of money has become a monetary transaction, even though there has been no change in the actual supply of goods and services becoming available. For example, the services of housewives are not included in the national income because they are not paid for; if housewives suddenly decided to have their washing done in laundries rather than do it themselves at home, the value of the output of the laundry industry (and therefore the estimated national income) would increase, although there might well be no increase in the total real value of the personal services being rendered in the economy as a whole.

A second point is closely related to this. Since national income is measured by current monetary values, a rise or fall in the general price level will cause the national income total to rise or fall, even if there has been no change in its real value. For example, between 1938 and 1955 the national income of the United Kingdom rose from £4,816 million to £15,226 million—an increase of more than 200 per cent. Does this mean that the British public were three times as well off in 1955 as they had been in 1938? Obviously it does not, since a large part of this rise was due solely to the rise in

the general price level, rather than in the supply of goods and services available. In fact, statistics of price movements indicate that in real terms—that is, excluding changes due solely to those in the general price level—the national income of the United Kingdom rose by only about 30 per cent between 1938 and 1955.

Thirdly, the total of the national income must always be interpreted with reference to the general economic environment in which it was produced. A rise in the national income might be accompanied by an even greater rise in population, so that the *per capita* income, and by implication the average standard of living, would have fallen rather than risen. Again, the national income of a country may rise during war-time, but the switch of resources to war purposes (which may contribute to social or political well-being but can hardly add to economic welfare) may mean that living standards have been reduced. Another possibility is that a higher national income may be the result of inadvisedly long hours of work, or the employment of housewives in industry to the detriment of family life, or of dictatorial powers of industrial conscription and direction of labour. In all cases of this kind one would have to be very cautious before assuming that a rise in the national income indicated a rise in economic welfare generally.

Finally, the absolute amount of the national income has to be considered in conjunction with its distribution, a subject which merits separate discussion. Here it is sufficient to observe that a rise or fall in the national income does not necessarily imply a rise or fall in *everyone's* income. If a rise in national income is associated with a redistribution of income in favour of the rich, it may be that the poorest members of the community are even poorer than before. *Some* people are better off than before, but the welfare lost by the poor may be greater than the increase in the welfare of the rich; hence economic welfare as a whole may be reduced. It is seldom possible to say dogmatically that economic welfare has increased or decreased, because the incomes of some people may have moved in an opposite direction to those of everyone else. As it is impossible to measure the welfare of different people on some objective scale, it can never be quite certain that the gains more than offset the losses, and vice versa.

Despite all these qualifications, the estimate of total national

income per head of the population remains the best single indicator of the wealth-producing powers of a country and of changes in its standard of life. The qualifications discussed above imply that such estimates should be interpreted with caution; this is especially true if a comparison is being made over a period in which prices, conditions of work, income distribution or other major features have undergone significant changes.

4. THE DISTRIBUTION OF THE NATIONAL INCOME

So far this Part has been concerned primarily with the distribution of the national income between the four factors of production. The analysis has shown that in a perfectly competitive world the wealth produced in an economy would be shared by the various factors in accordance with their marginal productivity. Price tends to equal marginal cost under perfect competition; this means that the value of all commodities would be equal to the value of the marginal product of all of the factors employed. That value, in turn, would be paid out in wages, rent etc. in accordance with the marginal productivities of the factors concerned. Since the real world is not a perfectly competitive world, there is not perfect knowledge or perfect mobility. Marginal productivity is seldom capable of precise measurement; obstacles exist to the free movement of factors which would equalise marginal productivity in all sectors; monopoly powers are exploited and rents obtained; Trade Unions and associations bring uneven bargaining pressures to bear.

It would be futile to attempt to embody all these complex and frequently immeasurable forces in one comprehensive analytical structure in the hope of explaining the actual world from *a priori* reasoning. It is necessary, therefore, to resort to empirical investigation to discover what is the actual distribution of the national product between factors. By studying changes in that distribution something may be learnt concerning the forces which are influencing it.

The national income accounts can throw a good deal of light here, and Table XII presents data compiled from them. As was mentioned earlier, interest is not shown as a separate item in the national income table, but is included in profits. Movements in the

relative shares of wages, profits and rent, however, can be traced fairly easily. (In Table XII income from self-employment has been combined with employment income, since the wages element is probably predominant; income from abroad has been combined

TABLE XII: *Distribution of U.K. Gross National Product between Factors, 1938–55*

Source of income	*£ million* 1938	1955	*% of total* 1938	1955
Employment and self-employment	3,669	12,868	70·9	76·7
Profits and interest* - - -	1,036	3,185	20·0	19·0
Rent - - - - - -	470	731	9·1	4·3
Total - - - -	5,175	16,784	100·0	100·0

* Less stock appreciation but including income from abroad and residual error.

Source: As for Table XI.

with profits and interest for similar reasons.) It will be seen that between 1938 and 1955 the estimates suggest a marked shift in the distribution of the national product in favour of labour and against rent-receivers; the share of profits and interest in the two years is much the same. Does this mean that the marginal productivity of labour has increased in relation to that of land? It seems unlikely that an explanation in these analytical terms would be a correct one. The major influences have almost certainly been the increased bargaining power of labour, and government intervention to hold down the level of rents. Labour's bargaining position undoubtedly improved between these years; 1938 was a year in which about 13 per cent of the insured labour force was unemployed (so that the supply of labour was plentiful in relation to the demand for it), while 1955 was a year of full employment in which there was generally a shortage of labour. Rents have been held down by the various Rent Restriction Acts passed by a succession of governments; as a result, the total of rent payments increased by little more than 50 per cent over a period in which the nominal value of the national income was trebled.

Another aspect of the sharing of the national income is that of the distribution amongst persons, rather than factors. The ultimate purpose of the production of wealth, after all, is its consumption

by people, rather than by the theoretical conception 'factors of production', and payments in respect of more than one factor of production are often included in the income of a single individual. How has the distribution of income as between rich and poor altered? From the fact that the share of labour has increased it might be inferred that the relatively poor section of the community had benefited at the expense of the well-to-do, but this would not be absolutely certain.

To obtain a reliable picture of what has been happening in recent years, reference can be made to the personal income tables in the national income Blue Book; the relevant data are summarised in Table XIII. The incomes which matter are those remaining after payment of income tax, since it is this *disposable income* which is a reflection of the wealth-consuming power of different sections of the community.

TABLE XIII: *Distribution of Personal Incomes after Tax in the U.K., 1938–55** (% *of total*)

Income Range (£)	1938		1955	
	No. of Incomes	*Amount of Incomes*	*No. of Incomes*	*Amount of Incomes*
0– 499	96·4†	79·0	64·1	40·1
500– 999	2·4	8·0	31·9	46·5
1000–1999	0·8	5·0	3·3	9·5
2000 +	0·4	8·0	0·7	3·9
Total	100·0	100·0	100·0	100·0

* Based on *National Income and Expenditure 1956* (H.M.S.O., London, 1956), Table 28.

† Partly estimated.

An examination of the data in Table XIII will bring home the significant change which has occurred in the distribution of disposable personal income in the United Kingdom since 1938. In that year about 96 per cent of the persons in receipt of separate incomes commanded, after payment of tax, less than £500 a year, and together received only 79 per cent of the total disposable personal income in the economy. By contrast, about 8 per cent of disposable income was received by a small number of persons, numbering less than 0·5 per cent of the total number, whose

income exceeded £2,000 a year. This marked disparity between the relative size of income groups in terms of numbers, and their relative shares of total disposable income, indicated a fairly pronounced inequality in the distribution of income. By 1955 this inequality had been much reduced. Only 64 per cent of income recipients possessed a disposable income of less than £500 a year and accounted for 40 per cent of total disposable income. Persons in receipt of disposable incomes exceeding £2,000 accounted for about 0·7 per cent of the total numbers and only about 4 per cent of total disposable income. The share of disposable income going to the middle section of the income group had risen (from 13 per cent in 1938 to 56 per cent in 1955) while the importance of the lower and upper ranges had fallen correspondingly. The general rise in money wages and prices between 1938 and 1945, of course, is a major cause of the decline in the size of the amounts received by persons in the very lowest income-group; that a redistribution, as well as a general lifting, of incomes has occurred, however, is evident from the decline in the proportion of the total going to people in the highest income-range.

Two questions are raised. First, why does a substantial, if reduced, degree of inequality persist? Second, what forces have been responsible for its reduction in the last fifteen years or so? The answer to the first question is relatively straightforward. Theoretically, inequality of personal incomes could arise *either* through differences of productivity within factor groups, *or* through an unequal ownership of factors by persons. Some inequality undoubtedly arises from the first cause; some men are naturally more able than others, some land is more fertile than the rest. It is difficult to believe, however, that there are many people whose productivity is so exceptional that they can earn incomes which even after tax were four or five times greater than the average for the community as a whole. Hence the main cause of inequality of personal incomes must be the concentration of the ownership of property—land and capital—in the hands of a relatively small number of people. It seems fairly safe to conclude from the statistics available that in 1954 at least two-thirds of the property in Great Britain was owned by a tenth or less of the adult population.

How has this come about? The main cause has certainly been the *inheritance* of property. Factors ranging from a capacity for hard work and thrift to sheer good luck (if nothing less admirable) have enabled a small number of persons to bequeath substantial sums to their descendants. Then, by investing their inheritances, the beneficiaries in turn have been able to leave even larger accumulations to their descendants—the larger the amount invested the greater the income it will bring in, and hence the more rapidly can the accumulation increase. Furthermore, the inequality of property tends to lead to inequality of earning-power. A man with capital can afford expensive training and education, or to become an entrepreneur; people of equal innate ability may be able to do neither because of their lack of capital. The unequal earning-power accentuates the inequality of property-ownership, and so the process continues in snowball fashion.

Turning to the second question of the causes behind the *reduction* of the inequality, it should be noted that the inequality of opportunity just referred to has been much lessened by the widening of educational facilities in recent years. It is no longer true to say that a university education, or a protracted course of technical training, requires wealthy parents or exceptional self-sacrifice and powers of endurance. It remains true to say, however, that a person will seldom be able in practice to adopt the role of an entrepreneur without substantial capital to invest in his own enterprise.

Nevertheless, the dominant factor at work has been that of progressive taxation—that is, a system under which large incomes or estates are taxed at a rate which is proportionately much higher than that levied on small incomes and estates. For example, in 1938 and 1955 a person with an income of less than £250 before tax would have paid, on average, considerably less than 1 per cent of his income in income-tax. In 1955, a person with an annual income of £15,000 would have paid away an average of about 68 per cent of it in the form of tax, compared with 49 per cent in 1938. Similarly, the proportion of an estate taken by the British Exchequer in the form of death duties increases rapidly as the size of the estate increases, and this rate of progression of death duties has become more and more pronounced in recent years.

It should be noted that the role of the state in the redistribution

of income is not confined to the process of revenue collection. In its expenditure, too, the modern British government redistributes income by spending large sums on social services of all kinds which tend to favour the relatively poor rather than the rich. Expenditures of this kind have become increasingly important in recent years. Between 1938 and 1955, the expenditure of British government authorities on goods and services (largely, but by no means wholly, social services), subsidies and grants to persons, such as pensions and unemployment benefits, rose from £1,084 million to £4,653 million—or from 23 to 31 per cent of the national income. It would be true to say, nevertheless, that the main way in which the modern state attempts to reduce inequality of income distribution is by a system of progressive taxation of income and property.

How far should this process continue? Should the state continue to increase the progressiveness of taxation until it is impossible for anyone in the community to retain a disposable income which is more than, say, 50 per cent greater than the national average, or to bequeath any significant capital to his descendants? On this type of question, as has been seen, the economist is not competent to pronounce. There is no way of comparing the utility enjoyed by different persons; hence it is impossible to discover by any objective test if the community as a whole is better off or worse off than before as a result of a further redistribution of income. This is an issue of politics or ethics, not economics.

Certain considerations of an economic character, however, must never be forgotten. It may be that the redistribution of income or property after a certain point will begin to have serious effects on incentives to work and save. An entrepreneur who is not allowed to retain more than a small fraction of the profits he makes may decide that, considering the risks of loss he runs and the mental strain he is under, the game is not worth the candle. A man forbidden to bequeath his capital to his son may prefer to relax his efforts and live on his capital, rather than leave it for the tax-gatherers. Innovation and risk-taking may decline and decay; capital accumulation may become unfashionable. The results of all this *may* be that the dynamic pulse of an economy will weaken, so that the national product will cease to grow as rapidly as it might otherwise have done. The process of stagnation may be slow and

gradual, but it will be none the less real for that. If all this occurs, the protagonists of the poor and weak may find that they have done harm rather than good, that they have stupified, if not killed, the goose that lays the golden eggs. It is better to have 5 per cent of £100 than 10 per cent of £40; the lower income-groups might be better off with a small share of a high and expanding national income than with a large share of a low and static income.

But many would dispute that such results would follow from any likely degree of redistribution. They would argue that a large element of existing profits, rent and interest is really rent in the sense of an economic surplus. Entrepreneurs, they would argue, would go on being entrepreneurs even if profits were a tenth of their current level; capitalists would still lend at ¼ per cent per annum, because there is nothing else they can do with their capital. Further, the ill-effects on production following from the impoverishment of the rich might be more than offset by the good effects of giving better conditions, education and opportunities to the relatively poor.

The argument is incapable of dogmatic resolution. Perhaps few economic problems of the time are more pressing than that of discovering the correct solution. Is there a point at which redistribution of income so saps the vitality of the capitalistic, or private property, system that the gains in equality are more than offset by the losses caused by disincentive effects on production? If there is such a point, has the United Kingdom reached it? If it is reached, is the price of further equality too high? Or is equality as such sufficiently desirable to compensate for losses in the absolute value of output? The discussion is moving into a realm where the questions are easy to pose, and where there are many of them; society as a whole, rather than the economist individually, will have to grope its way to what it conceives to be the truth in these fundamental issues.

FURTHER READING

MARSHALL, *Principles of Economics*, Bk. VI, Chs. XI–XIII.
PIGOU, *The Economics of Welfare*, Chs. III–VIII.
J. E. MEADE and R. STONE, *National Income and Expenditure* (3rd ed., Macmillan, London, 1952).

H. C. EDEY and A. T. PEACOCK, *National Income and Social Accounting* (Hutchinson, London, 1954), Pt. I.

CENTRAL STATISTICAL OFFICE, *National Income Statistics: Sources and Methods* (Studies in Official Statistics, No. 3, H.M.S.O., London, 1956).

PART VI

MONEY AND BANKING

CHAPTER 20

The Nature of Money and Banking

1. WHY HAVE MONEY?

Economic life revolves around the production of goods and services and their exchange for other goods and services. Previous analysis has shown why it is beneficial for individuals and countries to specialise in the production of whatever they can make best, rather than to attempt self-sufficiency; the gains of division and specialisation of labour are too great to be ignored. Exchange is a necessary consequence; if people are not self-sufficient so far as their economic needs are concerned, they must obtain the goods they do not themselves produce by offering in exchange the commodities in the production of which they are specialising. In this way they obtain not only the gains arising from specialisation but also the gains possible through exchange itself—it was seen in Chapter 4 that, from a given initial income, a higher degree of satisfaction can be obtained by entering into exchange than if people consumed their own produce and nothing else.

The process of exchange is entered into in order to exchange goods for other goods. In real life, however, people seldom do this. A producer exchanges his output not for goods but for money, and when he obtains goods from other people he does it by offering *money* for them, not goods he has produced. In other words, an extra stage is introduced into the process. Instead of exchanging goods for goods (a process which is called barter) people exchange their goods for money and then exchange the money for the goods they want. This is an additional complication and one which, as will be seen later, has given given rise to considerable difficulties and confusion. Why introduce such a complication?

There are three substantial advantages in conducting exchange indirectly through money, rather than directly in the form of

barter. First, the process of exchange is itself immensely facilitated by using money. When goods are bartered, it is necessary for each of the two persons involved to be prepared to accept whatever the other is offering, and this is much more troublesome than it might sound. If I am a watchmaker and want to barter a watch I have made for a pair of shoes, I have to find someone who not only has shoes to offer but who is also willing to exchange them for the kind of watch I make; obviously it might take me a long time to find such a person. It is easy to envisage the difficulties which would be encountered under a system of barter by a blacksmith wanting to obtain a dress for his wife, or a baker trying to obtain a bicycle.

Nor is this the end of the story. The process of exchange by barter would make impossible a world of specialisation and division of labour. If a man is one of five thousand helping to make railway locomotives in an engineering workshop, what commodity could he barter against the necessities of life? How could a policeman barter the product of his labour? What could a bus driver offer a farmer a hundred miles away so as to obtain food? Clearly, the process of barter is inconsistent with an economic system which has attained any substantial degree of specialisation, or with one in which the production taking place results in any but the simplest and commonest necessities of life.

The first and vital role played by money in economic life, therefore, is that of a *medium of exchange*. It permits the separation of exchange into the two distinct acts of buying and selling, without requiring that the seller should purchase goods from the person who buys his product, or vice versa. Hence producers are enabled to concentrate on finding the most suitable outlet for their goods, while buyers can concentrate on finding the cheapest market for the things they wish to purchase. Specialisation is encouraged, because people whose output is not a complete commodity but an intangible contribution to the manufacture of one in which many others are involved can be paid an amount equivalent to their share of the product, and can use this money income to purchase whatever they wish.

Another function of money is to act as a *measure of value*—that is, it serves as a unit in terms of which the relative values of different commodities can be expressed. In a barter economy it would be

a difficult and time-consuming task, to say the least, to determine how many electric-light bulbs were worth one hundredweight of raw cotton, or how many fountain pens should be exchanged for a ton of coal. The process of establishing relative values would have to be undertaken for every act of exchange, according to whatever commodities were being offered against one another, and according to the particular persons entering into the exchange. If I am trying to barter fish for cups and saucers, for example, a great deal would turn on whether the person willing to exchange cups and saucers on these terms is or is not keen on fish.

The establishment of relative values by the process of barter would not only waste a great deal of time and effort. Under such a system the outcome of exchanges would be so uncertain that few entrepreneurs would be prepared to undertake production in advance of demand. Specialisation and large-scale manufacture require that an entrepreneur should pay out in wages, interest and rent a large part of the value of a product long before the product is sold. If the amount for which the product will be sold could not be anticipated within reasonably narrow limits, the risk involved in production would be enormous; the entrepreneur, in any case, would have little idea of how costs of production compared with probable selling price, so that there would be no means of deciding if the production of a commodity was an economic proposition.

A third, and related, function of money is that of acting as a *store of wealth.* It is difficult to envisage saving occurring under a barter system. Obviously there would be no way in which persons rendering services could set aside some of their output to meet possible emergencies in the future. Nor could anyone engaged on only one stage in the production of a commodity save part of his output, since he would be producing nothing tangible. Even when a person actually produced a complete commodity the difficulties would be overwhelming. Most commodities deteriorate fairly rapidly, either physically or in value, as a result of long storage; even if storage were possible, the practice of storing commodities for years on end would involve obvious disadvantages—imagine a coal-miner attempting to save enough coal to keep him in his old age, for example. Yet if wealth cannot be set aside, or can be accumulated only with great difficulty, how can future contin-

gencies be provided for, or capital formation undertaken so as to raise productivity? The use of money disposes of these difficulties; instead of having to store physical commodities, a producer can set aside sums of money which can be used in the future for whatever purposes he chooses. Without money, debts could be incurred only in terms of specific commodities; the modern system of financing capital accumulation from funds derived from a large number of different people and institutions would be impossible.

It is clear, then, that many of the essential characteristics of a highly-developed economic system—division of labour, large-scale production in advance of demand, widespread exchange, and the accumulation of capital—are intimately linked with the use of money. Indeed, without money to facilitate exchange, production and saving, it would be impossible for an economy to develop beyond the primitive level which survives in communities still conducting their economic affairs on a barter basis.

2. REQUIREMENTS FOR AN EFFICIENT MONEY

Money can be defined only in terms of its functions. By that is meant that *anything* is money which serves as a means of exchange, a measure of value, and a store of wealth. Money is as money does. Put more formally, if the word *liquidity* is defined as the ability to turn wealth into any form without loss or delay, then money can be said to be *anything which confers complete liquidity on its holder.* This means that to be money a commodity must be generally acceptable in exchange for goods or services of any kind and for the settlement of debts. The precise substance which is used as money, or its shape and size, is a matter of convenience and convention; all that really matters is that it should confer complete liquidity.

In the course of history an extraordinary variety of substances and commodities has been used as money by different communities and at different times. Among them are shells, animal teeth, particular types of stones, cattle and all kinds of metal; in post-war Germany, cigarettes were widely used as an accepted medium of exchange. In recent centuries the use of gold and silver was adopted by most Western countries until their replacement by paper money in the present century. The only quality which all these things have shared in common has been that they com-

manded the *confidence* of the community in which they were used. That is the only essential requirement if something is to fulfil the functions of money; the community must have confidence in it as a means of exchange and of settling debts. If that confidence is lost, then no commodity, be it stamped with every conceivable mark by the monetary authorities, can survive as money. History is full of examples of monetary units which were once perfectly acceptable and stable but which have been forced into disuse because for one reason or another, the general public lost confidence in them. So long as a person believes that he can obtain goods in return for the commodity used as money, he will be perfectly willing to accept money, whatever its form, in return for his own goods and services. If doubt arises, however, as to the possibility of disposing of money for goods of roughly the same value as those against which it was accepted, people will be unwilling to part with goods and services for it and hence it will cease to be money.

There are other, less important, qualities which are desirable in a money material. It should be fairly *portable*; lead money for example would be inconvenient because of its great weight. Similarly, it should be *homogeneous*: that is, one unit of it should be identical with any other unit. If this were not so, people would almost certainly place a greater value on some units than on others, and there would be a tendency for different prices to be demanded when different units of money were being tendered in exchange. Thirdly, it should be *divisible*, so that it can be used in convenient units. Again, it must be *durable*, so that its quality does not deteriorate with keeping. People would be reluctant to use, say, cows as money because of the risk that a cow might die before it could be used for a purchase.

Gold and silver possess these incidental qualities to such a marked degree that for many centuries they were regarded as the only form of money worth consideration. They are durable and homogeneous metals, and their value is high in relation to their weight; further, their annual output from mines is small in relation to the total stock in existence, so that unsettling fluctuations in the total supply of them were unlikely. These and other incidentals are matters of convenience, however, rather than essential to the nature

of money. Many materials have in fact been used as money for considerable periods of time despite their marked inconvenience in many respects; the people of the Pacific island of Uap, for example, have for centuries used as money enormous stones of such weight that their physical transport is a matter of extreme difficulty.

Nowadays paper money has replaced gold and silver in most highly-developed economies. Such money has all the good qualities of gold and silver, and in addition has two which the precious metals do not possess; it costs virtually nothing to produce, and its total supply is completely within the control of the State. How the use of paper money originally developed is a matter which will be considered in the following section.

3. THE NATURE OF BANK CREDIT

When the word 'money' is used in ordinary conversation people have in mind the notes and coin handled in everyday transactions. These things *are* money, of course, but in a modern economy something like nine-tenths of all transactions carried out—by businesses and government agencies, as well as by private persons —are effected not with the use of notes and coin but through *bank credit*. What exactly is bank credit, and how is it that it has come to be the most important means of carrying out exchanges in the modern economy?

So far as Britain is concerned, the history of bank notes and bank credit can be said to have begun in the seventeenth century. Because of the insecurity generated by the Civil War, the habit grew up amongst the wealthy of depositing their valuable assets (especially gold plate) with goldsmiths; the latter were men accustomed to handling such things, were equipped with vaults and strong-rooms and were generally regarded as being people of integrity. When these assets were deposited, the goldsmiths naturally gave the person concerned a receipt indicating the value of the assets held by them on behalf of the depositor, and embodying an undertaking to hand the assets back to their owner when required.

During this period, gold coins were the most important form of money in circulation, and gold plate could be handed into the Royal Mint for melting down and conversion into coins of the currently required gold content. If people who had deposited gold

plate with a goldsmith required money to pay debts, therefore, they would frequently present their receipt to the goldsmith concerned, have plate minted into coins and pass the coins over to the creditor. Before long it became obvious that a lot of trouble and expense would be saved if the debtors, instead of reclaiming and minting their gold plate, simply handed the goldsmiths' receipt over to the creditor, endorsing it to the effect that ownership of the wealth involved had now passed from the debtor to the creditor. Provided he had confidence in the honesty and reliability of the goldsmith, the creditor would find it just as satisfactory to have payment in gold held safely in the goldsmith's vault as in the form of gold coin. Indeed, more often than not this procedure was more convenient for both creditor and debtor.

At this stage, then, the bank note which is nowadays familiar was beginning to emerge. The goldsmith's *receipt* (not the gold for which it was a receipt) was being accepted in payment of debts, so that it had become a form of money. It will be noted that the factor of confidence was a vital element—the goldsmith's receipt was accepted only if creditors had confidence in his solvency and honesty. Given this confidence, the receipts might pass from hand to hand in settlement of debts, without anyone at any stage actually presenting the receipt and demanding the gold plate in respect of which it was originally issued. Fairly soon the goldsmiths realised that it would be more convenient for all concerned if the receipts were issued for fixed sums—£50, £100, and so on—so that they would be more easily used for settling transactions of different amounts. This brought the receipts even closer to the modern bank note.

So far, however, *credit* had not really entered into the picture, except in the sense that the goldsmith's 'credit', or business standing, was an important factor determining the acceptability of his receipts. But each one of his receipts was still backed by gold in his vaults, and (assuming his honesty) every one could have been honoured simultaneously. In a sense, it was still the gold which was circulating, the receipts being used merely as a convenient way of handing it over. Before long, however, the goldsmith realised that his receipts were passing from hand to hand, and only occasionally did anyone actually come to him demanding the repay-

ment of gold in exchange for the receipt—either because the depositor concerned was of an unusually suspicious temperament or, more probably, because he needed to mint the plate into coins of relatively small denominations for transactions (such as wage payments) for which the use of the gold receipts was unsuitable. It followed, therefore, that goldsmiths could issue notes to customers considerably in excess of the value of the gold in their possession, *provided* that they always had sufficient gold in their vaults to meet any demands likely to be expressed for it. In order to issue these excess notes, the goldsmiths simply adopted the role of money-lenders. There are always people wanting to borrow money for one thing or another and willing to pay interest on their borrowings. The fact that the notes lent by the goldsmiths had (unlike their original notes) no gold 'backing' was of little consequence. All the notes were identical, and circulated on equal terms with one another.

This is a crucial point, so it is worth looking into it in a little more detail. Suppose that the acceptance of gold receipts or certificates has become so widespread that only one person in ten will actually insist that the receipt should be honoured and the gold handed over. If a goldsmith has accepted gold to the value of £1,000 for safe keeping, then he is only likely to be asked to repay £100 of it at any one time. Knowing this, he lends out £9,000 in the form of gold receipts bearing his signature, duly receiving a rate of interest from the borrowers. Receipts bearing his signature are now circulating to the tune of £10,000. As and when they come into the hands of one of the mistrustful 10 per cent who prefer to see gold rather than a mere receipt, he has the £1,000 in gold in his vaults to meet their demands. Since on average only 10 per cent of the £10,000-worth of his receipts will be coming in for conversion into gold, £1,000 will be sufficient to meet all likely claims. The mistrustful one will go away with his £1,000-worth of gold. In due course, however, he will spend it, and it will pass into the hands of one of the majority who prefer to keep their wealth in the goldsmith's vaults; so the £1,000-worth of gold comes back to the goldsmith, ready for the next customer from the suspicious ten-per-cent group. And so it can go on, the majority passing their wealth to one another in the form of the goldsmith's receipts, and

the goldsmith earning an income from the loans he has made. It is important to note that the 90 per cent majority are *not* leaving their wealth idle, any more than the minority; they are simply effecting their exchanges through the medium of the goldsmith's receipts rather than through gold coin.

Once again *confidence* is of the essence in the whole business. The majority are content to accept the gold receipts in return for goods they have sold only because they are confident that the gold is obtainable from the goldsmith and that other people, in turn, will be prepared to accept the receipt from *them* in exchange for goods. If for any reason that confidence is lost, then no-one will be prepared to accept the goldsmith's receipt (no longer believing that the goldsmith is willing or able to honour it); everyone will try to obtain gold from the goldsmith in return for receipts which they can no longer use to settle debts. If the goldsmith has in fact issued receipts in excess of his holdings of gold, then someone is going to be disappointed and the goldsmith goes out of business, unless the latter can persuade all his debtors to repay their borrowings immediately in gold. The fundamental difference between the use of gold receipts in this stage of development and the preceding stage is that receipts (or notes) for which there is no corresponding deposit of gold in the goldsmith's vaults are now in circulation. In the earlier stage it was the gold itself which was being used to settle transactions; the receipts were simply the means by which the transference of gold from person to person was affected. In the second stage it is goldsmith's *credit* which is being used to settle the majority of the exchanges. So long as confidence is maintained in the goldsmith's receipts, then his credit can be used to settle debts just as well as gold; hence it is money.

The goldsmiths have now more or less disappeared, and their place has been taken by banks so far as the business of credit-creation is concerned. Apart from this, the version of events just given is applicable to the situation at the present time. Nowadays the reserve, or backing, underlying credit consists of Bank of England notes, rather than of gold—more will be said about this development in the next chapter—and the receipts of the goldsmiths have been replaced by cheques. The latter change is a refinement of considerable convenience, but nothing more. The

goldsmith's receipt was a promise to repay a certain amount of gold; when it was endorsed by the original depositor, the benefit of the promise was in effect transferred to a third party. Fundamentally, a cheque is exactly the same: it is a request to a bank to transfer to a stated person a certain amount of the bank's indebtedness to the depositor. It is more convenient than notes because the exact sum involved can be written on it, so that transactions can be settled precisely.

Otherwise, a modern bank is following the same procedure as the goldsmiths of the seventeenth century. It has a basic reserve of cash which is acceptable as money to everyone. From experience, it knows that only a small proportion of its depositors will use *in the form of cash* the wealth deposited with it; the majority will instead transfer claims on it to their creditors (written in the form of cheques) which the creditors, in turn, will be prepared to accept and transfer in payment of *their* debts. So long as the bank has enough reserves to meet the needs of the minority who will want to draw out their wealth in cash (about one-twelfth nowadays), it can give credit to customers far in excess of its actual cash reserves. The bank credit used to settle the vast majority of transactions occurring in the economy is of this kind; it rests on a small cash reserve and the confidence of the general public in the banking system. In December of 1955 the total indebtedness of the 11 banks forming the bulk of the British banking system amounted to nearly £6,500 million; the banks actually held £565 million in cash.

4. BANK CREDIT AS MONEY

Since the modern banking system could honour at any one time only a small fraction of its total indebtedness, can a cheque, which transfers this indebtedness from one person to another, be called 'money'? It has been seen that money is anything which performs the functions of money—anything, that is, which is generally acceptable in settlement of a debt. Since the majority of transactions are in fact carried out through bank credit—that is, by cheque—clearly bank debt *must* be money. The fact that, if all the creditors of a bank rushed in at the same time to demand the repayment of their deposits in cash, the vast majority would be refused is quite irrelevant. Given confidence in the banks, the public do *not*

all demand cash at the same time; only a small proportion of bank debt is ever required to be changed into cash, and the banks keep sufficient for the purpose. For most people, the transfer to a creditor of the debt which a bank owes them is a method of settling transactions which is an effective as, and considerably more convenient than, the use of notes and coin. Admittedly the confidence of the public in the banking system is absolutely vital, and without it bank credit would be worthless. But this is true of *any* form of money; if the British public were suddenly assailed by doubts as to their ability to purchase goods with pound notes, no-one would accept any more pound notes and the Bank of England pound note would cease to be money.

The total supply of money in Britain at the present time, then, exists in a number of forms. First, coins made largely of copper and silver; secondly, notes of various denominations issued by the Bank of England and a small number issued by banks in Scotland and Northern Ireland; finally, the deposits (or indebtedness to customers) of the commercial banking system. During 1955 the actual amounts of these various components were as follows (in £ million):

Coin	163
Bank of England notes	1,795
Other notes	110
Bank deposits	6,454
	8,522

Of these only Bank of England notes are *legal tender* in unlimited amounts, while coins are legal tender up to £2; by legal tender is meant a form of currency which the law requires creditors to accept in settlement of a debt. A bank cheque is not legal tender, and a creditor could obtain legal enforcement of a demand to be paid in Bank of England notes. In practice, as has been stated, bank cheques are generally accepted by creditors, and commercial bank credit in fact constitutes far and away the most important element in the total money supply.

One last point is of great importance. In ordinary conversation the word 'money' is frequently used when it is *income* which is really meant: people say that a man earns a lot of money, or has

little money to live on. But money and income are two entirely separate and distinct concepts, and confusion will result if the terms are treated as interchangeable. Money is a definite *stock* of notes, coin and bank deposits existing at a particular moment of time; income is a *flow of wealth* accruing to a person or institution over a period of time. For example, a man may have a high income but hold little money, preferring to spend his income rapidly, or to hold his savings in some form other than money. The quantity of money and the level of incomes may thus move in opposite directions; between 1930 and 1933, for example, the national product of the United Kingdom (i.e. the level of incomes) fell by about 8 per cent, while the total stock of money in existence increased by about 9 per cent. The fact that incomes are usually (although not invariably) paid in the form of money should not be allowed to obscure the important distinction between the two things; electric current is conducted through wires, but this does not mean that electricity and wire are interchangeable concepts.

FURTHER READING

R. S. SAYERS, *Modern Banking* (3rd ed., Oxford University Press, London, 1951), Ch. 1.

SIR DENNIS ROBERTSON, *Money* (Cambridge University Press, London, 1948), Chs. I–III.

* J. M. KEYNES, *A Treatise on Money* (Macmillan, London, 1930), Vol. I, Chs. 1–2.

CHAPTER 21

The British Banking System

1. HISTORICAL DEVELOPMENT

In the previous chapter a brief account was given of how goldsmiths had developed the business of note-issue up to the end of the seventeenth century. So profitable was this pursuit that the goldsmiths (and other tradespeople who saw the possibilities of issuing notes against their own capital reserves) began to concentrate on this side of their business to the neglect of their original functions until ultimately they became full-time bankers, inviting deposits from outside sources. The Bank of England, which was to play an especially important role in the development of British banking, was formed in 1694; it was conducted as a commercial bank like any other, but in return for a loan to the government of the day it received the unique privilege of a charter of incorporation. This meant that the Bank became a joint-stock company, able to combine the resources of a large number of shareholders, whereas no other bank was able to operate on a joint-stock basis until 1826. For more than a century, therefore, the Bank of England had access to considerably greater resources than any other bank, and inevitably acquired a dominant position in the British banking system. So strong was its relative position, in fact, that the private banks soon came to regard the notes issued by the Bank of England as substitutes for a gold reserve.

The ability of banks to issue notes in excess of their reserves unhappily constituted a standing temptation for them to overdo the process and issue far more notes than was advisable. As a result, the eighteenth and early nineteenth centuries were characterised by periodic banking crises. One bank would be found unable to honour its commitment to convert its notes into gold on demand; the news of this would cause panic amongst the depositors not only

of the bank involved but also amongst those of other private banks. When these 'runs' on the banks developed (i.e. when everyone holding bank notes rushed to the bank of issue demanding conversion into gold) it was inevitably found that the total commitments of any bank could not all be honoured simultaneously; bank collapses on a wide scale frequently resulted and these often provoked depression in industries connected, in the role of lender or borrower, with the banks involved.

Two major consequences gradually followed from these periodic crises. First, a succession of legislative measures, together with the natural reactions of the banks themselves to these unhappy experiences, resulted in a strengthening of the private banks through amalgamations. More important still, the policy of these private banks underwent a fundamental change; the emphasis was placed to an increasing degree on the financing of specific commercial transactions by means of credit, while the issue of notes became less and less important. Secondly, the Bank of England, as the bank which possessed greatest resources (by virtue of being a joint-stock organisation rather than a mere partnership) and which had closest relations with the government, came more and more to play the part of the hub of the banking system. When banking crises occurred, the Bank of England frequently lent reserves in the form of notes or gold bullion to banks which were under pressure, allowing the latter to honour their obligations and so stemming the tide of panic before the entire banking system collapsed. On the occasions when the reserves of the Bank of England had become exhausted in this process, the government of the day suspended convertibility of Bank of England notes—that is, it relieved the Bank of the liability to redeem its notes in gold on demand—and declared the Bank's notes to be legal tender. As the government could then authorise the Bank to issue notes in unlimited quantities, there was no danger of any bank being unable to repay depositors in legal tender; hence the crises subsided.

These two tendencies culminated in the *Bank Charter Act* of 1844 which established the Bank of England as the Central Bank of the United Kingdom, and gave to the British banking system the essential features which have continued to the present day. Under the 1844 Act, the Bank of England was given the sole right to issue

notes, except that private banks already having an issue of notes were allowed to retain (but not increase or transfer) it. The remainder of the banking system was therefore forced to gradually abandon the practice of issuing notes, and great impetus was given to the previous trend towards deposit banking—i.e. the acceptance of funds deposited by the general public and the extension of credit to specific persons or for the purchase of relatively liquid assets. The broad structure of the Bank of England has since undergone no change from that provided for in the 1844 Act. The other banks have become stronger and fewer through the process of amalgamation, so that less than a dozen large banks are now responsible for something over 95 per cent of the commercial banking business carried out in Britain. Although it was not foreseen by the framers of the Bank Charter Act, the special position given by it to the Bank of England proved to be incompatible with the pursuit of ordinary banking business in competition with the other banks, and by the end of the nineteenth century the Bank of England had ceased to carry out any substantial amount of business for members of the general public.

By the early part of the present century, then, the main features of the modern British banking system had been formed. At its centre stands the Bank of England, which controls and guides the system and which has ceased to conduct any substantial amount of ordinary commercial banking. Around it is a small number of large commercial banks, carrying on business with the general public but no longer issuing their own notes. How the two parts of the system are linked together, and the functions which each perform, will be explored in the following sections.

2. THE BANK OF ENGLAND

From this brief survey of the historical development of the British banking system, four special features of the Bank of England can be seen, and these correspond with the four functions of the Bank in the present-day monetary system of the United Kingdom. First, it is responsible for the *issue of notes*, of which it now has a monopoly. Until 1844, as has been seen, notes could be issued by any bank. The Bank Charter Act provided, however, that the Bank of England should be divided into two parts one of which,

the Issue Department, was responsible for issuing notes and given the sole right of further new issues. During the war of 1914–18 the Treasury itself issued notes, but in 1928 these notes were withdrawn and the entire note-issue was left in the hands of the Bank of England. At the present time, then, the only notes in circulation are those of the Bank of England except for the small issues of certain banks in Scotland and Northern Ireland (which are, in any case, required to hold a Bank of England note in their vaults for every note they themselves issue).

Until 1931, and except for occasional periods of crisis when convertibility was suspended, the Bank of England was required to convert its notes into gold on demand at a rate laid down by statute. Just as an ordinary bank has to maintain a certain reserve to meet demands for repayment of deposits in cash, so the Bank of England needed to hold sufficient gold reserves to enable it to meet its obligation to convert its notes into gold at any time. Since 1931, however, this obligation has been removed from it, and the note issue is entirely a *fiduciary* one—that is, there is no 'backing' of gold or anything else into which its notes can be converted. Since public confidence in the Bank is complete, however, this in no way affects the acceptability of the Bank's notes, and no-one needs to convert them into some more acceptable form of money. Under the provisions of various Currency and Bank Notes Acts since 1928, the Treasury has complete control over the volume of the notes issued by the Bank, and can give the Bank authority to increase or reduce its note issue to any extent, subject to more or less formal parliamentary approval of increases over a specified limit which are maintained for more than two years.

The 1844 Act required the Bank to publish a weekly return in respect of each Department, showing the position as at the close of business on each Wednesday. Taking the return for 8th August, 1956 as an example, the position of the Issue Department was as follows:

(*£ million*)

Assets		*Liabilities*	
Government securities	1996·2	Notes issued:	
Other securities	0·8	In circulation	1977·7
Gold coins and bullion	0·4	In Banking Department	22·7
Other coin	3·0		
	2000·4		2000·4

The 'liabilities' side of this return needs little explanation, since the only liabilities of the Issue Department are the notes it has issued. Historically, these were liabilities in the sense that the Bank could be called on to exchange them for gold at a fixed rate; this is no longer the case, but the traditional form of balance-sheet accounting has been retained, presumably from sheer reluctance to change. The 'assets' side is a similar archaic survival. When the Bank was obliged to redeem notes in gold on demand it was required to show that it had assets to offset this liability, so as to demonstrate its ability to honour its obligations. Nowadays, with the exception of a small amount of coin and bullion retained in the interests of tradition and administrative convenience, the only assets which the Bank holds are securities—i.e., promises by a borrower to repay a sum of money on a specified or unspecified date and to pay a certain rate of interest in the meantime. Since the overwhelming majority of these are debts of the British government which could be repaid only in the form of Bank of England notes, they have little real significance as any kind of 'backing' for the note issue. In fact, the whole system of accounting adopted by the Issue Department in its weekly return owes everything to historical tradition and nothing whatever to logic.

The second major feature of the Bank's development was its special relationship with the government (it owed its origin to a loan to the government) and it now acts as the *government's bank*. It is the government's agent in the handling of tax revenue, payments of interest on the national debt, raising and repayment of loans etc. For this purpose the Bank maintains what are called *Public Deposits*, which are the bank accounts of the major revenue-collecting and spending departments. All government receipts are paid into these accounts and all government payments are made from them. The Bank of England is to the government, then, what a commercial bank is to a private individual.

The third factor of the Bank's history was its emergence as the *bankers' bank*. From an early stage the London private banks began to keep part of their reserves in the form of a deposit at the Bank of England, treating these deposits and their Bank of England notes as cash equally with gold. The Bank has continued to perform this role. All the commercial banks now hold an account at the Bank of

England—this group of accounts being called *Bankers' Deposits*—and use these accounts to settle indebtedness amongst themselves. Thus, if at the end of a day's business the customers of Lloyd's Bank on balance owe the customers of Barclay's £1 million, Lloyd's will settle this debt by writing a cheque to the value of £1 million on its Bank of England account and handing it to Barclay's; the Bank of England will then transfer £1 million from Lloyd's deposit to Barclay's deposit.

Bearing in mind the second and third functions of the Bank of England, it will be instructive to examine the weekly return of the other section of the Bank, the Banking Department; again the return for 8th August, 1956 is taken as an example:

(£ *million*)

Assets		*Liabilities*	
Government securities	216·8	Capital	14·6
Other securities:		Rest	3·8
Discounts & advances	36·2	Public Deposits	16·7
Securities	26·8	Banker's Deposits	189·1
Notes	22·7	Other Deposits	70·3
Coin	2·0		
	294·5		294·5

Consider first the 'Liabilities' side of the statement. The two small items 'Capital' and 'Rest' deserve brief explanation. The 'Capital' refers to the amounts subscribed by the original stockholders of the Bank, and therefore legally due to them in the event of its winding-up; the curious item 'Rest' represents the undistributed profits of the Bank, which are also in law the property of the Bank's stockholders. Until 1946 the Bank was a purely private corporation owned by its own stock-holders (although for a great many years the pursuit of profit had ceased to be the aim of its policy), so that these items had some theoretical significance. Under the Bank of England (Nationalisation) Act of 1946, however, the ownership of the Bank was vested entirely in the State, so that these items no longer have any real meaning.

The next two items, 'Public Deposits' and 'Bankers' Deposits', have already been discussed; they correspond with the Bank's functions of government bank and banker's bank respectively. The

final item 'Other Deposits', comprises the accounts of customers other than the British government and banks—mainly Commonwealth governments and banks and some foreign Central Banks. Some old-established business houses still possess accounts at the Bank of England, these dating from the period in which the Bank still competed with private banks for commercial business; they are not important in magnitude.

On the assets side—every banker must acquire an asset against each liability, since the sums deposited with him (his liabilities) must be put to some use—the first and easily most important item is 'Government Securities', or debts of the British Government on which the Bank receives interest. The second item 'Other Securities' is divided into 'Discounts and Advances' and 'Other'. The former refers to bills discounted for, and short-term loans made to, the institutions comprising the *money market*; the nature of this item will become clearer in the next chapter, in which the operation of the money market is considered. The item 'Other Securities' comprises securities of bodies other than the British government—particularly securities issued by Commonwealth and foreign governments and certain agencies set up with the participation of the Bank for the purpose of helping to finance development in British industry. The item 'Notes' corresponds with the item 'Notes in the Banking Department' in the return of the Issue Department; together with the item 'Coin', this indicates the reserve held by the Bank to meet demands from the general public (especially the commercial banks) for cash in exchange for deposits. In the past, when the Bank was required to convert its notes or deposits into gold on demand, the size of the Bank's cash (i.e. gold) reserve was a matter of considerable importance. Nowadays the Bank has no obligation to convert its notes or deposits into anything (except new Bank of England notes), so that the cash holdings of the Banking Department have little more than administrative significance.

The three functions of the Bank reflected in the weekly returns of both Departments, then, are those of controlling the note issue, acting as the government's banker and as the banker's bank. The combination of these functions gives the Bank peculiar strength in fulfilling its fourth, and most important function—that of

controlling the monetary and banking system as a whole. How it carries out this function will be described in a later chapter.

3. THE BRITISH COMMERCIAL BANKS

It was seen from the brief summary of the history of British banking that during the nineteenth century there occurred a process of amalgamation amongst the hundreds of private banks which had sprung up all over the country. This has proceeded so far that nowadays 95 per cent or more of the ordinary banking business of the country is carried out by the eleven banks which belong to the London Clearing House, the nature of which will be explained shortly. Of these eleven banks, the so-called Big Five (Barclays, Lloyds, Midland, National Provincial and Westminster) occupy a dominant position. These five banks probably hold more than 80 per cent of the total bank deposits in the country, and each of them has a large number of branches.

What are the functions of the commercial banks? First, they undertake the safe-keeping of cash deposited with them by their customers. Secondly, the possession of these cash reserves enables them to expand credit in the way described earlier—that is, the banks incur debts which they know from experience will be passed from person to person by means of cheques and only in a minority of cases be drawn out in cash. The modern British banker knows that only about one-twelfth of his total debts will be used by the creditor in the form of cash; so long as he has £100 in cash in his vaults, therefore, he can safely incur liabilities to a total of some £1,200. By far the greater part of this £1,200 will be used by depositors to make payments by cheque, requiring only a transfer from one account to another within the bank. But how does a bank incur these liabilities? If a customer deposits £100 in cash, the bank automatically incurs a liability of £100 in the shape of the depositor's account on which he can call at any time. But how can the bank incur extra liabilities amounting to £1,100 so as to bring the total to the £1,200 which the bank knows can be safely incurred with their £100 cash reserve?

The answer to this question may be found by examining the statements of assets and liabilities published each month by the eleven Clearing Banks; an example is given by the combined return

of the Clearing Banks for 30th June, 1956. The assets held by the banks were:

	£*mn*	% *of total*
Cash in hand & at the Bank of England	517	8·5
Money at call and short notice -	447	7·4
Bills discounted:		
Treasury Bills - - - -	1031	18·9
Other Bills - - - -	114	
Investments - - - -	1963	32·5
Advances - - - - -	1975	32·7
	6047	100·0

In a full bank balance-sheet certain other relatively minor items (capital subscribed by shareholders, value of bank premises etc.) are also shown, but the important items are those contained in the monthly returns such as that shown above. The first item on the assets side is the *cash reserve*. This comprises the notes and coin held in the tills of the banks and their deposits at the Bank of England; as a deposit at the Bank of England, like a Bank of England note, constitutes a claim on the Central Bank, it can be treated as cash just as much as notes and coin. At the present time the British banks adopt a firm convention (unlike the position in some other countries, there is no legislation governing the matter) that their cash reserves must always equal a minimum of 8 per cent of total liabilities. Since the banks naturally profit from acquiring the maximum volume of assets consistent with safety, they do not normally allow the cash ratio to rise significantly above 8 per cent. If the cash reserves fall by £1 million, therefore, the banks must reduce their total liabilities by £12 million, or thereabouts; conversely, a rise of £1 million in their cash would allow them to expand their total liabilities by about £12 million.

The next two items appearing on the assets side are *money at call and short notice* and *bills discounted*. Both of these involve the money market, of which more will have to be said in the following chapter. For the moment it is sufficient to say that the former item comprises loans of short duration to persons and institutions in the money market, while the latter comprises short-term securities (or acknowledgements of debt) which are bought and sold on the money market. The next item, *investments*, consists of longer-term

securities which are quoted on the Stock Exchange and can be bought and sold there. The majority of the securities held by British banks are those of the British government, although some securities issued by Commonwealth governments and by the various nationalised industries are also held. Finally, there is the item *advances*—loans to businesses, public corporations, or private persons—which are usually made in the form of an overdraft (the right to overdraw on an account up to an agreed maximum, interest being charged only on the amount actually used). British banks dislike making loans for a long period, so that advances made to businesses are usually in respect of working capital (raw materials, stocks of goods etc.) and not fixed capital, which can pay for itself only after a considerable period of time.

If a bank finds itself able to expand credit, then (i.e. if its cash reserves are more than 8 per cent of its existing liabilities) it can do so by acquiring any of these assets; it can go into the money market or Stock Exchange to buy bills or long-term securities, or it can make loans to money-market dealers, business-men or private individuals who are wishing to raise funds for the purchase of capital goods of some sort. On either type of asset, of course, the bank earns interest. An important consideration which must always influence the decisions of banks, however, is that of liquidity—which, in this context, generally means the ability to convert assets into cash. Although a cash reserve of 8 per cent is sufficient to meet the normal demands for cash which a banker is likely to encounter, the banker has the obligation to repay *all* his depositors in cash on demand, or at short notice, if they should require it. Hence, in the interests of safety, a banker should be able to convert some of his other assets into cash should an unusual demand threaten to exhaust his reserves. A compromise has therefore to be struck between liquidity on the one hand and profitability on the other; for reasons which will become clear after an examination of the theory of interest, the more liquid an asset the lower the rate of interest it will earn. Advances to businesses, for example, are the most lucrative of the banks' assets but are relatively illiquid; if an entrepreneur cannot repay his loan at the particular moment when a bank needs extra cash there is not much the banker can do about it. Cash, on the other hand, is perfectly liquid but brings in no

revenue at all. (Hence, in the bank return shown on page 271, the assets are shown in the order of their liquidity and the reverse order of their profitability.) As a rough working rule it is generally said that British banks like to keep at least 30 per cent of their resources (the 'secondary liquidity ratio', as it is sometimes called) in the relatively liquid forms of cash, money at call and short notice and bills, because both of the latter items can be turned into cash fairly rapidly and without significant loss. This is a much vaguer rule than that governing the minimum cash ratio of 8 per cent, however; indeed, the banks' concepts of what is the desired ratio which these relatively liquid assets should bear to the total not only varies from one part of the year to another, but also appears to depend on the price-level of quoted securities, the prevailing level of business optimism, and a good many other factors.

Apart from the capital subscribed by the stockholders of a bank (and an item called 'acceptances' which is both an asset and a liability, and can be ignored as a self-balancing item), the *liabilities* of a bank are represented by its deposits—the amounts due to its customers. When a bank acquires an asset it must automatically incur an equal liability. If it acquires cash, the persons from whom the cash is acquired are credited with an equal amount in their deposits. If a bank makes a loan, it does so by opening a deposit in the borrower's name of the amount concerned. If it buys securities, it pays for them by crediting the seller's account with the amount of the purchase. The purchase of assets is the obverse of the creation of bank credit; one is achieved by means of the other. Bank deposits may be of two kinds. First, there are *current accounts*, the holders of which may use their deposits (to obtain cash or transfer them by cheque) at any time and without notice; secondly, there are *deposit accounts*, for the use of which notice of a week or more must first be given to the bank, and on which (to compensate for this restriction) the depositor is paid interest.

So far, little has been said about the complications introduced by the existence of more than one commercial bank. Although in eleven cases out of twelve bank deposits will be passed from person to person by means of cheques, rather than drawn out in cash, a good many of the cheques written by a depositor of one bank will pass into the hands of customers of other banks. The other banks,

having credited their customer's accounts with the sums concerned, will then demand payment of the amounts due. The first bank will be unable to meet this demand by offering to open an account for the amount owing, of course, since the other banks will have no desire to hold a deposit with one of its competitors.

It was in order to deal with this problem of inter-bank indebtedness that the *London Clearing House* was instituted, and it continues nowadays together with similar clearing houses in the main provincial cities. The representatives of the members of the London Clearing House (the eleven 'Clearing Banks') meet daily for the purpose of cancelling-out debts arising between the different banks. For example, if the customers of Barclays Bank have written cheques to the value of £10 million payable to customers of the Westminster Bank, while customers of the Westminster have written £10 million worth of cheques to the benefit of depositors in Barclays, then the two sets of claims can be cancelled against each other and no cash at all need change hands. In general, the cheques paid out to other banks by the depositors of any given bank will usually be offset over any long period of time by cheques coming in to its depositors from other banks, since few people can go on for long disposing of more funds than they have coming in. During any one day, however, this balance will not usually occur, and, after all possible cancelling-out has been effected, each member of the clearing house will be left at the end of the day with a net balance due to other banks or owed to it by other banks. Each bank will then settle up these balances by means of cheques drawn on its account at the Bank of England.

The extensive use of the cheque as a means of payment, then, allows British commercial banks to create credit considerably in excess of the cash they hold at any given moment; they are able to rely on the fact that all but a small proportion of the debts they incur will be paid into another bank account. If the accounts concerned are in the same bank, nothing more than a mere bookkeeping entry is involved; if they are in different banks, the transfer will be offset over any fairly long period by corresponding transfers in the opposite direction. Does this mean that the banks have unlimited powers of credit creation? Clearly it does not; there are in fact three distinct limits to the power of any single bank to expand

its liabilities. First, and most important, each bank *must* maintain a cash reserve of 8 per cent, and unless its cash is increased a bank will be unable to increase its total liabilities beyond a total which is twelve times its cash. Secondly, each bank must keep in step with the others. If one of the five big banks tried to expand more rapidly than the others, roughly four-fifths of the cheques drawn on the additional deposits created by it would come into the hands of the other four banks; hence it would persistently lose cash to the others through the settlements in the Clearing House.Thirdly, the banks have some rough working rules as to the distribution of their resources amongst liquid and non-liquid assets, and if they were unable for any reason to obtain the right types of asset in the right proportions they might feel themselves unable to add to their total liabilities, however comfortable their cash reserves.

The minimum cash ratio, however, is the factor which is crucial to modern banking, and it is because of this, as will be shown in Chapter 24, that the Bank of England is able to maintain a close and continuing control over the operations of the commercial banking system.

FURTHER READING

R. S. SAYERS, *Modern Banking*, Chs. 2, 4, 5 and 9.

R. J. TRUPTIL, *British Banks and the London Money Market* (Cape, London, 1936), Chs. I and III.

T. BALOGH, *Studies in Financial Organisation* (Cambridge University Press, London, 1947), Chs. 1–6.

CHAPTER 22

The Money and Capital Markets

1. THE BILL OF EXCHANGE

In the preceding chapter more than one reference was made to a rather mysterious thing called the Bill of Exchange—it featured in the returns of both the Banking Department of the Bank of England and the commercial banks. What, then, is a bill of exchange? Essentially it is a post-dated cheque drawn not on a bank but on a particular person, and its purpose is to secure a loan for a relatively short period of time. Its function is best understood by referring back to the period of history in which its widespread use developed. It is not much more than a century since the railway system of Britain was developed. Before that time, the transport of goods from the manufacturer to the merchant purchasing them, and the disposal of them by the merchant in the market was a process which took a good deal of time. This was true even if the goods were produced and sold within Britain; the time involved was even greater if the goods were either produced or sold abroad. Hence a problem of capital was raised: if the merchant paid for the goods immediately they were produced, his capital would be locked up in them throughout the period during which they were being transported and finally sold, while if the manufacturer waited for payment until the goods were finally sold, *his* capital would be locked up for some time.

Now capital is a scarce factor of production, and neither a merchant nor a manufacturer regards the provision of capital as being part of his function. Hence a conflict of interests arises: the seller wishes to be paid as soon as he has despatched the goods, while the ideal situation from the buyer's point of view would be one in which payment could be deferred until he has resold the goods. These interests were reconciled by the Bill of Exchange,

which passed the job of capital-provision in these trade transactions to persons or institutions specially fitted for it. The bill consisted of a statement by the merchant that he had purchased goods to a stated value from the manufacturer concerned, and of a promise by him to pay that sum in three, or six, months' time, or after whatever period the merchant expected would elapse before the goods were finally sold. When dispatching the goods, the manufacturer would send the bill by post to the merchant, who would sign and return it. The manufacturer would then take it to some person or institution specialising in trade finance (more will be said about such specialists in the following section) and ask for it to be *discounted*—i.e. purchased. If the merchant's reputation was a sound one, the bill would be purchased from the manufacturer at something less than its full value (e.g. a bill for £100 might be bought for £99) and the person discounting the bill would then collect its full value from the merchant after the specified period had elapsed.

In this way, all parties were satisfied. The merchant obtained credit for long enough to enable him to sell the goods and pay for them from the proceeds; the manufacturer was able to recover, at only a small cost, the working capital (cost of labour, materials etc.) locked up in the goods and could thus devote it to further production; the person discounting the bill profited by the difference between the price at which the bill was bought and its full value at maturity. This latter difference, expressed as a percentage of the value of the bill at maturity, was really a rate of interest on the capital advanced on the transaction—it is called the *discount rate.* A three-months bill worth £100 and bought at £99 would carry a discount rate of 4 per cent—the difference (£1) divided by the value at maturity (£100) expressed as a rate *per annum.*

Bills of this kind, arising from actual transactions involving goods, are known as *trade bills*; they are classified as inland or foreign bills according to whether the movement of goods concerned is an internal one or one which involves exports or imports. Their attractiveness as a means of raising short-term capital was so great that another type, the *finance bill,* was soon developed. This is similar to the trade bill except that goods are not involved; it is, in fact, simply a promise to pay a specified sum of money on a date

in the future by a person seeking to raise capital for a short period. One important type of finance bill is the *Treasury bill*; this is used by the British government on a large scale as a means of meeting an excess of expenditure over revenue. Generally, the expenditure of the government is occurring more or less evenly throughout the year, but its receipts from taxation are heavily concentrated in the final three months of the financial year. Consequently, there is a need to borrow funds for the months when expenditure exceeds revenue, and towards the end of the nineteenth century the British government hit on the idea of raising such funds by means of the type of bill which had become a familiar instrument of short-term trade credit.

The Treasury bill is therefore a promise on the part of the government to pay the holder of the bill a specified sum on a date in the future. Nowadays these bills are usually 91-day bills—i.e. they are repaid 91 days after the date of issue—and are issued in denominations of £5,000 or multiples of £5,000. The Treasury bills offered in the market are known as *tender* bills; the Treasury announces every week the total amount of bills which are being offered and the institutions of the money market are invited to tender (i.e. bid a price) for whatever volume they are prepared to buy. The higher the price obtained for the bills, of course, the lower the discount rate which the government is forced to pay on its borrowings, and the Treasury disposes of as many bills as possible to the highest bidders. Other Treasury bills, known as *tap bills*, are issued directly to official agencies of various kinds which have idle funds they are prepared to lend to the Exchequer; the rate of discount paid on tap bills and the quantity of them issued each week are for some reason regarded by the Treasury as matters of the greatest secrecy.

During the present century the relative importance of the various types of bills of exchange has altered considerably. Immediately before the First World War, the volume of trade bills in circulation in London was about £500 million, compared with a figure of about £16 million for tender Treasury bills. By 1936 the volume of commercial bills had dwindled to less than £300 million while that of tender Treasury bills had risen to nearly £600 million. In the 1950's the volume of commercial bills circulating in the

London money market has probably averaged substantially less than £200 million; the issue of tender Treasury bills, on the other hand, has risen well above £3,000 million. The Treasury bill is now unquestionably the dominant element in the market.

What is the explanation of this development? The enormous rise in the volume of Treasury bills is a reflection of the expansion in the British national debt as a whole from less than £640 million in 1900 to about £27,000 million in 1954; the government borrowing necessitated by two World Wars has been the major factor at work in this. The decline of the commercial bill has occurred mainly as a result of the tremendous improvements in transport and in banking during the present century. The former has much reduced the time-lag between production and sale, and hence the amount of capital needed to finance trade; the latter has enabled business-men to finance their transactions through the more flexible and convenient medium of bank credit. This change in the character of the supply of bills to the London money market, as will be seen later, has important implications for the role of the market itself.

2. THE NATURE AND FUNCTIONS OF THE MONEY MARKET

Put briefly, the purpose of the money market is to supply capital for short periods of time and to carry the risks involved in this process; the discount rate is the reward earned by the market for performing these functions. The risks in question are of two kinds: the *trade* risk and the *monetary* risk. By the trade risk is meant the chance of loss because of the inability, or unwillingness, of the borrower to honour his obligation to repay the capital advanced. For example, if a bill is discounted, there is always the possibility that the person on whom the bill is drawn (i.e. the person promising to pay the stated sum in 3 or 6 months' time) may prove to be bankrupt when the bill matures. If this happens, the person who has discounted the bill will claim the due amount from the person who brought the bill in for discounting—i.e. the original creditor who in effect borrowed the capital on the strength of the debtor's promise to pay. If this fails to result in the repayment of the bill, then the discounter of the bill has lost his capital. The institutions in the money market are highly specialised in their trade and make it their business to know whether a bill is a good risk or not, but

whatever precautions are taken the element of risk is always present to some degree in the discounting of a commercial bill.

The monetary risk is of rather a different kind. Most institutions which discount bills do so with the aid of borrowed money (most of it being lent by the commercial banks). On these funds the institutions concerned have to pay interest, and they are usually liable to have to repay the loans at very short notice if required. They make their profits by discounting bills at a rate which is slightly higher than that which they have to pay on their own borrowings. If they should be called on to repay funds shortly after the current market rate of discount has risen (why this should happen will be seen in Chapter 24) one of two possibilities will be open to them. The first is that they can hold on to the bills they have discounted with the borrowed money, and repay their creditors with money borrowed elsewhere at the current (i.e. higher) rate of interest; if this procedure is adopted they may well have to pay to the new source of funds a higher rate than that which they are earning on the bills they hold, consequently suffering a loss instead of making a profit. The second possibility is that they can repay their debts by selling the bills they have discounted—i.e. by *re-discounting* them; if the current discount rate has risen, however, they will obtain for the bills a lower price than that at which they originally bought them. For example, say the discount rate rises from 4 to 5 per cent; a 3-month bill discounted at the old rate would have been bought for approximately £99 (£1 interest for 3-month credit being roughly 4 per cent per annum), but when re-discounted at the higher rate of 5 per cent it would fetch only £98 15s., assuming it still has three months to run. In either case, then, the institution discounting a bill suffers a loss because of the change in market interest rates.

The institutions in the money market which specialise in this business are four in number. First are the firms of *brokers*, either one-man businesses or partnerships, which use their own capital plus bank loans. Most of them act as mere agents, discounting bills for merchants seeking to raise credit and then immediately re-selling the bills to an institution wishing to hold bills until their maturity. Except in a minority of cases, brokers do not command sufficient resources to hold bills for their full life of three or six months, and act merely as intermediaries between buyers and sellers of bills.

The bulk of discounting properly so called is carried out by *discount houses* (there are about a dozen of them in the London market with three dominant joint-stock houses); these obtain resources from (*a*) the capital subscribed by the partners or shareholders, (*b*) the commercial banks and (*c*) deposits from the general public. With these resources they are able to hold a large volume of bills until they fall due for payment.

A third set of institutions are known as *acceptance houses*. By virtue of long experience in trade finance, and especially international finance, they are able to judge the quality of a bill and to add their own credit-standing to a bill by *accepting* it in return for a commission. That is, a borrower who fears that the name on a bill of exchange is not well-known, so that it would be regarded by the market as a rather risky proposition, can go to an acceptance house having special knowledge of the trade or country from which the bill has originated, and ask it to accept the bill. The acceptance house, with its specialised knowledge, will be familiar with the bill's credit-value (or at least will have the means of investigating it) and, if it is satisfied, it will accept the bill—i.e. guarantee to pay the amount due if the person named on the bill fails to pay up at the date of maturity. Such an acceptance by a firm of standing will naturally reduce the risk attached to the bill; hence it will be discounted at a relatively low rate.

The final set of institutions are the *commercial banks*, which enter into the money market in three ways. First, they discount bills (the third asset in the return shown in the previous chapter) either directly, as does a discount house, or by purchasing bills from discount houses after a part of the life of the bills has elapsed. (Nowadays all the Treasury bills held by the banks are obtained through discount houses in this way.) Secondly, the banks carry on acceptance business, especially for their own customers who are naturally well known to them. Finally, they lend money to brokers and discount houses (money at call and short notice) for periods ranging from overnight to a week or ten days, the loans usually being renewed as and when they fall due for repayment.

It may be asked why the banks should lend money to discount houses, rather than discount the bills themselves and so enjoy the profits otherwise made by the discount houses. So far as trade bills

are concerned, the reason advanced in past years was that by leaving some bills in the hands of the market, the banks were able to acquire a liquid asset (money at call) which brought in some revenue but carried none of the commercial risks attached to trade bills. With the decline of the importance of trade bills, however, this argument has lost a great deal of its force. Treasury bills carry no such risks, and it is in buying Treasury bills, and other short-term government securities, that the market currently uses the bulk of its borrowings from the banks.

Given the dominance in the present-day money market of government securities carrying no risk of default whatever, is there any justification for the retention of the various institutions which were designed primarily to handle trade and commercial risks? If the banks decided to purchase *all* their bills directly, rather than through discount houses, and to discount bills with funds at present lent out to the money market, there is no doubt that the money market would disappear. Why should it be retained? To some extent, it performs a useful function in dealing with such commercial bills of a risky nature as still circulate, and it is possible (although most unlikely) that such bills will regain in the future the importance they have had in the past. Much more important, however, is the consideration that the money market still retains its position as an intermediary between the commercial banks and the Bank of England—the actual way in which this relationship operates in monetary control will be described in Chapter 24. If this link between the Central and commercial banks were to vanish, it is by no means certain from the experience of other countries in which the money market link is absent that any other system of connecting the Central and commercial banks would be as satisfactory.

3. LONG-TERM SECURITIES

It has been shown that the money market comprises a number of institutions which specialise in the provision of funds for the financing of the flow of trade, usually through the medium of the bill of exchange. The funds involved are essentially short-term; most bills are repaid within three months, although a small minority may run for six or nine months. Loans of this duration—

say, anything up to twelve months—are sufficient for financing the provision of working capital which will be embodied in finished goods and sold on the market within a few months.

Vast amounts of capital are required for periods considerably longer than this, however. A firm wishing to build a new factory, or to replace expensive plant and machinery, will need to borrow funds unless it has sufficient reserves of its own, and will hardly be in a position to undertake repayment of the loan within a few months. It may be ten or twenty years before an entrepreneur will have earned sufficient revenue from new machinery or a new factory to repay their entire cost, and he will therefore need to raise funds in a form which allows him a considerable time before repayment is necessary.

The most important means of raising long-term capital of this kind is that of the issue of securities. A *security* is a document embodying a promise by a borrower to pay periodically a fixed or variable return to the owner of the security; often, but by no means always, it also embodies a promise to repay a fixed sum (the nominal value of the security) at or by a specified date in the future. If the return offered to the holder of the security is a fixed amount per annum (usually expressed as a percentage of the security's nominal value), the security would be called a *fixed-interest* security or a *bond*. Such a security involves the supply of capital only; it is a mere title to a debt. If the return is not fixed, but is dependent on the earning of profits, then (following the logic of Chapter 18) the supply of entrepreneurship, as well as of capital, is involved; the security is evidence of a share in the ownership of an enterprise, and not a debt. As was pointed out in Chapter 9, such securities are called *shares*.

A security, then, being a claim to a fixed income or a share in profits, is a marketable asset which can be bought and sold. By selling securities an entrepreneur will obtain funds to purchase capital equipment; in return he incurs the obligations imposed by the terms of the security.

The institutions which deal in the purchase and sale of such securities are together known as the *capital market*. They can be divided into two groups, one (the new issue market) dealing with securities being issued for the first time, the other (the Stock

Exchange) dealing in securities which have already been issued to the market and which are subsequently transferred from one owner to another.

4. THE NEW ISSUE MARKET

The main institutions of the new issue market are *Issuing Houses* and firms of *underwriters*. When an enterprise is seeking to raise funds from the general public, a good deal of preparation is required either by law or by the need to attract potential lenders. A prospectus must be drawn up, indicating the nature of the company, its record of profits in recent years and its plans for the future. Advice has to be sought concerning the terms on which the money is to be raised; the borrower naturally wants to raise funds on the easiest terms possible, but if all the required capital is to be raised the public has to be sufficiently attracted. The issuing houses carry out all these tasks and all the administrative work such as that involved in advertising the issue, collecting the subscriptions and dispatching securities. They are a good deal more than administrators, however; their business is to know the state of the market for new loans and to ensure the success of the issue so far as they can. They are usually paid on a commission basis for their services.

The underwriters provide borrowers with a form of insurance. If a borrower is anxious to ensure that he will succeed in raising the full amount of the loan he is seeking, the issuing house will arrange for the issue to be underwritten by one or more of the firms of underwriters in the market. In return for the payment of a commission, the underwriters will promise to buy a stated number of the securities at an agreed price. If the public purchases the whole issue, the underwriters will not be called on; if some of the securities are left in the hands of the issuing house, however, the underwriters will have to purchase them out of their own resources. They will subsequently sell them on the Stock Exchange, but they run the risk that the price at which they dispose of them will be lower than the price at which they took them up.

Who are the purchasers of newly-issued securities? Individual investors may take up some with their personal savings, but nowadays institutions such as insurance companies, pension funds, and industrial concerns with accumulated reserves provide the bulk of

the funds entering the market. If government securities are being issued of course, the banks may be heavy subscribers.

Certain specialised institutions in the new issue market are worth mention. *Building societies* collect deposits from the public and use most of them to finance the purchase of property by private persons and companies. Several organisations supported by the banks and insurance companies, such as the *Finance Corporation for Industry* and the *Industrial and Commercial Finance Corporation*, specialise in providing capital for concerns which are seeking sums too small to be raised by the relatively expensive process of a new issue on the market, or firms which are too new to have a sufficiently long profit record to attract the ordinary investor. The British commercial banks, as was noted earlier, are reluctant to invest in relatively illiquid industrial capital, but they do lend to businesses at short-term; such loans are frequently renewed from time to time, becoming in effect relatively long-term loans.

5. THE STOCK EXCHANGE

Few investors would be happy about sinking resources in a company if there was no way of recovering them, short of winding-up the entire concern and selling its assets—a process which in any case could hardly be brought about by an individual investor. The Stock Exchange exists in order that existing securities may be disposed of by a sale to someone else. Suppose Mr. X invests £1,000 in Company A and subsequently wishes to get his £1,000 back. He can withdraw his funds from the bonds, or shares, by selling the securities in the Stock Exchange to Mr. Y; X recovers his money and Y becomes, in effect, the person who has invested £1,000 in Company A. The ability to recover funds in this way, of course, makes the placing of capital in industry a much less risky affair than it would otherwise be, and accordingly the flow of funds to industry is immensely stimulated.

The major Stock Exchange in Britain is that established in London; there are Exchanges in several provincial cities, but they are of much less importance, and in any case all of them are modelled on the London Stock Exchange. The *brokers* are the people who transact business with the general public, taking orders

and delivering securities or the proceeds of them. The *jobbers* are the people who actually deal in securities on the floor of the Exchange. They deal only with brokers, having no direct contact with the public.

When an order for the purchase or sale of securities is received by a broker, he goes into the Stock Exchange and seeks out a jobber specialising in the kind of security involved in the order. Some jobbers deal only in gilt-edged stocks (those issued by the British government and a few other very high-grade bonds); some deal in oil shares, some in foreign government stocks, and so on. The broker will ask the jobber to quote a price for the security concerned, without disclosing whether he is proposing to buy or to sell securities. The jobber, in return, will quote two prices; a higher price at which he is willing to sell the security and a lower price at which he is willing to buy. If and when the broker obtains what he considers to be a satisfactory quotation, the deal will be closed by means of an informal but none the less binding agreement between broker and jobber.

The Stock Exchange year is divided into a series of periods known as 'accounts'; most of these last two weeks, but a few last three weeks. At the end of each account there occurs a 'settlement day' on which all transactions have to be concluded—i.e. the securities must be delivered and the purchase price handed over. In certain circumstances a buyer or seller wishing to postpone the final delivery of and payment for the securities can arrange with the other party for the agreement to be carried over to the next settlement day; for this privilege a special fee known by the curious titles of 'contango' or 'backwardation' must be paid.

Does the Stock Exchange perform any useful function other than that of bringing in commissions for the brokers and jobbers engaged in its business? There can be no doubt that it does, and in two respects. First, as has been seen, it greatly increases the mobility of invested funds, and so reduces the risk attached to the lending of capital. Through the Stock Exchange, any investor can buy or sell securities as he pleases, provided he is prepared to accept whatever price is currently ruling for them. Furthermore, the existence of a highly-developed and specialised securities market gives an investor the opportunity of seeking expert advice

and guidance on a wide range of securities; the Stock Exchange Committee's practice of requiring companies to fulfil certain conditions, such as those concerning the availability of information, before allowing the shares to be quoted on the Exchange is a further protection for the investor against fraudulent borrowers.

Secondly, through the Stock Exchange a current assessment of the value of different types of asset is constantly available. For example, if the demand for shipping shares declines, the price quoted for them on the Exchange will fall, and the fall may proceed so far that the cost of acquiring a shipping line through the purchase of securities will fall below the cost of starting a new one. In this way, the economy is given an indication that ships are in excess supply, and the use of capital in the construction of new ones will be discouraged. Conversely, an expectation that the demand for motor-cars is going to exceed the supply will cause a rise in the price of motor shares; the scarcity will be expected to increase profits and hence dividends. This rise in the price of motor shares will serve as an indicator to those contemplating capital construction that the community is in need of more motor-car factories.

The possibility of changes in the price of securities gives rise to *speculation*—that is, the purchase of securities in the hope of reselling at a profit rather than as a permanent investment. If such speculation is based on considerations of the kind just discussed—genuine assessments of the changing economic values of different types of assets—it is thoroughly useful and advantageous, since it anticipates future scarcities or surpluses of different types of capital goods and thus guides current capital formation.

Sometimes, however, speculation is more akin to gambling; dealers may buy or sell securities (i.e. they may be 'bulls' or 'bears', respectively) in anticipation of price movements whether or not they believe such movements to be well-founded and justified. Experience shows that speculation of this kind, often being based on completely irrational forces, can become so out-of-hand as to deteriorate into sheer hysteria. Security prices are driven up or down by the spread of what can only be called mob emotion, in which the optimism or pessimism of one person in itself generates similar sentiments in other people in a cumulative fashion. Under

such circumstances the prices quoted for securities on the market can lose touch completely with the real value of the assets underlying them, and the resulting wild swings of prices can only damage the economy as a whole even though the more expert (or luckier) speculators succeed in achieving substantial capital gains.

Like most human institutions, then, the Stock Exchange can perform services of considerable value to society, but can also be abused by a minority acting in their own interests who lose their heads in the process. In the United States the experience of the great Wall Street crash of 1929 stimulated the monetary authorities of that country to acquire powers of control over the extension of bank loans for the purpose of dealing on the Stock Exchange. In Britain the authorities have not attempted to acquire statutory powers of a similar kind. Perhaps this is because the banks are satisfied that within the existing framework they could prevent excessive speculation financed with bank credit; perhaps it is because the authorities have confidence in the stability and caution of the modern generation of dealers on the London Stock Exchange.

FURTHER READING

R. S. SAYERS, *Modern Banking*, Ch. 3.
T. BALOGH, *Studies in Financial Organisation*, Chs. 8–11.
N. MACRAE, *The London Capital Market* (Staples Press, London, 1955), Chs. VIII–XIV and XX.
R. J. TRUPTIL, *British Banks and the London Money Market*, Chs. III–IV.

CHAPTER 23

The Theory of the Rate of Interest

1. THE NATURE OF LIQUIDITY PREFERENCE

The functions of money and the institutions which comprise the monetary system having been examined, it is now possible to return to a problem which was left on one side when the various factor-incomes were discussed in Part V—the determination of the rate of interest. Interest is the payment made to owners of capital; but *why* has it to be paid, and what factors determine it?

The analysis can begin by considering the question: why should anyone want money? This may appear at first sight an extremely odd question to ask. Anyone knows why he wants money; it is the one thing which practically everyone would be glad to have more of. But why? If the reasons are examined a little further it will be found that people want money to buy a new car, or a holiday abroad, or a house or a thousand other things—all the material things which can help to make life more comfortable and pleasant. In other words they do not want to hold money at all—they want *to get rid of it*, to spend it on all the things they *do* want. Money as such is useless; it is the one form of wealth which can yield no satisfaction or pleasure. People cannot eat or drink or wear it; it is a barren and sterile asset. The man who *does* derive pleasure from holding money—the miser—is the exception who proves the rule; he is a man who is regarded by the community as being a little mad. Yet, as has been seen, at any moment of time there is in existence an enormous stock of money—coins, notes and bank deposits—every penny of which must be held by someone; and they cannot all be misers. Why should this huge stock of wealth be held in the sterile form of money, rather than used to purchase goods and services which can add directly to well-being?

The answer lies in the functions of money. Money, it was noted

earlier, confers on its holder complete liquidity—the ability to turn wealth into any form without loss or delay. Now this quality which money possesses corresponds to the desire to possess liquidity felt in some degree by every person or institution; this psychological characteristic is usually called *liquidity preference*, meaning the relative partiality for liquid rather than illiquid assets. Liquidity preference will be felt more strongly by some people than by others, and more at some times than at others, but to some degree it is felt by everyone. What motives underly it?

There is first the *transactions motive*. For most persons and institutions expenditure is occurring constantly—daily purchases are made in shops, cinemas and so on, while businesses have to pay out wages and raw material costs. Income, however, is received only at fixed intervals; persons are paid once a week or once a month, while businesses may receive revenues from the sale of goods months, or even years, after incurring the first costs of production. Some cash must therefore be held to bridge this interval of time between each receipt of income, and the further apart these receipts are, the larger the balances which will be required. For example, if a man earns £7 a week and spends £1 each day, he will need to hold an average of £3 10s. in cash—£7 at the beginning of the first day of the week and nothing at all at the end of the last day. If he earned exactly the same income and maintained exactly the same rate of expenditure but was paid monthly instead of weekly, his average cash holding would be about £15—say £30 at the beginning of the month and nothing at the end. Similarly, a business will require larger cash balances the longer is the interval between the commencement of the production of a unit of output and its final sale. Apart from this time-interval, the transactions demand for cash will depend on the level of incomes and prices ruling in the economy. If prices and wages are suddenly doubled, for example, the average wage-earner will obviously need an average holding of cash which is roughly double the previous average, assuming that he maintains the same rate of consumption of goods and services. Given a more or less stable price-level, however, the demand for cash for transactions purposes would not be expected to oscillate much in any short period, because income-payment and expenditure habits do not change quickly.

A second motive for holding cash is that known as the *speculative motive*. To understand this motive it is necessary to consider the relationship between current market rates of interest and the price level of securities. Consider a bond carrying a promise to pay an income of £5 a year in perpetuity. If the rate of interest payable on new securities currently being offered on the market is 5 per cent —i.e. borrowers are prepared to pay £5 a year interest for every £100 they borrow—then the price of that bond will obviously be £100, whatever its nominal value (which is usually the amount for which it was sold when it was first issued). On the other hand, if the rate of interest offered by new securities coming on to the market rises to 6 per cent (i.e. borrowers are now willing to pay £6 a year interest for a loan of £100) then the bond will be worth only $\frac{5}{6} \times 100$, or about £83. This must be so, since the bond carries an income of £5 a year, and when £6 a year can be obtained on a new security costing £100, people will be willing to pay for an income of £5 a year only a sum which makes that income equivalent to 6 per cent —i.e. £83. Conversely, if the rate of interest currently obtainable on new securities falls to 4 per cent, the price obtainable for the bond will rise to $\frac{5}{4} \times$ £100, or £125; an income of £5 per year from a security purchased at £125 is equivalent to 4 per cent, the current rate.

In other words, the price of a fixed-interest security moves inversely with the current market rate of interest. The original nominal value of a bond (the amount for which it was originally sold) has no relevance to the determination of its *current* market price. Its quoted price will depend on (*a*) the income it offers and (*b*) the current market rate of interest. If the current market rate is the same as the rate prevailing when the bond was originally issued, then the price at which the bond will be quoted will be the same as its nominal value. But if the current rate of interest is different from the rate prevailing at the time of issue, the bond will now command a higher or lower price than it did then, depending on whether the rate of interest has fallen or risen since the bond was first issued.

Bearing these considerations in mind, it is easy to see why the speculative demand for cash (i.e. the demand to meet the speculative motive) arises. It arises insofar as people believe that the rate

of interest is going to *rise*. Assume that the current market rate of interest is 3 per cent, that a person holds a bond bringing in an income of £4 per annum and that he believes that the market rate of interest will shortly rise to 5 per cent. Inevitably he will prefer to hold his wealth in the form of cash rather than securities, despite the fact that he will earn nothing on wealth held as a cash balance. His bond must have a current price of about £133 (the value of an income of £4 a year when the current rate is 3 per cent); if he is right about the rise in interest rates, its price will shortly fall to £80 (£4 a year valued at a current rate of 5 per cent). If he sells now, therefore, and buys back again *after* the rate of interest has risen, he will make a profit of about £53 (selling at £133 and buying back at £80) which will more than offset the loss of interest. The stronger the expectation that interest rates are going to rise, therefore, the greater will be the speculative demand for cash. There will almost always be *some* demand for cash under this heading, since many investors will wish to keep some of their wealth in the form of cash just *in case* interest rates rise, even though they may think this unlikely.

It may be objected that the investor could switch his capital to *any* form if he thinks interest rates may rise; why should he necessarily want to keep it in *cash*? The answer is that fluctuations in interest rates will cause inverse movements in the market value of *any* asset—such as a house, or a machine, or a piece of land—which yields a fixed income. If a house brings in a net rent of £50 a year, the value of that house (assuming a free market) will be £1,000 if the current rate of interest is 5 per cent: this will be so because £50 on an investment of £1,000 is equal to 5 per cent. If the current rate of interest rises to 5½ per cent, the value of the house will fall to about £909; an income of £50 a year on an investment of £909 is equivalent to 5½ per cent. If the owner tried to sell the house for more, no-one would buy it (for investment purposes, anyway) because the yield obtainable from it would be smaller than that offered by securities currently being offered on the market. Hence, only when wealth is in the form of *cash* is it completely free from this possibility of fluctuation in capital value.

Because interest rates may rise, then, a demand will exist for cash balances on the part of people seeking to avoid capital losses

on income-yielding assets. Other things being equal, the lower the level of interest-rates the greater the possibility of a rise in the future, and hence the larger the speculative demand for cash. On the other hand, if current interest rates are high, the chance of a further rise is relatively slight, and the speculative demand for cash will be small.

There is a third motive for holding cash which is something of a mixture of the transactions and speculative motives. This is usually called the *precautionary motive*. It arises because people believe that certain personal or business contingencies may arise in the future which require cash: for example, an unexpected journey away from home, or the possibility of a bargain if it can be snapped up quickly, or an accident which involves expenditure on repairs. The wealth set aside to meet these contingencies has to be held in cash; if it was held in the form of some other asset (e.g. a security), that asset might be standing at an unfavourable price at the moment when funds are needed, so that it could be turned into cash only at a loss. Since, by definition, the contingencies which the wealth would be used to meet are urgent, it would be impossible to wait for a more favourable market before selling out.

This element in the demand for cash is similar to the transactions demand in that the balances are required to meet current expenditures; it differs from it in that the expenditures concerned are unpredictable, whereas the transactions motive is concerned with fairly regular and predictable expenditures. It is similar to the speculative demand in that it is because of the danger of fluctuations in capital value that the wealth set aside is held in the form of cash, rather than of income-yielding assets. It differs from the speculative demand in that while the speculator *anticipates* an unfavourable rise in interest rates, the precautionary balance is not held because any particular view is held concerning the future trend of interest rates; it arises simply because an asset may have to be sold at an unfortunate moment, and the nature of the emergency precludes the possibility of waiting for a temporary and random drop in security prices to disappear.

These three motives constitute the total demand for cash. They have nothing to do with the motives for *saving*; they are concerned only with the form in which wealth is held. Many influences may

explain why a person saves a given fraction of his income; an entirely separate issue is raised, however, by the question of the *form* in which he holds his savings. Liquidity preference explains why people wish to hold some of their wealth in the form of cash balances, rather than in the form of assets which will yield utility or income.

Given these constituents of the demand for cash, a schedule, known as a liquidity preference schedule, can be drawn up; this shows the amounts of cash which the public wish to hold at different rates of interest. The graph of this schedule will probably be shaped like the curve *LP* in Fig. 33, indicating a small demand for cash when interest rates are high and a large demand for cash when interest rates are low. In other words, the smaller the quantity of cash which is available to the public, the greater will be the reward people will have to be offered if they are to sacrifice more of their cash balances. At very low interest rates, the schedule may well become completely elastic—when interest rates become very low, the possibility of a future rise (and therefore of a capital loss) is extremely strong, and everyone will want to move out of securities and into cash before security prices start to fall.

This schedule gives the demand for money. The preceding chapters have described how the total supply of money is determined. Ignoring notes and coin (which are an unimportant element in the total) the total money supply depends on the cash reserves of the banks (fixed by the monetary authorities) and on the ratio of cash to total deposits which the banks adopt. The supply of money

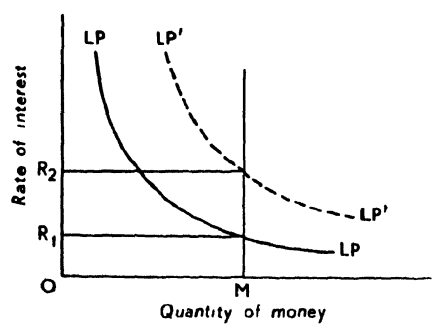


FIG. 33

is therefore fixed at any moment, and does not fluctuate in response to changes in the rate of interest; hence it can be shown as a vertical (completely inelastic) curve as in Fig. 33. The rate of interest will settle at the point of intersection of the two curves.

Like all demand and supply curves, both the liquidity preference schedule and the money-supply curve are liable to shifts to right or left. If the public liquidity preference suddenly increases (because of any factor which enhances the attractiveness of having wealth in the form of cash) the schedule will shift to the right: LP' in Fig. 33. If the quantity of money remains constant, then the rate of interest will rise—in Fig. 33 it rises from R_1 to R_2. Similarly, a shift in the money-supply curve to the right (i.e. an expansion of the quantity of money) with a constant liquidity preference schedule will result in a fall in the rate of interest.

Interest, then, is the reward for parting with liquidity. Given the existing supply of money, which is fixed by the monetary authorities, a person can obtain cash only by persuading someone who already possesses cash to exchange it for some other asset; the rate of interest is the instrument of persuasion. The more reluctant people are to give up the advantages of cash—i.e. the stronger are the forces underlying the demand for cash which have just been described—then the higher the rate of interest which will be needed to induce them to exchange cash for other assets.

2. SHORT-TERM AND LONG-TERM RATES

So far the discussion has been concerned with 'the' rate of interest; it has been assumed that there is only a single rate quoted in the market. In fact there is a whole range of interest rates covering all the different types and grades of securities. One especially important cause of differences in rates of interest is the different length of life for which securities may have to be held before their maturity—i.e. before they are repaid in cash. Normally, a short-term security will bear a lower rate of interest than a long-term security which is identical in all other respects. Why is this? The answer is that the longer the life of a security the more liable it will be to fluctuations in price, so that its liquidity is reduced. The liability to price fluctuations can best be appreciated from a simple example. Suppose the average current market rate of interest rises

from 4 per cent to 5 per cent. The price of a bond with no fixed repayment date (an 'irredeemable' security) carrying an income of £4 per annum will now fall, as was noted earlier, from £100 to $\frac{4}{5} \times$ £100, or £80. Now imagine that there exists a bond similar in all respects except that it is due for repayment at the end of one year from now. If the price of *this* bond fell to £80, an enormous demand would arise for it. At the end of a year, the possessor of the bond would receive £100 as cash repayment *plus* £4 in interest, a total of £104, or £24 more than was paid for it. A gain of £24 in one year on an investment of £80 would be equivalent to a rate of interest of 30 per cent—considerably more than the 5 per cent currently offered on new securities.

Everyone would therefore rush to buy such an attractive security, and its price would inevitably be driven up. In fact, its price would continue to rise until it reached just over £99. People buying it at that price would receive at the end of the year £100 in repayment *plus* £4 in interest, a total of £104 and a net gain of just less than £5. This gain, on an investment of £99, would be equivalent to 5 per cent per annum, the rate now prevailing in the market, and equilibrium would be reached.

Now consider a similar bond repayable in 10 years' time. If the price of *this* fell to £80, at the end of ten years the purchaser would have received £100 in cash repayment and £40 in interest payments—a total of £140 and a gain of £60 over the original purchase price. Ignoring compound interest, a gain of £60 over 10 years would be equivalent to an annual return of £6 on the original investment of £80—i.e. a rate of 7·5 per cent per annum. Once again, the attractiveness of such an investment, in comparison with the ruling rate of 5 per cent on new securities, would force up its price until it reached about £93. By purchasing at this price, and receiving £140 over the security's 10-year life, the purchaser would receive a total gain of about £47, or £4·7 per annum. An annual income of £4·7 on an investment of £93 would be equivalent to about 5 per cent, the rate currently prevailing in the market.

It can be seen, then, that securities differing only in the length of time which must pass before their maturity will fluctuate in price to different degrees as a result of any given change in the current rate of interest. In the hypothetical example just given, a

rise of the rate of interest from 4 to 5 per cent reduced the price of a one-year bond to £99, that of a ten-year bond to £93 and that of an irredeemable bond (one with an infinitely long life) to £80. Long-term securities, therefore, are less liquid than short-term; there is less certainty of being able to convert them into cash without significant loss. Admittedly, a *fall* in interest rates will have precisely opposite results, and in a period when the market is predominantly of the opinion that interest rates will be lower in the future than in the present, long-term interest rates may fall *below* short-term rates. In normal times, however, there is no great degree of certainty attached to expectations about the future, and the *risk* of a given capital loss will weigh more heavily in the minds of investors than the possibility of an equal capital gain. (If it is assumed that the law of diminishing marginal utility applies to income, a reduction in a person's wealth from £100 to £90 will involve more utility than a rise from £100 to £110.)

3. OTHER RISKS AND UNCERTAINTIES

The liquidity of assets is affected by factors other than the length of their life. A debenture in a small and little-known company, for example, will not find a market as readily as one in I.C.I. or British Motor Corporation if its holder should suddenly require to sell it. A foreign bond is subject to special risks because of the possibility of expropriation, or of the freezing of foreign funds by the government of the country concerned, factors which would drastically affect the market value of the security. Moving outside the range of securities, the liquidity of assets such as land or houses will be affected by the expense and inconvenience of selling, or the imperfection of the markets for them. A great many considerations of various kinds influence the liquidity of assets, and hence the rate of interest which will have to be offered in order to induce investors to hold them.

When assets carry a *variable* return—such as that from preference or ordinary shares of all kinds—further complications arise because they involve the factor of enterprise as well as that of capital. The purchase of an ordinary share, like that of a debenture, involves the placing of capital in a relatively illiquid form; the return on the shares must therefore include some reward for the

sacrifice of liquidity. But the shareholder incurs uncertainties over and above those involved in the sacrifice of liquidity. He is not merely *lending* his wealth, as the debenture holder is; he is purchasing part of a company and thus shouldering the responsibilities of entrepreneurship. The debenture-holder is paid (and his capital ultimately repaid, if the debenture has a maturity date) whatever happens to profits. As was seen in Chapter 18, however, the ordinary shareholder receives a dividend only if and insofar as profits are made. The income received on shares, then, is only *partly* interest; part of it is profit, which may be negative. If no dividend is earned on an ordinary share, the holder has not merely received nothing but has also lost the interest he could have earned if he had put his capital into a fixed-interest asset rather than shares. And as the discussion of profit showed, the earnings on ordinary shares will vary widely according to the degree of uncertainty which surrounds the outcome of the activities of different industries.

FURTHER READING

J. M. KEYNES, *The General Theory of Employment, Interest and Money* (Macmillan, London, 1936), Chs. 13–15.

A. H. HANSEN, *A Guide to Keynes* (McGraw-Hill, New York, 1953), Chs. 6–7.

D. DILLARD, *The Economics of J. M. Keynes* (Crosby & Lockwood, London, 1950), Ch. 8.

E. NEVIN, *The Mechanism of Cheap Money* (University of Wales Press, Cardiff, 1955), Ch. II.

CHAPTER 24

The Mechanism of Monetary Policy

1. THE WEAPONS OF THE CENTRAL BANK

In an earlier chapter it was seen that the quantity of money in the United Kingdom, as in most highly-developed economies at the present time, is controlled by the Central Bank as the agent of the government itself. It is one of the main virtues of modern monetary systems that the supply of money is amenable to close and conscious control, and is not subject to the chance influences of the mining of so-called precious metals or the discovery of deposits of stones or shells. The power to influence the supply of money in a community is a most important one—too important, in fact, to be left in the hands of anyone other than the government itself. It has been seen that the rate of interest is dependent to a large degree on the quantity of money; in examining the theory of employment, it will be found that the rate of interest in turn plays an important part in influencing the level of prices, incomes and output. The present task is to examine the way in which the authorities exercise control over monetary conditions; Part VII will be concerned with the influence of monetary changes in the wider context of the economic system as a whole.

On the Bank of England, it was noted in Chapter 21, rests the responsibility for controlling the general monetary system through its influence over the level of bank credit. To carry out this task it has three instruments which are usually operated in conjunction with one another. The first is *Bank rate*, the rate at which the Bank is prepared to discount or re-discount first-class Bills of Exchange. The Bank of England has the traditional role of lender of the last resort; it is the one bank which will never refuse to extend credit to money-market dealers if it is called for. At the same time, it will lend only at its own price, and that price is Bank rate. The Bank

does not undertake ordinary banking business to any significant extent, but the brokers and discount houses of the money market have the privilege of being able to take bills to the Bank of England and sell them—i.e. re-discount them—to the Bank, or, alternatively, to deposit bills as security for loans (of at least seven days' duration) on which they will have to pay interest at ½ per cent above Bank rate.

The Bank of England is careful to keep Bank rate above the current market rate of discount. For example, if bills are being discounted in the market at 3 per cent, Bank rate will be at least 3½ per cent; this means that a 3-month bill worth £100 would be sold to the discount houses for about £99 5s. but could be sold to (re-discounted at) the Bank of England for only about £99 2s. If the situation was reversed, so that Bank rate was *below* current market rates, the discount houses would be able to use their privilege of having re-discounting facilities at the Bank in order to buy bills in the market and re-sell them immediately to the Bank at a higher price; as a result, the Bank would find itself inundated with bills.

The second weapon in the Central Bank armoury is known as *open-market operations*. By this is meant simply the purchase or sale by the Bank of securities, including Treasury Bills, in the ordinary securities market. Such purchases or sales will have important effects. In the first place, by increasing or decreasing the demand for securities the current rate of interest will be directly influenced. More important, such operations will cause changes in the cash reserves of the commercial banks. If the Bank buys securities, the seller will receive in exchange for them a cheque drawn on the Bank of England. He will deposit this cheque with the commercial bank of which he is a customer, and the commercial bank concerned will duly present the Bank of England cheque for payment. The Bank will handle this claim just as the commercial bank handled its customer's—that is, the amount of the cheque will be added to the commercial bank's deposit at the Bank of England. The cash reserves of the commercial bank are therefore increased by the amount of the original purchase of securities. Given a certain increase in its cash reserves, of course, the commercial bank will be able to expand its total deposits by roughly

twelve times as much. Conversely, if the Bank of England *sells* securities, it will receive in exchange for them a cheque drawn on the buyer's account with one of the commercial banks; the Bank will obtain payment of this cheque by reducing the commercial bank's deposit by a similar amount, and the cash reserves of the commercial bank will therefore be reduced. Once again the total deposits of the commercial bank will have to undergo a change which is twelve times as great as that in its cash reserves in order to maintain the 8 per cent cash ratio.

The third instrument at the disposal of the Central Bank is that of *moral suasion*. Because of its long-established reputation as leader and controller of the banking system, any hints or suggestions that the Bank may make to the other banks are treated with great respect; they may even lead to the desired result without the adoption of any other measures. The commercial banks realise that in the last resort the Bank can always enforce its will; hence, if the Bank lets it be known that bank advances are rising too rapidly, for example, or that securities are being sold on such a scale as to damage the market in gilt-edged securities, the commercial banks may prefer to take the hint forthwith rather than wait for action to be forced on them through the use of Bank rate and open-market operations.

This moral power of the Bank of England was given a firm backing in the provisions of the Act of 1946 by which the Bank was nationalised. Clause 4, sub-clause (3) of the Act provided that

> The Bank, if they think it necessary in the public interest, may request information from and make recommendations to bankers, and may, if so authorised by the Treasury, issue directions to any banker for the purpose of securing that effect is given to any such request or recommendation.

The powers given by this clause could hardly be wider. So far as is known, no specific use has been made of them, but they must immensely strengthen the influence of any hints the Bank may give to the banking system; the commercial banks now know that the Bank of England can secure statutory enforcement of its wishes if it proves necessary.

2. CENTRAL BANK POWERS IN ACTION

Having described the instruments which the Bank of England can employ to influence the monetary environment, it is necessary to examine the way in which they operate. Suppose that the Bank thinks it necessary to raise interest rates so as to check spending; *why* the latter should be a result of higher interest rates is a question best left over to Part VII. The Bank will probably first advise the commercial banks that in its view credit is being given too freely and too cheaply. If this advice fails to check the expansion of credit, or if the Bank considers the situation so serious as to justify its immediate enforcement, the other weapons will be brought into play—Bank rate will be raised and securities sold on the open market.

The raising of the Bank rate will usually cause an immediate rise in other short-term interest rates. The discount rate on bills will probably rise, because the discount houses realise that should it prove necessary to re-discount bills at the Bank they will be able to do so only at the high rate—i.e. they will be able to sell their bills at a lower price than was previously likely. Their own discount rates will therefore be raised to cover this increased risk of loss. Further, the commercial banks conventionally maintain a fairly stable relationship between Bank rate and the rates they themselves pay to depositors and charge on loans. Thus the discount houses will have to pay more for their loans at call and short notice, a further stimulus to the upward movement of bill rates. Businessmen will also be charged a higher rate on loans and overdrafts, and this will discourage the borrowing of money (for the purchase of new machinery and factories, for example, or for holding stocks of goods and raw materials).

At the same time, the Bank will reinforce the increased Bank rate by the sale of securities in the open market. As has been noted, this will reduce the cash reserves of the commercial banks, and in order to maintain the 8 per cent cash ratio the banks will be forced to reduce their deposits by twelve times the amount of the reduction in their cash. How will they reduce their deposits? In the first place, loans to the money market will be called in, and discount houses will have to re-discount some of their bills in order to repay

their loans from the commercial banks. As the commercial banks will be trying to restrict credit, the discount houses will be able to raise cash only by taking bills to the Bank of England for re-discount; the market will be 'in the Bank', as it is said. The Bank of England will have made Bank rate effective, because the discount houses are being forced to re-discount bills at Bank rate, or to borrow on the security of bills at $\frac{1}{2}$ per cent above it. Again, the banks may reduce deposits by buying fewer bills and by selling some of their securities, thus causing security prices to fall and interest rates to rise. Finally, the banks may recall some advances and will certainly be reluctant to make any new loans to borrowers. Both short-term and long-term credit will therefore have become more expensive and more difficult to obtain, so that expenditure depending on borrowed money (which covers a great deal of investment expenditure) will be discouraged.

The tightening of credit will not be confined to the banks and money market. The sale of securities by the banks will be augmented by sales on the part of would-be borrowers from the banks who have been refused accommodation and who will therefore have to sell securities if they are to obtain money. The price-level of existing securities will consequently fall, and their yields will rise. This will react in turn on the new issue market. No-one will buy a new security offering a lower rate than comparable securities already obtainable on the Stock Exchange; hence the rate on current new issues will have to rise. Furthermore, issuing houses and underwriters will be less willing to sponsor new issues at a time of falling security prices; the risk involved in having part of a new issue left on their hands is naturally more serious in a falling market than when security prices are buoyant. New capital issues will thus become more difficult to arrange as well as more expensive. Specialised credit institutions, such as hire-purchase houses and building societies, will also have to make their credit more expensive and more difficult to obtain. To attract the resources which they use to extend credit, they will have to offer higher rates so as to compete with the rates currently prevailing in the securities market; naturally their charges will have to be raised to cover this, while the volume of their business will inevitably be curtailed by the general shortage of credit.

It can be seen, then, that the impulse originating in the raising of Bank rate and the sale of securities by the Bank of England is carried to the outer fringes of the capital market and influences the entire credit system. It is fairly easily imagined how the process could operate in the reverse direction. A fall in Bank rate will reduce the market rates on bills; security purchases by the Bank will increase the cash reserves of the banking system and encourage the banks to expand credit, making loans cheaper and easier to obtain. Security prices will rise as the banks add to their investments, and the new issue market will be stimulated.

3. THE EFFECTIVENESS OF MONETARY POLICY

A good deal of controversy has raged over the question of how far the Bank of England can actually influence the pace of economic activity through the use of the policy instruments the operation of which has just been described. Theoretically the Bank can so contract or expand the volume of credit, and so manipulate the level of interest rates, that it can achieve almost any desired effect on economic conditions. There is a substantial degree of doubt, however, concerning the real power of monetary policy to influence the economy in practice. Put in simple terms, these doubts concern, first, the power of the Bank to regulate the volume of credit and, secondly, the importance of changes in the volume of credit as a means of influencing economic activity.

On the first question it is agreed that the Bank can *contract* credit to any degree it wishes, and thus raise market interest rates as high as it desires. With a contraction in its cash basis, the commercial banking system must reduce total deposits; it has no choice in the matter. To some extent the banks could postpone such a reduction by continually calling in money market loans to replace the cash they were losing; the discount market would borrow from the Bank of England and hand over the proceeds to the banks. Obviously this could go on only so long as the Bank of England was willing to see a policy of credit contraction nullified in this way; if Bank rate was raised enough, the discount houses would soon find borrowing at the Bank so ruinous that they would allow all their bills to mature and make no attempt to replace them. In this way the discount houses would pay off all their debts to the banks, and

the latter would no longer be able to prevent the Bank of England from reducing their cash reserves.

A more important consideration is that the commercial banks could prevent a decline in their cash reserves by demanding repayment of their maturing Treasury bills rather than replacing them (as they usually do) with newly-issued bills. To meet this demand, the Exchequer would have to repay the banks by transferring the due sums from its account at the Bank of England ('Public Deposits') to the banks' account ('Bankers' Deposits'). Hence the banks would be able to force the Exchequer to replenish their cash reserves (i.e. their Bank of England deposits) as rapidly as the Bank of England depleted them by open-market operations. In practice, however, this is a process which could not long continue. The decline in the banks' Treasury bills would cause a fall in their secondary liquidity ratio (i.e. the ratio of cash *plus* money at call *plus* bills to total assets). When this secondary ratio fell to its permissible minimum, the banks would be unable to reduce their holdings of Treasury bills any further, and therefore would no longer be able to force the Exchequer to replenish their cash reserves through the repayment of maturing bills. (Precisely because the minimum liquidity ratio of British banks had previously been a somewhat nebulous entity—as was remarked in Chapter 21—the authorities 'requested' the British commercial banks in 1955 not to reduce their liquid assets below the 30 per cent. level, at least in the general credit conditions then prevailing.) Although the banks could delay the impact of open-market operations on their cash reserves if they initially held a large volume of liquid assets, therefore, the delay could only be a temporary one; ultimately the policy of the Bank of England would be bound to prevail, and the cash reserves of the commercial banks would begin to fall.

In *expanding* credit, however, the power of the Bank of England is less absolute. It can add to the banks' cash reserves, but there is no way by which it can compel the banks to take advantage of their increased reserves. If the banks considered the economic outlook to be so black that they were unwilling to take up securities or grant any further loans, the increased cash reserves would simply not be used as the basis for a further expansion of deposits.

Similarly, although interest rates are bound to rise as the volume of money grows smaller, it does not follow that they can be driven down to zero by a continual expansion in the money supply. The liquidity-preference curve may become perfectly elastic at low interest rates, as it did in Fig. 33. Security prices may rise so high (i.e., interest rates fall so low) that the market may become absolutely convinced that they are bound to fall soon. If such a view does prevail, the only effect of credit expansion will be a wholesale movement out of securities (as their price tends to rise even further under the pressure of buying by the banks) and into idle cash balances.

Much the same is true of the effects of changes in monetary conditions on the rate of borrowing for capital formation. Credit contraction is usually attempted in periods of rising prices and profits, and in such periods entrepreneurs may be so optimistic that no practicable rise in the rate of interest will deter them from borrowing. But as credit becomes harder and harder to obtain, borrowing is bound to be restricted eventually. If the Bank of England reduces the cash reserves of the commercial banks the latter will be faced with the necessity of reducing deposits; hence they will simply refuse to lend, whatever rate of interest is offered for a loan, and entrepreneurs will be unable to obtain funds however optimistic they may be.

Once again the process is less compelling when working in the opposite direction as the result of an expansionary policy on the part of the Central Bank. If business prospects are bad, entrepreneurs will not borrow even if there is abundant credit available free of interest. What is the point of borrowing if new enterprises hold out only the prospect of loss? Even if the banks were anxious to expand credit, therefore, a lack of borrowers might prevent them from doing so.

There are impressive grounds for caution, then, in assessing the power of the Bank of England to influence economic activity through the use of monetary policy. Such doubts are at their strongest, however, in connection with periods of extreme prosperity or extreme depression—that is, when profit expectations are so good as to be undeterred by high interest rates, or so bad as to prevent investment even when interest rates are low. But these are

exceptional situations, and the whole aim of monetary policy should be to correct tendencies before these extremes have been reached. Monetary policy is much more likely to be effective if it is applied *before* movements towards boom or slump have proceeded very far. Even then its precise power is open to dispute; few people, however, would deny that appropriate action by the Central Bank, *if taken in time* (which is easier to say than to do) can play an important part in contributing to the maintenance of stability in the level of activity and prices.

4. SELECTIVE CREDIT CONTROLS

Apart from the question of its power, or lack of power, monetary policy—meaning the adjustment of the overall volume of credit—has been criticised on the grounds that it is *indiscriminate* in its operation. By general credit restriction, *all* borrowers are discouraged, without regard to the social value of the projects for which they wish to raise funds. It makes no distinction between a local authority trying to finance house-building in a slum-clearance project and a speculative entrepreneur wishing to build a new stadium for greyhound racing; most reasonable people, however, would agree that there is a good deal of difference (in terms of social desirability) between such uses of resources.

In fact it is not strictly true to say that general credit control makes *no* distinction between borrowers. What monetary policy does is to expand or contract the overall volume of funds; it leaves to individual bank managers, financial houses and investors the decision as to who is to have the extra funds in a period of credit expansion and who is to be denied credit in a period of contraction. It is by no means obvious that all these individuals through whom monetary policy operates will be unaware of, or indifferent to, the need for giving priority to the projects of greatest social value. But it is undeniable that they are liable to give more attention to the profitability of a project than to its inherent social value.

Because of this, there has been a development in most Western countries of what may be called *selective credit controls*; the authorities have acquired powers to limit or encourage the use of credit in particular, specified types of activity. The object of these powers is the gaining of control over sectors of the economy which for one

reason or another are considered to be of critical significance—they may be of strategic importance, or a major source of exports, or especially vulnerable to damaging speculative booms.

In the United Kingdom, such controls have taken two main forms. The first is the regulation by the Board of Trade of the terms on which hire-purchase transactions may be carried out, and in particular of the proportion of the purchase price which must be found by the buyers as a deposit and the length of time over which the hire-purchase instalments may be spread. As hire purchase plays an important role in the sale of many durable consumer goods —for example, cars, electrical equipment and furniture—these regulatory powers can be used to affect consumer spending to a significant degree.

The second, and more important, selective measure is the control exercised by the Treasury over the raising of new capital. A permanent Capital Issues Committee exists to which all applications for the raising of capital in the new issue market in excess of £10,000 must be referred, and the Committee advises the Treasury as to whether any particular issue should be allowed or disallowed. In its deliberations the Committee is guided by memoranda periodically sent to it by the Chancellor of the Exchequer; these indicate not only whether new issues in general should be encouraged or discouraged in the light of the current situation, but also the industries to which priority should be given and those which the Chancellor feels have no strong claim on capital resources. Copies of these memoranda are also sent to the commercial banks, with a request that the banks should pay due regard to the Chancellor's views when considering application for advances. Although the banks are merely 'requested' to do this, they do in fact follow closely the policy indicated in the memoranda.

There can be no doubt that both of these measures of selective credit control are potentially extremely effective, certainly in the negative sense that they could undoubtedly reduce or prevent expenditure in particular directions to a much greater degree than would be possible with only general credit policy. Like the latter, however, they are much weaker in the expansionary direction. They can do no more than simply remove restrictions on hire-purchase transactions or on the raising of new capital; if spenders or

borrowers prefer to stay out of the market, the controls cannot prevent them.

Experience shows that selective controls are more rapid and effective when used in conjunction with overall credit policy. Although selective measures have proved too valuable for their abandonment to be likely in the future, therefore, they strengthen and sharpen general monetary policy rather than replace it.

FURTHER READING

R. S. SAYERS, *Modern Banking*, Ch. 11.
N. MACRAE, *The London Capital Market*, Chs. XV–XIX, XXI–XXII.
T. WILSON (Ed.), *Oxford Studies in the Price Mechanism* (Oxford University Press, London, 1951).
E. NEVIN, *The Mechanism of Cheap Money*, Ch. IX.

PART VII

THE THEORY OF EMPLOYMENT

CHAPTER 25

Savings and Investment

1. AGGREGATE DEMAND AND EMPLOYMENT

In the economic world the essential problem is always that of infinite wants faced by limited resources, which are the means whereby those wants may be satisfied. It would appear to follow from this that a society could never experience any difficulty in finding suitable uses for the productive resources at its disposal. Land, labour, capital and enterprise could be in excess supply only if the wealth of the society had become so great that all human wants had been satiated, a situation which has never yet arisen and is not even conceivable at the present time. Nevertheless, resources have been in excess supply during many periods in the past, and factors of production have been unemployed for considerable lengths of time; men and machines have been kept in enforced idleness, despite real poverty and need in the society to which they belonged. Somehow the economic machine had broken down, or at least was operating at something well below its proper capacity. Why?

One answer is that although the factors of production would be completely mobile in the theoretical world of perfect competition, in the actual world the mobility of resources is far from perfect. Because of inertia, or limited knowledge, or any one of a hundred other causes, labour and capital no longer needed in a declining industry may not be taken up immediately by industries whose output is expanding—geographical factors may hinder movement, time may be required for adaptation to the specialised requirements of the receiving industries, or the owners of factors may hope that the decline in the industry previously employing them is merely temporary. Imperfections of factor mobility, therefore, are bound to result in *some* unemployment of factors in a dynamic

economy, and it must be regarded as an unfortunate growing-pain which has to be borne in order to secure the benefits of improvements in technology.

Considerations of this kind can explain the unemployment of factors in a particular industry or limited group of industries. But at intervals in modern history there has occurred widespread and severe unemployment throughout the economic system as a whole, and a shift in demand from one product to another obviously cannot explain a situation of this sort. A redistribution of demand may explain temporary unemployment in a particular sector; but a drop in *aggregate* or total demand must underly a situation of general unemployment. The analysis of price determination has shown that an entrepreneur will employ factors up to that point at which his marginal costs—the rewards to the various factors—are just equal to marginal revenue. Unless there is a sudden and general rise in factor-prices, he will reduce his demand for land, labour and capital only if marginal revenue should fall—i.e. only if the demand curve for his product should shift to the left.

This familiar conclusion can be generalised and applied to the whole economy, Entrepreneurs as a whole pay out in the form of wages, rent and interest the incomes of the factors they employ, receiving also profits as their own incomes. With a constant flow of total incomes there may occur a redistribution of factors as between one entrepreneur and another, but this will not affect the overall level of employment. These factor-costs—the sum of the incomes received in the community, or the national income—must be equalled by the total revenue derived by entrepreneurs from the sale of their products. The price of a commodity must necessarily equal the incomes—including profits or losses—derived from its production. The total revenue received by entrepreneurs, of course, is the same thing as the total expenditure on goods and services ('aggregate demand') by the community as a whole. Aggregate demand therefore equals the national income; general unemployment of factors can occur only if aggregate demand declines, causing the marginal revenue of entrepreneurs to fall below marginal cost, and thus forcing entrepreneurs to dispense with factors of production and reduce output. A redistribution of demand between commodities, taken in conjunction with imperfect factor

mobility, will explain unemployment in particular industries; only a fall in *aggregate* demand can explain unemployment over a wide range of industries simultaneously.

Aggregate demand, then, determines how much is available in any period for the payment of factors of production. This is merely another way of expressing the identity referred to in Chapter 19—national expenditure (aggregate demand) is necessarily the same as national income (total payments to factors). In the last resort, every penny of every income is derived from the proceeds of production. If the total expenditure of a community rises, its total income will rise, either through the employment of more factors of production at current factor-prices (wages, rent etc.) or, if all factors of production are already in use, through a rise in factor-incomes. The existence of international trade introduces certain complications which will be discussed in Part VIII, but these do not affect this fundamental truth that the total amounts paid out as factor-incomes, and therefore the level of demand for factors, is dependent on the aggregate demand for goods and services.

2. THE COMPONENTS OF AGGREGATE DEMAND

The total amount being spent on goods and services over any given period plays a vital role, then, in determining the sums which entrepreneurs are prepared to offer for the services of factors of production. Because of this, it is desirable to separate aggregate demand into its constituent parts in order to see more clearly the possible causes of the fluctuations in demand which can have significant effects on the level of employment and of factor-prices.

It is usual to divide aggregate demand into three components—consumers' expenditure, investment expenditure and government expenditure—since each of these broad groups is subject to rather special influences. Leaving the question of exports and imports aside for the moment, every item of expenditure taking place in an economy will fall under one or other of these three headings. The everyday purchases in the shops of food, clothing, tobacco, furniture etc. and payments for services such as entertainments, laundries and travel, constitute *consumers' expenditure*; the heading covers, in other words, all the spending which people do in their

capacity of private citizens. By *investment expenditure*, it will be remembered, is meant the construction of producers' goods such as factories, shipyards and machines. The purchase of securities, or existing assets such as houses, is sometimes spoken of as investment in ordinary conversation; the economist, however, studying the overall economic activity of society, excludes such transactions from investment because they are merely transfers from one person to another, and therefore cancel out for the economy as a whole. *Government expenditure* refers to all those purchases of goods and services by government agencies on behalf of the community, such as the provision of health and educational services, expenditures on personnel and equipment required for defence and the payment of national and local civil servants.

These three items, then, make up total domestic expenditure which, together with the net revenue from exports, represents the flow from which factor-incomes are paid. For the United Kingdom in 1955 the orders of magnitude (at market prices) were:

Gross national product	=	*Consumers' expenditure*	+ *Investment*	+ *Government expenditure*	+	*Net Exports*
£19,300 mn.	=	£12,800 mn.	+ £3,200 mn.	+ £3,200 mn.	+	£100 mn.

Of the third type of expenditure—i.e. that on goods and services by the government—little need be said here. Until recent years it has been an element in the total which was determined by political rather than economic factors; its magnitude depended on the extent of government activity, the requirements of defence, the size of the civil service etc., while its movements from year to year have been determined by current political policy. More recently, however, the potentialities of government expenditure as an element which can compensate for fluctuations in private expenditure have been increasingly realised, and more will be said on this subject in Chapter 27. For the moment it is best regarded as an element in total expenditure which varies according to political rather than economic motives, and which, furthermore, is relatively stable from year to year. The question of exports and imports can also be left on one side for the moment. Attention can therefore be concentrated on the two main constituents of aggregate demand which

are wholly economic in nature—consumers' expenditure and investment. Could movements in one or both of these account for the fluctuations in the level of employment and factor-incomes which are an important feature of modern economic life?

3. CONSUMPTION BEHAVIOUR

A great deal of this book has been devoted to the analysis of the way in which a consumer will distribute a given amount of income between all the various goods and services available to him. In considering consumer expenditure as an element in aggregate demand, however, the economist is faced with the problem of the factors determining the *total* amount which people spend, rather than the distribution of that total amongst different commodities.

The first and most obvious factor is the income received by a person. Most people would spend more than they do now if they had the opportunity, but their total income is a limiting factor. If a person's income rises, his purchases of goods and services will almost always increase; if his income falls, he is forced to spend less. A second factor, however, is the thriftiness of consumers; this is generally referred to as the *propensity to save* and expressed in the form of the fraction of income which is saved. The determinants of the propensity to save are numerous: examples are the attitude to future as opposed to present needs, the desire to accumulate resources to meet possible emergencies, the attraction of the interest to be earned on funds used to buy securities and the wish to provide for dependants.

One especially important determinant of the propensity to save is the amount of a person's income. When people are so poor that they can afford only the barest necessities of life, saving is impossible; all the wealth available to them is needed to preserve life itself. As income increases, however, the more urgent wants can be satisfied; some wealth can be used in the enjoyment of commodities which make life more comfortable but which are not absolutely essential. It becomes at least *possible* for a person to save. As the standard of living rises further, the urgency of wants still unsatisfied becomes less and less, and consequently less sacrifice is involved in saving. If a person is very rich indeed, it may be difficult for him to avoid saving at a fairly high rate; the number of

meals which a person can eat, or suits he can wear, or houses he can live in, is not easily pushed beyond a certain point. As a general conclusion, then, it can be said that as income rises the propensity to save (i.e. the fraction of income which is saved) will also rise. Put alternatively, the *marginal propensity to consume* (the proportion of increments of income devoted to the purchase of consumer goods) will fall as income rises; as a person becomes richer, a smaller and smaller fraction of each addition to his income will be spent. Like all generalisations in economics, this is at best the statement of a tendency amongst a large group of people; it is liable to be falsified in individual cases, or by changes in conventional standards, the appearance of new products and a good many other causes.

It can be said, then, that so long as the level of incomes is stable consumer expenditure is unlikely to fluctuate significantly. Human beings are largely creatures of habit and become accustomed to certain standards and patterns of consumption. Although the propensity to save may be substantially changed over a short period of time by special causes (such as the savings campaigns waged during wartime, or, conversely, the sudden appearance of television in the post-war years), these changes will be exceptional. If incomes remain constant, the level of consumer expenditure will generally remain constant. The analysis must look elsewhere, therefore, for the most likely cause of independent fluctuations in aggregate demand which result in the unemployment of resources; a change in consumer expenditure will be an *effect* of such fluctuations (since the unemployed will be forced to reduce their expenditure) but is unlikely to be a *cause* of them.

The preceding discussion has been conducted in terms of personal saving. It must be remembered, however, that *business* saving (i.e. the 'ploughing back' of profits instead of their distribution to shareholders) is at least as important an element of total saving in a modern economy as that of persons. Much the same sort of conclusions apply, however. Companies accumulate profits primarily to build up a reserve to meet possible emergencies in the future—to tide over bad times, to meet unexpected outlays, and so on—or to finance the further extension of the business. The practices and policies of companies in these respects are unlikely to change

suddenly or frequently, so that, provided companies are earning a fairly stable level of profits, their rate of saving is unlikely to show any great variations over a period of a year or two. Business saving, that is, will fluctuate primarily as a *result* of changes in the level of prices and profits; it is unlikely to be an independent *cause* of such changes.

4. THE DETERMINANTS OF INVESTMENT

An analysis of the factors guiding entrepreneurs in their decisions concerning investment shows that this component in aggregate demand will have no stabilising element comparable with the influence of habit and convention on consumer behaviour. As in all other productive activity, an entrepreneur will purchase new capital equipment only if the revenue he expects to derive from it is at least equal to its cost of production. A sudden change in *either* the expected profitability of an investment project, *or* in its cost to the entrepreneur can therefore result in the abandonment of a project which was previously attractive or, conversely, in a decision to go ahead with an investment scheme which had previously been considered to be uneconomic. How could such sudden changes occur?

On the cost side, the main items are the cost of purchasing or constructing the asset itself and the price paid for the funds locked up in it. The former item, the physical costs of production, are no more likely to vary in the short run than are those of any other commodity (although oscillations in the demand for capital goods may sometimes have substantial effects on their prices). More important, however, are the changes that are possible in the price of funds—i.e. the rate of interest. If an entrepreneur purchases a machine costing £1,000 and raises the funds by borrowing, he will naturally have to pay the lender the current rate of interest as a reward for sacrificing the liquidity represented by £1,000 in cash. Even if the entrepreneur uses his own funds, he still has to reckon the current rate of interest as a part of the cost; he *could* obtain a return by using the £1,000 to purchase securities rather than a machine, and this loss of interest is part of the price he pays for the machine. As was seen in Chapter 24, significant changes do occur in interest rates during short periods, and indeed are frequently

brought about deliberately by the monetary authorities. Fluctuations in investment may well occur, therefore, because of sudden variations in its cost.

Even greater instability can be expected in the other major determinant of investment decisions—the expected revenue attributable to any given investment project, usually referred to as the *marginal efficiency of capital.* A factory or machine will go on yielding revenue, via the goods produced with its aid, for a number of years. An entrepreneur has to look a considerable distance into the future, therefore, in assessing the extra receipts which an addition to his capital equipment may bring in. Only rarely will capital goods bring in enough revenue to cover their cost within less than two or three years; many 'heavy' capital goods, such as generating stations or steel mills, may have to operate for twenty or thirty years before the revenue they earn is sufficient to cover their cost. Now a great many things can happen even in two or three years to falsify the expectations of entrepreneurs; demand may change so as to ruin the market for the product concerned, new competitors may unexpectedly enter the market, while technological developments may render obsolete overnight what currently appears to be the most modern machinery. Entrepreneurs' expectations are therefore necessarily rather vague and highly speculative. One event (a collapse of a competitor, a sudden new vogue, a temporary dip in prices etc.) may cause a widespread revision of expectations involving new investment; changes in prevailing business views will tend to be contagious, since views of the future are largely coloured by the events of the present, and a wave of optimism or pessimism can spread rapidly throughout the business world. Above all, any increase in uncertainty about the future course of trade, or politics, or tastes, will inevitably have a dampening effect on investment decisions; an investment project once undertaken cannot be lightly abandoned, and an entrepreneur will stand to lose heavily during many years because of a single error of judgement.

On both the cost and revenue sides, then, the investment process contains a good many unstable elements; consequently, significant changes in the rate of investment expenditure can occur because of a number of possible causes. An example may make this point clearer. Suppose that an entrepreneur is considering the installa-

tion of a machine costing £200; he believes that it will have a working life of five years and that (after allowing for the costs of the labour and raw material used in conjunction with it) it will add £60 a year to his total revenue. Over its working life, then, it will earn £300; of this, £200 will need to be set aside to cover its initial cost, leaving £100 as the gross profit accruing on the investment. On an outlay of £200, this represents an annual rate of roughly 10 per cent. If the current rate of interest on borrowed money is 5 per cent the entrepreneur will make a net profit of 5 per cent per annum, and the proposition will be an attractive one. Any one of a number of factors can transform it into an unattractive one, however. A rise in the rate of interest to 10 per cent or above will wipe out the entire net profit, while a lesser rise will significantly reduce it. Similarly, if the entrepreneur was forced to revise his estimate of the probable working life of the machine from five to four years, because of the possibility of some technological development, the expected revenue from the machine would be reduced to £240. After allowing for the cost of the machine, the gross profit of £40 earned during the machine's working life of four years would amount to an annual rate of only about 5 per cent, and this would be insufficient to do more than meet the interest charges of 5 per cent, leaving the entrepreneur no net profit. Or again, if the price of the machine's product were to fall by 20 per cent, the total revenue expected over five years would similarly be reduced from £300 to £240, once again leaving insufficient to pay interest charges and repay the capital cost.

A great many factors, therefore, can materially influence investment decisions, and a relatively small change in any one of them can have a marked effect on the total level of investment expenditure. The significance of these factors making for instability of investment should not be exaggerated. Much investment is of a 'social' nature—houses, roads, schools etc.—and is nowadays carried out by the central or local government; it is therefore not subject to oscillations resulting from increased or diminished profit expectations. Even within the broad range of commercial investment projects, many are of a relatively routine nature and are not much affected by short-term movements in profits or interest rates. Nevertheless, at any one time there are inevitably more

'marginal' projects which can be brought within, or excluded from, the range of profitability by sudden changes in the economic and political environment. And for reasons discussed in the next chapter, even a relatively small initial change in the level of investment expenditure can have widespread and cumulative effects on the economy as a whole.

5. SAVINGS AND INVESTMENT

In the two preceding sections the factors governing the expenditure of consumers and of entrepreneurs have been discussed. By and large, the two sets of factors are entirely different and unrelated; the decisions of consumers in allocating income as between spending and saving will have no direct connection with those of entrepreneurs in deciding the rate of investment expenditure. Despite this complete separation of saving decisions on the part of ordinary people and investment decisions on the part of entrepreneurs, however, there is an identity of a fundamental and inescapable kind between the total *amount* saved in an economy during any period and the *amount* invested (in its economic sense of expenditure on the construction of capital goods) during the same period. In fact, the amount saved and the amount invested will always be equal.

This may seem a rather remarkable phenomenon—the entirely unrelated and disconnected decisions of two separate sets of people leading to identical results. On further consideration of the nature of the investing process, however, the apparent strangeness disappears. In an economy, the act of adding to wealth consists of using a part of current output for some purpose other than consumption. That is what investment *means*—the setting aside of resources from current consumption. Now saving is exactly the same process looked at from the other side, so to speak. If a pound's worth of resources is devoted to adding to a society's stock of capital goods, someone, somewhere, must *own* the result; nothing can float about the economic system belonging to no-one. That someone, whoever or whatever he is, has used wealth for the purpose of capital formation rather than for current consumption. And that is what *saving* means. The amounts that people save—in whatever form they may happen to hold their saving—are

nothing more nor less than *the monetary equivalents of the growth in the value of the stock of capital goods.* When people save they refrain from using all their income for consumption; but all income originates in current production, so that the income not consumed must be offset somewhere by *production* which has not been consumed—i.e. which has been used to add to the stock of capital. Investment cannot occur without saving, nor saving without investment, any more than selling can occur without buying or lending without borrowing; they are the two sides of a single event in the economic world.

Some elementary algebra may clarify this vitally important fact. Let Y represent the total national income, C total consumers expenditure, I total investment expenditure and S total savings. (Government expenditure and foreign trade are being ignored for the moment for the reasons given earlier.) All income must arise from production, so that the value of the national income must be equal to the value of all goods produced—consumer goods and capital goods. That is,

$$Y=C+I \quad \text{.....(1)}$$

Again, all income must either be consumed (spent on consumer goods) or saved; there is no other alternative. Hence

$$Y=C+S \quad \text{.....(2)}$$

Since the value of the national income is the same in both equations (1) and (2), the expressions on the right-hand side of the equations must also be equal to one another—they both equal the same thing (Y) and must therefore equal one another. Hence it follows that

$$C+I=C+S \quad \text{.....(3)}$$

Since consumers' expenditure (C) appears on both sides of equation (3), it can be subtracted from each side, leaving the statement that

$$I=S$$

This merely puts in symbolic language the identity discussed in the previous paragraph. Goods produced but not consumed constitute capital formation, or investment; income received but not consumed constitutes saving. One is no more than the reflection of the other.

It is clear that the amount saved *must* equal the amount invested; what are not obvious are the implications of this equality when the relevant decisions are taken by entirely different groups of people, savers and entrepreneurs. These are only to be seen through an examination of the process of income-formation to which the discussion will turn in the next chapter.

FURTHER READING

J. M. KEYNES, *The General Theory of Employment, Interest and Money*, Ch. 3, 6–12.
A. H. HANSEN, *A Guide to Keynes*, Chs. 2–3, 5.
D. DILLARD, *The Economics of J. M. Keynes*, Chs. 4 and 7.
J. ROBINSON, *An Introduction to the Theory of Employment* (Macmillan, London, 1937), Chs. I–II.

CHAPTER 26

The Process of Income Formation

I. THE NATURE OF THE INCOME FLOW

Demand must always equal supply, but the analysis of Part IV has shown that this statement is true only because a particular mechanism exists which ensures that the amount of a given commodity taken off the market by consumers is always equal to the amount put on it by producers. That mechanism is price. If prospective demand tended to exceed prospective supply, price would rise, cutting down the demand and stimulating supply. However great prospective demand might be, there is always *some* price which will be high enough to reduce it to any given level; however large prospective supply might be, there is always *some* price which will be low enough to reduce it to any given level. In any market, at any time, actual demand *must* equal supply, since for every act of sale (supply) there is a corresponding act of purchase (demand).

Now savings and investment, defined in the way outlined in the previous chapter, must necessarily always be equal; there is no means by which one can differ from the other in the way that demand and supply at a given market price may be different. Hence there is no mechanism corresponding to price which *brings* saving into equality with investment. There *is* a parallel with supply and demand, however, in that the identity of savings and investment implies repercussions on the level of *income* when the decisions underlying them change, just as repercussions are felt on market price when the supply of or demand for a single commodity changes.

To understand the nature of these repercussions, it is necessary to appreciate the interdependence of all the members of an economic system so far as their incomes are concerned. Although

each individual might think that the amount he earns has no connection whatever with the amount earned by people a hundred miles away, he would deceive himself in so thinking. It is obvious that the income of a shopkeeper is closely linked with the incomes of the people living in his neighbourhood; they are his customers, and if their wages go up his trade will be brisk and his profits will consequently be raised. What is less obvious is that the wages of the customers are in turn intimately linked with the profits of the shopkeeper; when the shop prospers and profits are good the shopkeeper will buy more of the clothes, or electric irons, or books, or chocolates which his customers help to make in their factories. The coal miner may not see any connection between his wages and the earnings of a country doctor, but it is there all the same; if the doctor has a good year he may buy a new car, which will help to stimulate the motor-manufacturer's demand for steel, which in turn will stimulate the demand for coal, and hence for coal-miners, which may tend to increase miners wages.

What this all comes to, of course, is that the economic system is one enormous circular flow of income, with income passing from hand to hand in exchange for goods and services. *Everyone's income is someone else's expenditure.* When I buy petrol for my car the transaction represents expenditure to me but income to the garage proprietor; when the garage proprietor buys a new house it is expenditure to him but income for the builder; when the builder buys my text-book on economics for his son, it is expenditure to him but income for my publishers and me. Economically, people are utterly and hopelessly dependent on one another; by altering his expenditure a person will thereby alter the income of someone else, and by changing his expenditure someone quite unknown to me may be influencing my income.

Bearing this in mind, the significance of saving and investment can be recognised. Saving is, by definition, the act of not-spending; a person who saves simply withholds part of his income from expenditure on goods and services. (He may place the funds so saved at the disposal of someone else, of course, but that is another part of the story altogether; in any case, he may not part with his savings at all.) What effect will this have on the circular flow of income? Here is some income which has come into the hands of a

person and which he does not pass on to someone else through the purchase of goods. The flow of income is therefore diminished to that extent, just as a leak in a water-pipe will diminish the pressure of water flowing through the rest of the system. If the saving is achieved through a reduction in the consumption of (say) magazines, then the income of the publishers, and all those dependent on them, is reduced to the same extent. Nor is this the end of the affair. The magazine publishers, being somewhat poorer than they would have been if this saving had not taken place, are in turn forced to spend less on their own consumption of (say) cigars, so that the incomes of cigar-manufacturers are in turn diminished. So it goes on, the act of saving reverberating throughout the whole economic system in the form of reduced incomes.

An act of investment has precisely reverse effects. If a company decides to build a new factory, using either borrowed funds or their own accumulated reserves, it *adds* to the flow of income. Men are engaged in constructional work and are paid wages; raw materials or machines are purchased, adding to the revenues of their producers, and so on. These extra incomes will result in extra spending by their recipients, so that the demand for clothes, or entertainments, or food, or whatever else the incomes are spent on, is stimulated and the revenues of their producers increased. The latter, in turn, will be encouraged by the increased demand to expand production, so increasing incomes through *their* extra expenditure on wages, raw materials etc. Once again, then, there is a chain reaction through the economy. Investment directly raises incomes through the rewards paid to factors producing the capital-goods involved, this increased income passing out again through increased expenditure.

The circular flow of income can be conceived of, then, as the passage of income from hand to hand through the production and sale of ordinary consumer goods and services. This basic income flow is *reduced* by savings, which comprise income received but not passed on again, and *increased* by investment, which represents the creation of incomes by the process of adding to the stock of capital goods. Two important questions about this income-flow remain to be answered. First, what limits, if any, are there to the changes in income which occur as a result of new acts of saving or investment?

Do the repercussions go on for ever, or will they tend to die out? Secondly, why and how is the level of income affected by the equality of saving and investment? These questions will be considered in turn.

2. THE MULTIPLIER

When the company embarked on a new factory in the example just given, incomes were directly generated; for the sake of simplicity only the wages of the men engaged on building the factory need be referred to, although obviously other types of income will be involved. These wages found their way into the tills of shopkeepers selling food, or clothes, or tobacco, and thence passed to the producers of those commodities. These in turn spent more on the goods *they* wanted, and so the sequence continued. Does it go on for ever? Will an initial injection into the income flow set in motion an unending rise in incomes?

It will not, and for the reason given in the discussion of saving. If every penny of everyone's income was spent, the chain reaction *would* go on for ever, an initial injection of income passing from hand to hand *ad infinitum*. But every penny of every income is *not* spent. People save, and, as has been seen, the effect of saving is to diminish the flow of income. If people on average save one-tenth of their income, an initial injection of £100 taking the form of additional wages of factory-builders will result in an extra £90 spent on, say, clothing; the extra £90 received by the producers of clothing will result in an extra £81 expenditure, and so on. The wave of income will spread further and further into the economy, but with diminishing force. In this example the actual sequence will be something like this:

$$100 + 90 + 81 + 73 + 66 + 59 + 53 + 48 + 43 + 39 \ldots$$

(Initial injection)

Clearly the series will eventually tail away into nothingness. Can the situation be described more precisely than this? Is it possible to say how much extra income will be generated by any initial increase in spending? It is indeed possible, because it is known that

the brake at work slowing down the rise in income is the propensity to save. If the proportion of any additional income people will put away in saving is known, it can be calculated how rapidly the extra income will vanish through the savings leak. In the example, it is known that people save one-tenth of their income; it follows that total incomes must rise by £1,000 before people's equilibrium, or 'normal', level of savings will have risen by £100. If there is an initial increase of £100 in incomes, then, this increase will be exactly offset by increased 'normal' saving only when total incomes have risen by £1,000. In the sequence given above, only the first ten items in the series of income changes originating in the initial injection of £100 are given. These actually add up to £652. If the *next* ten items were calculated and added, the total would come to about £868, if a further ten were calculated, it would be about £943. If the series were continually extended in this way, the cumulative total of the items would approach closer and closer to the sum of £1,000.

This does not mean, of course, that the total volume of savings in the economy will rise by £100 only after this process (which naturally takes time) has come to an end. Immediately investment rises by £100, saving must increase equally. The day the factory is built, its constructors are paid £100, and their income is that much greater. Until they, in turn, spend this extra income, their savings must be £100 greater—saving being the difference between income (which has risen by £100) and consumption (which is as yet unchanged). *After* they have spent part, or all, of the £100 it becomes someone else's income and (until it is passed on again through extra expenditure) forms part of someone else's saving. At any moment of time the extra saving must be equal to the extra investment. For example, suppose that the numerical sequence shown above is stopped immediately the third group have received their extra income, but before the fourth group have yet been affected (i.e. before the third group have had time to spend their additional income). The position will then be:

Additional incomes received	=	100	90	81
Additional expenditure	=	90	81	0
Hence additional saving	=	10	+ 9	+ 81 = 100.

Similarly, if the sequence were stopped just after the sixth group had received their extra incomes, the position would be:

Additional incomes received	=	100	90	81	73	66	59
Additional expenditure	=	90	81	73	66	59	0
Hence additional saving	=	10	+ 9	+ 8	+ 7	+ 7	+ 59 = 100.

But the vital factor in these cases is that the saving being performed by the final group in the sequences is not their *equilibrium* rate of saving—it is not the saving they will perform after they have had sufficient time to adjust their expenditure to their new level of income. Hence the process of change in the overall level of incomes will continue; it will end only when *everyone* (including the final recipients in the chain of income changes) is saving and spending at his normal rate from his new level of income. Savings and investment are equal throughout; but the movement of incomes upwards or downwards, as the case may be, only comes to an end when everyone has had time to reach his *equilibrium* rate of saving (i.e. the product of his new income and his normal propensity to save).

This can be put more generally. The sequence of items in the process of income formation constitute a geometric series, since each item is found by multiplying the preceding item by a constant factor. (In this context, the constant factor is the marginal propensity to consume, which is usually represented by the letter c.) From any text-book on algebra it can be found that the sum of any such series having a very large number of terms is

$$A\frac{1}{1-r},$$

where A is the first term in the series, and r the ratio by which each term is multiplied. Applying this formula to the income flow, it follows that the total change in incomes arising from an initial change in investment (I') will be

$$I'\frac{1}{1-c}.$$

In the numerical example quoted above, for example, I' was £100 and c was $\frac{9}{10}$. Hence the formula indicates that the total change in incomes will be

$$£100 \times \frac{1}{1-\frac{9}{10}} = £1{,}000.$$

This expression, $\frac{1}{1-c}$, which relates an initial change in expenditure to the total change in the general level of income which will result from it, is known as the *multiplier*.

The multiplier is clearly a most useful and important concept. Given a knowledge of the fraction of additional incomes which the public as a whole will devote to increasing its consumption, the ultimate effects on total incomes of any postulated change in expenditure can be estimated—for example, if it is known that the propensity to consume is $\frac{3}{4}$, a rise of £100 million in, say, defence expenditure could be expected to result, other things remaining equal, in a rise of £400 million in the overall level of incomes. At the same time, it is important to remember that in real life time-lags are liable to occur between a change in income and the full adjustment to expenditure which it will eventually cause. If my income rises today, it may be weeks or even months before I pass on the full effects of this to the rest of the economy in the form of increased expenditure. The propensity to consume in the short run, in other words, may differ substantially from that of the long run when everything has had time to work itself out. A multiplier based on long-run spending or saving propensities, therefore, can only indicate the changes in total income which will probably result from an initial change in spending after a lapse of time sufficient to allow for the full adjustment of people's expenditure to changes in their income. The change in total incomes which will occur in, say, the next month or two months may be considerably smaller than the ultimate long-run effect suggested by the multiplier.

3. INVESTMENT FLUCTUATIONS AND INCOME

It is now possible to analyse the implications of the necessary equality of saving and investment despite fluctuations in investment expenditure or in the thriftiness of the general public. Consider a decline of £50 million in investment expenditure in an economy where the marginal propensity to consume is $\frac{3}{4}$. The initial impact of the decline will be a loss of £50 million income on the part of those who were previously engaged directly or indirectly on the investment projects. This fall in income will in due course be passed on in the form of reduced expenditure; since the people

concerned would have saved, on average, one-quarter of this £50 million, however, their expenditure will not be reduced by the full £50 million but only by that fraction of it which they would have spent—i.e. about £37·5 million. This reduction in expenditure will in turn cause other incomes to be reduced, causing a further fall in expenditure, the process continuing in the familiar sequence. The series of repercussions (changes of income in £ millions) will be:

$$-50 - 37{\cdot}5 - 28{\cdot}1 - 21{\cdot}1 - 15{\cdot}8 \text{ etc.}$$

From the multiplier formula it is known that the total ultimate change in incomes (Y') will be

$$Y' \frac{1}{1-c}$$

$$= -50 \frac{1}{1-\frac{3}{4}}$$

$$= -50 \times \frac{1}{\frac{1}{4}}$$

$$= -200.$$

When the process has worked itself out, then, the overall level of incomes will be £200 million lower than before as a result of the fall of £50 million in investment expenditure.

Naturally the equality of saving and investment has not been affected by all this. If people save on average one-quarter of marginal increments to their income, it follows that when incomes fall £200 million total saving must fall £50 million—i.e. by exactly the same amount as the change in investment. If saving and investment were equal before the change, then, they must be equal during and after it. The level of incomes has altered *because* of this equality. A fall in the rate of investment has made everyone sufficiently poorer in money terms to ensure that total savings in the new equilibrium situation are reduced by exactly the same amount. Conversely, if a rise occurs in the rate of investment, extra incomes will be generated, so enabling people to increase their equilibrium rate of saving to exactly the same extent. Investment will always generate the savings which, as was noted earlier, are the monetary counterpart of investment. The whole process of adjustment, of

course, will take time, but the multiplier formula can indicate the equilibrium situation towards which the economy will move.

Given the rate of investment and the average propensity to consume, indeed, it is possible to calculate what the aggregate (national) income of the community will be, since it has been established that the national income must be such as to produce (with the given propensities to spend and save) an amount of saving equal to the rate of investment. (The *average*, rather than the *marginal*, propensity to consume is now involved, of course, since the analysis is concerned with the *total* level of income and not *changes* in the level of income.) This can most easily be seen in terms of the symbols used previously. Ignoring government expenditure and foreign trade, total income must be equal to the sum of the values of production of capital goods and consumer goods; i.e.

$$Y=I+C \text{ and } Y-C=I \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(7)$$

It is also known that total consumer expenditure must be equal to the product of income and the average propensity to consume (C'), i.e. $C=C'(Y)$.

Substituting this expression for C in (7) gives

$$Y-C'(Y)=I$$

hence

$$Y=I \cdot \frac{1}{1-C'}.$$

If investment expenditure equals £2,000 million and the average propensity to consume is $\frac{3}{4}$, then the total national income must be

$$\text{£2,000 million} \times \frac{1}{1-\frac{3}{4}} = \text{£8,000 million.}$$

If people consume three-quarters of their income they must save one-quarter—i.e. a total of £2,000 million. Saving must equal investment.

The important point about all this, of course, is not so much the equality between saving and investment as the fluctuations in the national income which must occur as a result of this equality, assuming that variations occur in the rate of investment. If the marginal propensity to consume is a relatively large fraction, as it usually is, the value of the multiplier will be high, implying that fairly small changes in the rate of investment will have a substantial impact on the total national income. If the marginal propensity

to consume is $\frac{95}{100}$, for example, the multiplier becomes 20; a change of £1 million in the rate of investment will therefore cause a change of £20 million in total incomes. Large fluctuations in income imply large fluctuations in demand and consequently in the amounts of factors of production which entrepreneurs will employ in order to meet that demand.

4. INCOME AND THE PROPENSITY TO SAVE

It was noted earlier that, in general, human beings tend to acquire fairly firm habits in the matter of saving and spending from a constant income, so that there are unlikely to be sudden and substantial changes in the propensity to consume. Nevertheless, such changes are not impossible. The growth of a 'war scare', for example, frequently leads to a rush on the part of consumers to buy up a stock of commodities which they anticipate will soon be in short supply, causing their current rate of savings to drop violently; an expectation of a fall in prices in the near future may have the opposite effect, since consumers will postpone expenditure as much as possible in order to obtain the benefit of the anticipated price reductions. Similarly, business saving may occasionally undergo sudden and substantial fluctuations; companies may change their views about the future availability of credit, for example, and place more (or less) of their current profits to reserve, while a change in the rate of tax on distributed or undistributed profits may have similar effects.

A change in the propensity to save will have effects on the national income comparable to those of a change in the rate of investment, but in the reverse direction. The level of incomes depends on (*a*) the rate of investment, and (*b*) the multiplier; a change in the propensity to save will cause the value of the multiplier to alter. If the rate of investment in an economy is held constant at £200 million, a rise in the propensity to save from one-tenth to one-fifth will cause the multiplier to fall from 10 to 5; hence the overall level of income will fall from £2,000 million to £1,000 million. Precisely similar effects on the level of incomes will follow from a rise in the propensity to save, therefore, as from a fall in the rate of investment.

Will a rise in the propensity to save disturb the equality of

saving and investment? Obviously it cannot. In the example just given, the society saved one-tenth of its income when the national income was £2,000 million and one-fifth when the national income was £1,000 million. The *total amount* saved was the same in both cases—£200 million. In other words, there is no direct and immediate connection between the *propensity* to save and the total *amount* saved. The former is determined by the psychological factors discussed earlier, such as the prevailing attitude towards the future and the number of dependants; the latter is determined entirely by the rate of investment. The role of the propensity to save in the economic system is not to govern the total volume of savings but to determine the level of income associated with any given rate of investment.

Hence the important fact is established that, although as individuals people can change their rate of saving whenever they wish, the community as a whole cannot change its saving unless it varies the rate of investment. For society as a whole, indeed, saving can *only* take the form of an increase in the stock of capital goods; an increase in the stock of monetary claims of one form or another adds nothing to the total wealth of society. An individual can increase his saving by adding to his holdings of notes or bank deposits, but he does so only by reducing the incomes (and therefore the saving) of the producers of the goods he *would* have bought had he not decided to save more.

Does this mean that thrift, as a social characteristic, has no economic importance? Since the total volume of saving is fixed by the rate of investment, is it therefore unimportant whether people save a large part of their income or not? Such conclusions would be unsound. Although the volume of savings is determined by investment, the propensity to save plays a crucial role in deciding the level of money incomes in an economy. As the next chapter will show, it often makes a great deal of difference if a certain rate of investment results in a high rather than a low level of total money incomes. Further, the propensity to save, through its influence on the level of incomes, may *indirectly* affect the rate of investment. Other things being equal, a fall in the level of incomes will result in falling demand for consumer goods; this will usually discourage investment in the industries producing them.

The process of income formation, then, is intimately connected with both the rate of investment expenditure and the habits of the individual members of a society with regard to the saving and spending of their incomes. The analysis can now be directed to the problem of how changes in the level of income can influence the level of prices and employment.

FURTHER READING

J. M. KEYNES, *The General Theory of Employment, Interest and Money*, Ch. 10.
A. H. HANSEN, *A Guide to Keynes*, Chs. 4, 9, 11 and 12.
D. DILLARD, *The Economics of J. M. Keynes*, Ch. 5.
J. ROBINSON, *An Introduction to the Theory of Employment*, Chs. III–V.

CHAPTER 27

Fluctuations in Prices and Employment

I. INFLATION AND DEFLATION

The theoretical tool-box assembled in the two preceding chapters enables a fair understanding to be gained of the damaging oscillations in economic activity which have plagued most Western societies since their industrialisation. It has been seen that the demand of entrepreneurs in general for factors of production depends first and foremost on the strength of aggregate demand for their finished products. Second, it was seen that the demand for consumer goods is dependent on the level of incomes which potential consumers are receiving, and that investment expenditure is closely linked with entrepreneurs' profit expectations and therefore with the state of demand for the final goods they are producing. Finally, the level of incomes was seen to be a circular flow from hand to hand, its magnitude at any time being determined by the rate of investment and the marginal propensity to consume.

Now at a much earlier stage it was established that the price of any particular commodity is governed by demand (the amount which consumers were prepared to spend on it) and supply (the quantity currently being produced). What is true of particular commodities is also true of commodities as a whole—i.e. the *general* price level will be determined by the relationship between aggregate demand (the expenditure of all purchasers of capital and consumer goods) and aggregate supply (the physical volume of goods and services currently coming on to the market). If aggregate money demand rises in relation to the physical supply of goods and services available, more income (in money terms) is being spent on given real quantities of goods and services, and normally the only consequence possible is a rise in the prices of goods in general.

Such a rise in the level of money income relative to the current supply of goods is known as *inflation*. The opposite process—a *fall* in the level of money incomes relative to the supply of goods and services—is known as *deflation*.

In general, the word inflation is used, rather loosely, to denote a situation in which the excess pressure of monetary demand is forcing up prices at a rapid rate, while deflation is used to denote a situation in which the general price level is falling because monetary demand has fallen below the current supply of goods and services. This identification of inflation with rising prices and deflation with falling prices is convenient enough for most purposes, but has its dangers. Inflation, for example, *can* occur without a rise in the general price level; money incomes may be rising in relation to output, but the effects may be offset by contemporaneous reductions in costs of production which would otherwise have resulted in *falling* prices. Similarly, the effects of a deflationary fall in money income may be concealed by rises in costs of production, the conflicting pressures on prices offsetting one another so as to result in a stable price level. Rising prices, in other words, are not *solely* caused by inflation; they may be due to changes in costs (higher prices of imported raw materials, for example) rather than the pressure of demand. Nor are falling prices necessarily indicative of a deflationary situation; greater productive efficiency, or lower raw material costs, may reduce production costs so substantially that prices fall despite a rise in money incomes relatively to output. More important, there is a danger of confusing the disease with the symptom. A rise in the general price level is one probable *effect* of inflation, but it is not in itself inflation; inflation is a disturbance of the relationship between money incomes and the volume of supply, and the rise in prices is simply the evidence of that disturbance.

Nevertheless, in most cases inflation is associated with a pronounced rise in the general price-level and deflation with a pronounced fall in prices, and before going on to consider how such movements originate in practice it will be as well to establish why such changes in the general price-level are undesirable. The major disadvantage is that substantial changes in the general level of prices involve a redistribution of income of an indiscriminate nature. Suppose that, at a time of full employment, government

expenditure is increased without a corresponding increase in taxation, so that the overall level of money incomes is raised at a time when the available supply of goods and services (at the initial level of prices) cannot be enlarged. Prices will be forced up under the pressure of this excess demand, and the real value of current wage-rates is naturally reduced. Trade Unions will certainly press for higher wages in order to restore the real income of their members, but the processes of wage negotiation and adjustment inevitably take time. Until wages *are* adjusted, the real income of labour has been reduced; when the adjustment is finally secured the pressure of demand will probably be raised further by the extra spending of wage-earners, so that prices rise even further and the whole process of wage negotiation will have to start all over again.

What is true of wages is even more true of salaries, which are generally even slower to change than wages. The time-lag in adjustment is greatest of all, however, for the other cost-elements, interest and rent, and the incomes of people like pensioners. These incomes are usually fixed in money terms and for considerable periods at a time, so that people in receipt of them find their money income fixed despite the rise in the general price-level; as a result, they sustain a decline in their real income. Given this relative inflexibility of costs, a rise in prices must inevitably result in a disproportionate rise in profits. Hence the effect of an inflationary rise in prices is to cause a redistribution of real income in favour of profit-earners and away from fixed-income receivers such as lenders and landlords. The wage-earner will be able to adjust his income to keep pace with the rise in prices, but there will usually be a time-lag between a rise in prices and the compensating rise in wages, causing a worsening in the position of labour. In a deflationary period, of course, precisely the opposite occurs; people in receipt of fixed incomes benefit by the increased real value of their incomes when prices fall, while profit-receivers suffer because costs of production do not fall as rapidly as prices. The wage-earner, again, is in an intermediate position, with the real value of his wages rising so long as there is a lag in the process of wage adjustment.

Hence, in general, in periods of inflation and deflation there occurs a redistribution of real income between profit-earners and

fixed-income receivers. Whether such a redistribution of income is desirable or not will depend on the relative wealth or poverty of the persons falling in these different groups. Whatever the situation may be in this respect, of course, the redistribution involved is bound to be haphazard and fortuitous, and this is scarcely a defensible characteristic.

Other dangers are involved. A marked rise in the general price level may result in a country's exports being priced out of foreign markets, which would be a serious development for a country like the United Kingdom which is highly dependent on foreign trade. Efficiency also tends to decline in a period of rising prices; entrepreneurs have little incentive to hold down costs firmly if their profits are being boosted by rising prices. Most important of all, a rapid rise in prices may undermine public confidence in the national currency. If prices rise so violently that money held for any length of time loses a substantial part of its real value, ultimately no-one will be willing to accept money in return for goods, and this can only result in the collapse of a specialised economy resting on the basis of monetary exchanges. Such a collapse of the economic system did in fact result from runaway inflation in Germany in the early 1920's and in several European countries (for example Hungary) in the years immediately succeeding the Second World War. Theoretically, all these factors will work in the opposite direction during a period of falling prices. A decline in the level of prices will be particularly dangerous in that no-one will want to *spend* money unless they cannot avoid it, since the real value of money will be rising as prices fall. This holding-back from expenditure will cause demand, employment and incomes to fall even further.

Inflation and deflation, then, insofar as they lead to fluctuations in the general price-level, are in themselves undesirable. Their inherent evils have been overshadowed in the past, however, by the effects of inflationary and deflationary movements on the level of employment, and it is to these that the discussion must now turn.

2. THE HISTORICAL TRADE CYCLE

For something over a century before the outbreak of the Second World War, almost every industrialised economy in the world was

subject to an alternation of inflationary and deflationary periods—a continued succession of booms and slumps. Apart from the intervention of wars, the alternation was sufficiently regular—each slump and each boom being some eight to ten years separated from its predecessor—for the sequence to be given the name of the *trade cycle*, and to be regarded as an inherent characteristic of the capitalistic economies of the Western world. There is some reason for supposing that this phenomenon, as a recurring and systematic pattern, has ceased to be of more than historical interest; at any rate, the regular oscillations have not been experienced in any economy since 1939. The features of the trade cycle are worth consideration, however, since they throw a good deal of light on the relationship between inflation and deflation, on the one hand, and the level of employment on the other.

Although each cycle in economic activity had its own peculiar features, certain broad phases were common to all of them. The first was usually known as the *recovery* phase. Starting from a position in which the level of employment, output and prices was low, an initial stimulus to investment would occur in some sector of the economy. (The analysis will have to return to the question of why this should happen.) This increase in investment would generate new incomes, raising demand and causing further rounds of income rises in the familiar multiplier sequence. As a good deal of unemployment usually existed at this stage of the cycle, there was no reason why the increase in demand should cause any significant rise in prices; by employing more labour and capital, entrepreneurs would be able to increase supply to meet the increased demand. The higher level of demand would further stimulate investment in both stocks of goods and fixed capital, since profit expectations would have been improved. This increased investment would generate a further rise in incomes, stimulating demand, and consequently investment, even more, and so the process would continue in a cumulative, snowball fashion.

Ultimately, however, the unemployed resources would have been absorbed into employment as far as was possible (given the imperfect mobility of factors, it is unlikely that *all* resources would be employed) and as a result further increases in demand could not be met with increased output. At this point the second, or *boom*,

phase of the cycle would begin. If investment continued to increase under the stimulus of buoyant profits, the rise in income and expenditure generated by it could only result in a rise in the general price-level, because increased demand (in money terms) would be pressing on a more or less constant supply (in real terms). In other words, an inflationary situation now developed in which increased investment resulted, not in a higher level of employment and output, but wholly in a rise in prices. This phase was unlikely to continue for very long. On the one hand, the cessation of the rise in the volume of sales would cause a drop in the demand for capital goods to support increasing output (entrepreneurs would not order machines to enlarge their capacity if sales had ceased to expand, or were expanding at a declining rate); on the other hand, the marginal propensity to consume would be falling as the level of incomes rose. Both elements in the income-formation equation (the rate of investment and the multiplier) were therefore ultimately bound to decline, so that the rise in incomes and prices was eventually reversed.

When this occurred, the third, or *recession*, phase of the cycle commenced. Business expectations being of a particularly volatile nature, the first sign of falling prices and profits, which must occur when demand begins to decline, was liable to set in motion a violent wave of pessimism which might border upon panic. In an attempt to cut their losses before prices fell even further, businessmen would drastically reduce the rate of investment, both by cutting down stocks of goods through forced sales on the market and by cancelling any outstanding orders for new capital equipment. This sharp fall in investment, magnified through the multiplier process, would result in a rapid fall in incomes, and also in the fall in prices which the entrepreneurs were seeking to escape. This, of course, merely added to the general panic. Like the boom phase, however, this recession period could not last long. There is a limit to the possible reduction of investment in stocks of goods, even if it is only attained when stocks have fallen to zero. Similarly, the rate of fixed capital formation cannot be reduced indefinitely. Having cancelled outstanding orders, entrepreneurs could reduce their existing capital equipment only by allowing it to wear out with the passage of time; there is nothing to be gained by physically destroy-

ing machines and factories. Also, as incomes fell the propensity to consume would be rising; people would become so poor that their savings would fall to zero, or even become negative as past savings were dipped into in order to meet current expenditure.

Eventually, then, the crisis period of violently falling incomes and prices would come to an end and the last, or *depression*, phase of the cycle began. Although incomes ceased to fall, they would remain at a low level and considerable unemployment of factors might exist; the depressed conditions of the market discouraged investment, and the economy dragged its feet along a floor of poverty and unemployment. Why should this phase have ever come to an end? The answer lies in the factor of obsolescence and wearing-out mentioned earlier. As time went by the existing capital stock depreciated until its volume was eventually adjusted to the prevailing low level of sales and income. After this point the continued wearing-out of capital would reduce the capital-stock *below* the level necessary for even the low current level of output, and some replacement capital would have to be ordered. The rate of investment therefore picked up, and the economy started once again on the recovery phase of the cycle.

Two points are worth noting in connection with this description of the typical trade cycle. First, although the various cycles of history shared in common these four broad phases, they differed widely in their individual patterns; particular factors (new products, political scares, financial crashes etc.) invariably entered into the course of events with the effect of moderating or aggravating the violence or duration of the various phases. Secondly, the trade cycle, it must be repeated, is a phenomenon of economic history, and it would be unwise to apply dogmatically to modern conditions any conclusions drawn from oscillations in the past. Having gained an understanding of the forces at work in past fluctuations, the problem now is to examine the possibilities of controlling those forces so as to maintain the economy at a high rate of activity with a more or less stable price-level.

3. THE PROBLEM OF FULL EMPLOYMENT

Since 1939 most Western economies have enjoyed an almost unbroken period of full employment. This does not necessarily mean

that in those countries no-one has ever been unemployed; there will always be elements of seasonal or frictional unemployment. But in general no more than a small percentage of the available labour force (usually less than 2 per cent in the United Kingdom) has been out of work at any one time. Why is this? As always, the answer must be sought in the two major determinants of the level of incomes—investment and the propensity to consume (which determines the value of the multiplier)—since it has been established that employment depends on the state of aggregate market demand for finished goods, and aggregate demand depends primarily on the level of incomes.

Investment has been high during the war and post-war years for various reasons. During the war, of course, expenditure on defence (which is investment so far as the effect on income is concerned—it generates income without adding to the current supply of consumer goods) was maintained at a high level. After the war, the need for reconstruction expenditure on roads, houses, factories and ships involved a great deal of investment expenditure. Subsequently, increased defence expenditure has been undertaken by many Western economies, while in most countries a high rate of expenditure on social services—hospitals, schools, homes for the aged etc.—has become an accepted feature of economic life. Finally, in Western Europe and most of the non-dollar world, the need to export goods has been an urgent one for reasons which will be discussed shortly; as will also be seen, exports are a form of investment and have an expansionary effect on incomes.

So far as the propensity to consume is concerned, it is probable that the redistribution of incomes in the United Kingdom and elsewhere has much reduced the overall propensity to save amongst private persons. As was noted in Chapter 19, one effect of budgetary policy in Britain has been to transfer income from the relatively rich to the relatively poor. If it is true to say that the propensity to save is higher among the rich than among the poor, it follows that the overall propensity to consume of the community as a whole must be increased by redistributive policies of this kind.

What then are the problems of full employment? Broadly, they fall under two headings; the problem of maintaining full employment in the future, and the problems created by the attainment of

ull employment itself. Both types of problem raise complex and ar-reaching issues which can only be touched upon here.

Under the first heading, the issues resolve themselves into those :oncerned with maintaining a high rate of investment and those nvolving the propensity to consume. Assuming that defence ex-)enditure will not continue at a high level, will future investment :xpenditure be sufficient to generate a full-employment level of ncomes?. Doubts have been felt about this for three main reasons. First, there is the decline of the rate of population growth in the West; as was seen in Chapter 6 a falling population may seriously reduce major forms of investment. Secondly, there exists a possibility that future inventions may require a smaller amount of capital than the commodities or processes they replace; examples often quoted to illustrate this development in the past are the replacement of wireless telegraphy by radio, or railways by road transport. The final source of anxiety is the virtual disappearance of new territories in the world calling for the development of ports, railways, public utilities etc.; these have constituted a tremendously important outlet for capital in the past. Those who do not believe that investment opportunities are likely to dry up would point to the enormous increase in capital usage which has been caused by inventions which were at first sight capital-saving; all the developments of radio, for example, or the growth of the motor-car industry. They draw attention, also, to the continuance of inventions, such as atomic power or automatic transfer machinery, which promise to absorb vast amounts of capital, and to the obvious capital scarcity over the greater part of the globe. The possibility of declining investment opportunity in the future, therefore, remains a matter of considerable dispute.

The fears involving the propensity to consume are based on the hypothesis that, as economies grow richer with the passage of time, people will want to save an increasing proportion of their income. If this is not to result in the fall of incomes which normally follows from an increased propensity to save, an even higher rate of investment will be required than would otherwise be the case, so adding to the difficulties discussed in the previous paragraph. Once again, however, the tendency for the propensity to consume to decline in the long run has been questioned. It can be argued that although

at any one time the rich save at a higher rate than the poor, human beings have a habit of revising upwards their concepts of a necessary standard of living from generation to generation; what one generation considers a luxury, having only a weak claim on income, the next generation considers to be more or less a necessity which must be enjoyed before any question of saving can arise. If this is true, and remains true in the future, it is by no means certain that the propensity to save will increase as economies grow richer.

Turning to the problems raised by the existence of full employment, the basic issue is that of maintaining something approaching stability in the general price level. Some of the evils of inflation have been touched on already: the indiscriminate redistribution of income, the danger to export markets, and the social frictions induced by the constant necessity for the revision of wages and other factor rewards. To these can be added the immobility of factors which is liable to occur when employment is always available in any given industry or locality; this may mitigate against the introduction of new products and processes which are essential to the full development of a healthy economy. Finally, efficiency on the part of both labour and entrepreneurs may not be sufficiently stimulated in an environment in which there is only a remote possibility of the loss of jobs or markets.

In all this, the crucial issue is that of the power given to Trade Unions by the fact of full employment. There is a danger that the demands advanced by them with regard to wages and conditions may tend to be somewhat extravagant; because labour is in short supply, they may believe that there is little risk of bargaining their members into unemployment. Employers, on the other hand, have a weakened inducement to resist; the scarcity of labour makes them wary of losing employees, while the buoyancy of the market enables them to pass on, in the form of higher prices, any increases in costs they may incur. The danger of the wage-price spiral is consequently always present: a rise in wages results in higher prices and the higher prices stimulate renewed demands for higher wages, the general price level rising continuously. It is true that wages and salaries are not the only element in costs, and that stability should be achieved in other types of income; the fact remains, however, that in an economy such as that of the United Kingdom wages and

salaries represent more than 70 per cent of total personal incomes, so that changes in them are bound to play a dominant role in the determination of the total.

Hence the fundamental problem is that of securing relative stability in employment incomes without resorting to the creation of unemployment as a means of moderating the demands of organised labour. In a country possessing a completely free and independent Trade Union system, this is not an easy task; although unions in general appreciate the fact that everyone would gain from the achievement of price stability, it is a long step from this hypothesis to its translation into an act of voluntary restraint by each individual Union concerning the pay of its own members in particular. Many observers have suggested the need for a national wages policy by which agreement would be reached between government and labour each year (in the light of current trends of demand, employment and productivity) as to the total wage increases which could be granted without causing a rise in prices; the Trade Union movement would be left to work out for itself the sharing of this total amongst individual unions. So far, however, schemes of this kind have failed to find favour with the Trade Union world. The only alternative to a policy claiming the free consent of the unions themselves would be one of state control of wages and direction of labour; most people would regard such an alternative as an even greater evil.

4. THE BUDGET AS A STABILISER

So far, government expenditure has been treated as a constant element in aggregate demand, and nothing has been said of the ways in which the state can influence the level of employment and prices. Most democratic governments, however, have now taken to themselves the responsibility for maintaining a high and stable level of employment, and it is necessary to consider the main ways in which governments can attempt to discharge this responsibility.

The first main weapon is that of *monetary policy*. As was seen in Chapter 24, the Central Bank, acting under the guidance of the Treasury or other government authorities, can exert considerable influence over the availability and cost of credit, and this in turn will influence the rate of investment in the manner indicated in

section 4 of Chapter 25. A great deal of controversy has raged over the power of monetary policy; it is frequently argued that profit expectations are often so optimistic or pessimistic that fluctuations in interest rates (a relatively small fraction of total costs) are unlikely to have much effect. How much force there is in this argument is open to dispute, but it is widely accepted that a change in the *availability* of funds, as opposed to their cost, is bound to influence investment decisions. However optimistic an entrepreneur may be, if he cannot get hold of funds to finance his project his optimism will avail him nothing.

The main weapon employed by governments in the field of employment, however, is that of *fiscal policy*—the use of the annual budget to influence the level of activity. The budget may achieve this end in one or all of three ways. First, the government's expenditure on goods and services (including that of local government) —such as that on defence, public administration, health services and education—is a significant element in aggregate demand, and by varying this expenditure the government may be able to offset tendencies in private expenditure. In a period when aggregate demand showed signs of slackening, for example, the deflationary tendency could be prevented by a rise in government purchases of goods and services; conversely, in inflationary periods a reduction in government expenditure could ease the pressure of demand. This process is easier in theory than in practice, however; the greater part of government expenditure is determined by political considerations and is not easily varied from year to year. Defence expenditure, for example, has to be undertaken when it appears necessary rather than when it happens to be convenient from the point of view of aggregate demand; houses and schools and hospitals cannot be shelved indefinitely against the possibility that overall demand may become inadequate at some unknown future date. Similar considerations would apply to the investment expenditures of nationalised industries. In theory at any rate, these could be controlled by the government so as to offset undesirable movements in private investment expenditure, but in practice the timing of capital investment is of necessity dictated largely by the current needs of the industries concerned.

The second element of fiscal policy which can play a stabilising

role is that of taxation. The rate of investment may be influenced to some extent by tax allowances on capital expenditure, or by a differential rate of profits tax favouring undistributed ('ploughed-back') profits in comparison with distributed profits. Consumers' expenditure may be influenced by changes in the rate of taxation on commodities carrying Purchase Tax or Excise duties, or by changes in the treatment of personal savings in income tax assessments.

The final, and perhaps most important, aspect of fiscal policy in this context is that of the overall surplus or deficit attained through the budget. In times of inflationary pressure, a Chancellor of the Exchequer can 'mop up' large amounts of income by adjusting government revenue and expenditure so as to secure a surplus of receipts over expenditures. If on balance the taxpayers of a country are paying over to the government in the form of taxation substantially more than is being paid out by the government through purchases of goods and services or transfer payments, the budget is obviously exerting a deflationary influence on the economy. When inflation or deflation threatens, therefore, the budget can be used in the wider context of overall economic stability, rather than in the traditional, and narrower, context of the mere raising of sufficient revenue to finance government expenditure. In times of inflation, a budget surplus would be desirable, since it will on balance reduce incomes and therefore demand; in periods of deflation, a budget deficit could be aimed at in order to augment the general level of incomes and so stimulate demand.

The modern budget, then, must be framed in a much wider context than was usual before 1939; in drawing it up, the Chancellor must have an eye to the overall economic situation rather than simply the current needs of the government itself. Because of this, the budget cannot be judged on the same basis as that appropriate to the income and expenditure accounts of a private person or company. Private individuals cannot increase their own incomes by spending more than they are currently earning; if pursued for any length of time, indeed, such a course can lead only to disaster. Governments, however, are in a different position; by spending more than they are currently receiving from taxation they can play an important part in raising the overall level of incomes and there-

fore the level of taxation receipts. Similarly, a Chancellor cannot concede tax remissions on the strength of a large budget surplus in the preceding fiscal year if an inflationary tendency still persists in the economy. Each budget must be framed in the light of the situation currently prevailing; sound budgeting nowadays may involve a series of budget surpluses or deficits over a period of several years if the prevailing economic situation demands it. With the aid of the annual estimates of national income and expenditure, the Chancellor's advisers attempt to assess the probable relationship of aggregate demand and aggregate supply during the year ahead. If it seems likely that demand will exceed the forthcoming supply of goods valued at current prices, then a budget surplus should be aimed at so as to remove this 'inflationary gap'; if demand is expected to fall short of aggregate supply at current prices, then a budget deficit should be used to remedy the deficiency.

Armed with all these weapons, can a modern government ensure that substantial unemployment never occurs? The truth is that this still cannot be regarded as anything like a certainty. A severe depression has not threatened the economies of the Western world since modern ideas of the role of fiscal policy in maintaining full employment have developed; the policy weapons have not yet been seriously tested. But clearly there is room for optimism. Modern governments can do a great deal to remove or prevent unemployment and to secure stability in the general level of incomes and prices; more important, they are *aware* of their powers in this respect to a much greater extent than ever previously. One thing is certain: a movement towards inflation or deflation can be checked with relative ease if corrective action is taken at an early stage and before the full cumulative operation of the multiplier process has gained momentum; action on even a heroic scale may be useless, however, if an upward or downward spiral has once gathered force. The essential pre-requisite for a sound monetary or fiscal policy, therefore, is continuing and accurate diagnosis of the current situation, plus the ability and moral courage to adopt corrective measures before the need for them has made itself obvious to the general public. That is a grave and immensely difficult task.

One more point remains. Countries such as the United Kingdom are heavily dependent on international trade for the foodstuffs

and raw materials without which life and industry cannot continue. Hence for such countries full employment is not merely a question of domestic policy; the inherently complex task of securing stability of incomes and employment is complicated even further by the necessity for reconciling internal stability with external solvency. To complete the picture of the functioning of the modern economy, then, attention must be turned to the issues raised by the exchange of goods and services amongst the various nations of the world.

FURTHER READING

J. M. KEYNES, *The General Theory of Employment, Interest and Money*, Chs. 18, 21, 22, 24.

A. H. HANSEN, *A Guide to Keynes*, Chs. 11 and 12.

D. DILLARD, *The Economics of J. M. Keynes*, Chs. 6, 9, 10 and 12.

J. E. MEADE, *Planning and the Price Mechanism* (Allen & Unwin, London, 1948), Ch. II.

Employment Policy (Cmd. 6527, H.M.S.O., London, 1944).

The Economic Implications of Full Employment (Cmd. 9725, H.M.S.O., London, 1956).

CHAPTER 28

Index Numbers

1. THE PROBLEM OF MEASUREMENT

Before concluding this discussion of the theory of the determination of the general level of incomes and prices, it may be helpful to digress a little in order to consider the technique adopted for the measurement of the general price level in practice. The expression 'price level' suggests an average, in some sense, of all individual prices, or at least of all the prices relevant in a particular context. The concept almost always occurs in connection with changes in prices over periods of time, when the discussion is concerned not so much with the price of one commodity relative to that of others as with the broad tendency of all prices in an absolute sense.

It is not difficult to see why it is impossible to compare *all* individual prices existing in an economy at one date with all prices existing at another. Every commodity has a price, and there are hundreds of thousands of commodities entering into exchange in any moderately-developed economy; the proposition becomes even more impracticable when it is remembered that any single commodity may be produced and sold in dozens of qualities, brands and types each having a separate price. The mere task of listing all these prices would be almost impossible; recorded prices would have changed before the full list could be compiled (just as one end of the Forth Bridge is reputed to require re-painting before the painters have worked their way to the other end); it would therefore be impossible to draw up a list of all prices prevailing at a given moment.

Even if such a list could be compiled, however, it would be of little real use. The human mind is incapable of arriving at a general impression of even moderate accuracy from a collection of hundreds of thousands of different prices varying to different degrees and possibly in different directions. Yet for many purposes it is useful, and for some it is essential, that a reliable idea should be obtained

of overall movements in prices as a whole, i.e. of their direction and magnitude. (Such knowledge is especially important in assessing prospects for the level of employment and analysing international economic developments.) The problem is solved well enough for ordinary working purposes by the application, in the particular context of price movements, of the technique of index numbers. The following section will therefore be devoted to the discussion of index numbers in general; afterwards, their application to the particular problem of price measurement will be outlined.

2. THE NATURE OF INDEX NUMBERS

Fundamentally, an *index number* is a means whereby the changes in a large number of related variables can be averaged, and the average change expressed through movements in a single number. To take a simple example, suppose there are five men whose production of different commodities in 1950 and 1955 (in physical terms) was as follows:

		Output	
Man	*Units*	*1950*	*1955*
1	Tons of coal	500	510
2	Bushels of wheat	2,000	1,800
3	Suits of clothes	60	65
4	Cwts. of bricks	800	800
5	Gallons of oil	9,000	9,500

How can the *average* growth of the output of this group over the period 1950–55 be expressed? The index number technique consists of taking as 100 the output of each man in the earlier year (the 'base-year' as it is called) and then calculating the output of each in the later year as a percentage of his output in 1950. The five figures obtained for 1955 can then be added up and the total divided by five to arrive at the average change. Thus, the calculation would be:

	Output	
Man	*1950*	*1955*
1	100	102·0
2	100	90·0
3	100	108·3
4	100	100·0
5	100	105·6
Total	500	505·9
Average	100	101·2

It could then be said that, on average, this group of men produced 1·2 per cent more in 1955 than in 1950.

In this very simple example, the variables concerned (the output of five different men) were regarded as being of equal importance, so that a straightforward arithmetical average could be used to arrive at the overall index number. In many cases, however, the variables making up the group are of *un*equal importance, and allowance has to be made for this. For example, suppose that an index number of industrial production is being constructed—i.e. a single number that will represent the average change in the output of industry as a whole over a given period. If there are in the country ten different industries of varying degrees of importance, then a system of *weights* will have to be used to ensure that a movement in a big industry is given more influence in the index than a similar movement in a minor industry. These weights will have to be based on some factor which can be taken to measure the relative importance of each of the industries entering into the index. In the case of industrial production, a tolerably reliable guide to the relative importance of different industries might be the numbers of people employed in each. For example, an unweighted and a weighted index might be calculated as follows:

Industry	*Output*		*Numbers employed in Year 1 (000's)*	*Output-relatives*	
	Year 1	*Year 2*		*Year 1* [(2) × (4)]	*Year 2* [(3) × (4)]
(1)	(2)	(3)	(4)	(5)	(6)
1	100	104	2·0	200·0	208·0
2	100	103	4·5	450·0	463·5
3	100	94	1·2	120·0	112·8
4	100	110	5·8	580·0	638·0
5	100	101	3·0	300·0	303·0
6	100	82	2·1	210·0	172·2
7	100	95	0·6	60·0	57·0
8	100	99	1·0	100·0	99·0
9	100	103	1·2	120·0	123·6
10	100	106	3·1	310·0	328·6
Total	1000	997	24·5	2450·0	2505·7

The unweighted index numbers would be calculated by simply dividing the totals of columns (2) and (3) by the number of variables included, i.e. 10. This would give 100 for Year 1 and 99·7 for Year 2, suggesting a slight fall in industrial production as a whole. To calculate a weighted index, however, the totals in columns (5) and (6) would be arrived at, and these would be divided by the sum of the weights used—i.e. the total of column (4). This would give index numbers of 100 for Year 1 and 102·3 for Year 2, suggesting a *rise* in overall industrial production.

The choice of the weights used in calculating an index number can be of great significance, therefore; given a certain set of changes in the variables whose overall movement is being measured, the use of two different sets of weights may well result in final index numbers which do not change even in the same direction, let alone to the same degree. Naturally this fact is of considerable practical importance.

3. THE RETAIL PRICE INDEX

Without any doubt, the index number most widely known and used in Britain is the index of retail prices, more commonly but less accurately spoken of as the cost-of-living index. Strictly speaking there is no such thing as *the* cost of living, just as there is no such thing as *the* standard of living. Conceptions of what constitutes a reasonable standard of life (which obviously must be defined before the measurement of changes in its cost can be attempted) differ widely as between one person and another, and from year to year. What is nowadays regarded as bare subsistence, previous generations would have regarded as something approaching luxury. Similarly, one man may regard cigarettes as an essential element in his customary standard of life and will be sensitive to changes in their price, while another man may be quite indifferent if the price of cigarettes is doubled or trebled.

Hence, what is frequently called the cost-of-living index is really something much more limited and well defined. It is an index number which seeks to measure the average change in the retail prices of a representative sample of the goods and services purchased by a particular section of the community. Three main issues are therefore raised in compiling such an index. First, the section of

the community which is to be covered by it; secondly, the commodities to be included; thirdly, the weights to be attached to the various commodities. There has been a series of retail price index numbers in Britain since 1892, but it will be best to confine the discussion to the ways in which the current index, which was first published in January 1956, deals with these three questions.

The first question—that of the people whose cost of living is to be measured—is open to a fairly straightforward solution. The index is compiled by government statisticians, and the government is naturally concerned with the position of the majority rather than that of small groups. Hence, the current index is based on the expenditure of households whose head had a gross income (in 1953) of up to £20 a week, but excludes households solely dependent on pensions or national assistance of some form. This means that the expenditures involved were representative of 9 out of 10 households in the country. Although the index cannot be said to measure changes in the cost of living for every single section of the community, therefore, it certainly takes account of the prices which are important to the vast majority.

The second problem is that of selecting the commodities to be included. By carrying out a survey of the expenditure of some 13,000 households over the year 1953, a mass of information about consumption patterns was obtained. As has been observed already, it would be an enormous task to collect information about the prices of every type of every individual commodity involved, especially when it is wished to compute the index every month. To make the work manageable, therefore, total expenditure by these households was classified into ten broad groups; certain commodities were taken as being representative of each group, and certain types or brands were taken as being representative of each commodity. In the current index, therefore, the prices of all commodities entering into the budget of the average household are assumed to be accurately represented by the prices of 350 individual commodities of specified type or quality. That the range of expenditure covered by this group of 350 commodities is a very wide one may be deduced from the inclusion in it not only of obvious items such as bread, milk, meat and fish but commodities such as television sets, washing-machines, ice-cream, chocolate biscuits and telephone rentals.

There remains the third problem, that of the weights to be attached to the different commodities included in the index. Obviously it would be absurd to treat a rise of, say, 10 per cent in the price of chocolate biscuits as being equal in importance to a similar change in the price of bread, which is what would happen if the index was calculated by taking an arithmetical (i.e. unweighted) average of all the separate price indices. But on what basis should different weights be allotted to the various items? The method adopted is the allocation to each item within a group, and to each group in the index, a weight which reflects the share of total household expenditure devoted to it. For example, if the average household spends three times as much on bread each week as it does on chocolate biscuits, then three times as much weight would be attached to a change in the price of bread as to a change in the price of chocolate biscuits. The survey of the expenditure of 13,000 households therefore indicates not only the commodities to be covered by the index but also the relative importance to be attached to them.

If an index is to reflect accurately changes in the cost of things actually bought by the average household (as opposed to changes in the cost of an unvarying bundle of goods) obviously it will need to be revised periodically as new items of expenditure become popular and old ones drop out, and as the relative importance of the items changes with shifts in spending habits. How far consumption patterns can change may be seen from the following comparison between the relative importance attached to the main types of commodity in the index published in Britain between 1914 and 1947, and their relative importance in the current index.

	Weights used in index numbers (%)	
	1914 *index*	1956 *index*
Food	60·0	35·0
Alcohol and tobacco	0·8	15·1
Housing, fuel and light	24·0	14·2
Clothing	12·0	10·6
Durable household goods	0·5	6·6
Services	1·2	12·6
Miscellaneous goods	1·5	5·9
Total	100·0	100·0

Because of these constant changes in expenditure patterns, the current retail price index departs from previous British practice in that the items included, and the weights attached to them, are kept under constant review in the light of a continuing investigation of household expenditure. This has the advantage of ensuring that the prices included are the prices which matter to the average modern family, and are not those of a group of commodities which have ceased to represent the prevailing pattern of expenditure. It has the disadvantage that it is difficult to see in what sense the values of the index number at two widely-separated dates can be compared, because the selection of commodities entering into the index may be quite different at the two dates. However, provided that the comparisons attempted are over, say, three or four years (during which period the selection of commodities is almost bound to be *substantially* unchanged) and not thirty or forty years, this disadvantage will be more than offset by the advantage of having the index on a realistic base.

Apart from this, there are certain dangers inherent in index numbers of this kind if they are taken (as they often are) to measure changes in the cost of the prevailing standard of living. First, it is difficult to make allowance for changes in quality which may offset, or more than offset, changes in price. For example, the average television set may be 10 per cent dearer in 1957 than in 1956, but if the quality of the average television set has improved between those years the cost of a given standard of television consumption may have fallen rather than risen.

Secondly, the index may not always reflect changes in taxation and social services. If indirect taxes on consumption goods are increased, prices will rise and the index will rise with it, suggesting that a given money income has a smaller real value. If *income tax* is increased, however, a similar fall occurs in the ultimate purchasing power of a given money income, but the index will give no indication of this change. Conversely, the index will suggest no increase in the real value of a given income if certain services (e.g. health services) are provided free instead of at a charge, but the standard of life associated with any given income will undoubtedly have risen.

Finally, it must be remembered that index numbers of this kind

can reflect the expenditure pattern only of broad social groups, within which there may be individual groups undergoing different experiences. For example, the retail price level may be constant for the average household, but the average household is, of necessity, something of an abstraction. Marked changes may be occurring in the prices of commodities of especial importance to particular groups—old-age pensioners, or families with small children, or professional classes. Substantial gains or hardships may therefore be concealed in an overall average.

Nevertheless, provided it is interpreted with care, the retail price index remains a useful device. It enables a broad impression of changes in the price level of consumers' goods as a whole to be formed, and such changes may be important symptoms of the forms of economic indigestion known as inflation and deflation. The index can give a guide to government agencies as to changes in the real value of pensions and grants, while a great many wage-agreements in Britain are related to the index of retail prices, so that workers do not have to renew wage negotiations every time the purchasing power of their wages is reduced by a rise in the general price level. (Although, as has already been indicated on more than one occasion, this system of automatically relating wages to the general price level is not without serious dangers.) Finally, the index enables comparisons to be made between the real purchasing power of wages and other incomes at different dates, and hence permits an assessment of the effects of given changes in economic policy without having the issue obscured by price movements.

4. OTHER INDEX NUMBERS

For the questions with which this book is concerned, the index number of greatest importance is the retail price index discussed in the preceding section, insofar as this can be used to measure changes in the general level of prices. Other kinds of price measures may be of at least as much interest in certain connections as the retail price index however, and it is therefore worthwhile to examine briefly some of the other index numbers which attempt to indicate changes in various groups of prices and which are published regularly in Britain.

To a producer, changes in his costs of production will be of as much importance as changes in the market value of his product. An index number of *wage-rates* reflects movements in one major cost element, and also indicates the development of the most important source of spending-power in the community. The construction of such an index is basically similar to that of the index of retail prices. Instead of averaging the prices of different commodities, the wage-rates of different types of labour are averaged; the weights attached to the individual wage-rates are based on the numbers in each labour-group. For some purposes, and certainly for that of examining movements in total personal incomes, an index of average *earnings* rather than wage-rates may be more illuminating, since it will take account of changes arising not only from adjustments in wage *rates* but also from changes in the amounts of overtime and bonuses. Both types of index number are published monthly, not only for the country as a whole but also for important industrial groups.

Another important cost element is that of the price of raw materials, and index numbers are published every month indicating movements in the *wholesale prices* of the various types of raw material used extensively in production. Similarly, monthly index numbers are published showing changes in the prices of the *output* of important industries which supply other industries with materials or semi-finished goods, and others measure changes in the average value of the goods which are *exports* from or *imports* into the country. As will be seen in Part VIII, changes in the relative prices of exports and imports can be of enormous importance to a country, such as Britain, which maintains a high level of international trade.

Apart from all these price indicators, monthly or quarterly index numbers are published which provide an indication of changes in *quantities* rather than prices. The most important of these is the monthly *index of industrial production*. A simple example of such an index was given earlier in this chapter; the official index of industrial production in the United Kingdom is a more complex version of this. Returns are sent to the official statisticians each month by a sample of firms which are representative of major industrial groups; these indicate the level of output produced by the firms concerned

in the preceding month, output being measured usually in physical terms (tons of steel, pairs of shoes and so on). From these data, a weighted index number is computed for each industrial group, the weight attached to the output of each firm being based on its relative importance in the group's total output. (This is discovered from an examination of the findings of the Census of Production, a survey which was briefly described in section 1 of Chapter 19.) Because the index of industrial production can be computed every month, it provides an extremely important indicator of current developments in a major segment of the national economy.

Other index numbers of a quantitative nature are published monthly or quarterly covering the physical volume (as distinct from the value) of imports and exports, the level of retail sales, volume of rail traffic and a great many other economic variables. Each represents a simple and useful guide to trends in different spheres of economic activity or over the economy as a whole. All of them have to be interpreted with care, since they are subject to statistical and other weaknesses which sometimes seriously qualify their significance.

FURTHER READING

A. C. PIGOU, *Income* (Macmillan, London, 1948), Ch. I.

R. G. D. ALLEN, *Statistics for Economists* (Hutchinson, London, 1948), Ch. VI.

L. R. CONNOR, *Statistics in Theory and Practice* (5th ed. Pitman, London, 1938), Ch. XVI.

PART VIII

INTERNATIONAL ECONOMICS

CHAPTER 29

The Significance of International Trade

1. THE LAW OF COMPARATIVE COSTS

Just as a modern economy is based upon internal specialisation, so the world as a whole has tended towards international specialisation as economic life has developed and expanded. Few countries aim at supplying all their needs directly from the productive activities of their own citizens. To some extent this is because self-sufficiency is seldom possible; apart from this, however, it is clear that even if it were possible it would not be to the best advantage of the countries concerned. Resources, human and material, are unevenly distributed over the face of the globe, just as talents and abilities are unevenly distributed amongst the population of any given country. Some countries are endowed with a climate favourable to agricultural production, some are not; some have vast mineral resources, others have none; some have an abundant supply of skilled labour, others suffer from a shortage of it. If resources were fully mobile, of course, they would all tend to gravitate to that region in which their combined efficiency was greatest. While factors are relatively immobile within single countries, however, their international mobility is infinitely smaller. As a result, the cost of producing any particular commodity will vary widely from one country to another, according to the supply in the different countries of the resources required in its production. Hence, it will be to the mutual advantage of all concerned that each commodity should be produced in those areas best suited to its production, and that the different areas should exchange products with one another.

International trade does not develop solely because a country is more efficient than others in the production of some commodities and less efficient in producing other commodities. If this were the

sole justification of international trade, it would be impossible to explain the foreign trade of many countries which in fact export to and import from others with vastly different levels of productivity; many countries can produce *all* the commodities they consume more efficiently than any of the others with which they in fact conduct trade. Why should this be so? Why should any country import from abroad goods which it could itself produce at a lower cost?

The explanation of this phenomenon is usually formalised in what is known as the *law of comparative costs*, which states that countries will derive mutual benefit from trading with one another so long as there is some difference in their *relative* efficiencies in the production of the commodities concerned. What this means is that trade will be beneficial, despite the overall productive superiority of some countries, if it enables the more efficient countries to concentrate on the production of goods in which their superiority is especially great, importing from the rest of the world the commodities in which their productive superiority is less marked. There is here an exact analogy with the division of labour. If a man is both a first-rate musician and a first-rate motor-mechanic, it will profit him to leave the repair of his car to a professional mechanic, even though this involves the use of more time and materials than he would have used if he had done the job himself. The reason is, of course, that by devoting the time to music he can earn more than enough to pay even the relatively high cost of having his car repaired. Similarly, a country with especially great efficiency in certain types of production will earn more than enough by concentrating on those lines to repay the extra cost of importing goods which it could have made itself rather more cheaply than its foreign suppliers.

An example may help to make this clear. Consider a hypothetical world in which there are only two countries, A and B, and only two commodities which will be called loaves and machines. Taking the simplest situation first, suppose that A is able to produce 20 loaves or 5 machines with each of its units of labour ('labour' being used to represent a composite bundle of given amounts of all factors of production), and that B is able to produce with each unit of labour 18 loaves or 6 machines. In other words, A is more efficient in the production of loaves and B in the production of machines. Obviously

it will pay these countries to specialise on their most efficient line and import the other commodity. If they aimed at self-sufficiency, and carried on no trade, with two units of labour in each country production would amount to

Country A -	20 loaves	5 machines
Country B - -	18 loaves	6 machines
Total - -	38 loaves	11 machines

If they now specialise in their best lines, production with the same outlay of two units of labour in each country would amount to

Country A - -	40 loaves	
Country B - -		12 machines

giving a net gain of two loaves and one machine which the two countries can share between them. (Exactly how the gain will be shared is a matter which will be examined a little later.)

Consider now the less simple situation in which one country is more efficient in the production of both commodities. Suppose that without specialisation and foreign trade the outlay of two units of labour in each country resulted in the following outputs:

Country A - -	20 loaves	8 machines
Country B - -	18 loaves	4 machines
Total - -	38 loaves	12 machines

Will there be any point in Country A specialising in one product and importing the other? After all, it can produce each more cheaply than B. But consider A's *relative* superiority. In the production of loaves its efficiency is greater in the ratio of 20 : 18, or $1\frac{1}{9}$ to 1; in the production of machines the ratio is 8 : 4, or 2 to 1. The law of comparative cost states that both countries will gain if A specialises in the commodity in which its relative efficiency is greatest—i.e. machines. If this happens, the outlay of two units of labour in each country will result in the following outputs:

Country A - -		16 machines
Country B - -	36 loaves	

The gain here is not so obvious, since more machines but fewer loaves are being produced. Assume that A exchanges 6 of its

machines for 20 loaves. Clearly it has benefited from trade: it now has 20 loaves and 10 machines compared with the 20 loaves and 8 machines it would have had without specialisation and trade. What of B? It is left with 16 loaves and 6 machines, compared with the 18 loaves and 4 machines it would have had without trade—a loss of 2 loaves and a gain of 2 machines. Has it profited on balance? To answer this question, consider B's gains and losses in terms of the resources B would have needed to make the goods concerned itself. With one unit of labour it could produce 18 loaves or 4 machines. Its loss of 2 loaves therefore represents the equivalent of $\frac{1}{9}$ of a unit of its labour, while the gain of 2 machines represents the equivalent of $\frac{1}{2}$ of a unit of its labour. On balance, then, B has gained $\frac{1}{2} - \frac{1}{9}$, or $\frac{7}{18}$ of a unit of labour. Both countries, therefore, have gained from specialisation and foreign trade.

The principle of comparative costs can be generalised, of course, to cover a world in which there are more than two countries and considerably more than two commodities. National specialisation and foreign trade perform for the international economy the role performed by division of labour and geographical specialisation within an individual economy. So long as differences in relative costs of production exist between countries, specialisation and trade enable the world as a whole to maximise the output obtained from its scarce resources, given the inability or unwillingness of factors to move across national frontiers so as to be everywhere combined with one another in the optimum proportions.

This is a theoretical conclusion, but it does not follow that in the actual world the productive activities of each country will be solely governed by the operation of the law of comparative costs. First, the law assumes the existence of complete freedom of trade; in reality, comparative production costs are distorted, and their effects on trade stultified, by the use of tariffs, quotas and other hindrances to the international exchange of goods and services. Secondly, the effect of transport costs must be allowed for; although a commodity may be more efficiently produced in A than in B, the costs of transporting it from A to B may be so great as to more than offset A's productive superiority, and B will find it preferable to produce the commodity itself despite its lower efficiency. Thirdly, the existence of different levels of efficiency in the production of

any given commodity need not result in complete specialisation within a particular country, since diminishing returns may set in if output exceeds a certain point; a country may retain its superiority in the production of the commodity only so long as it limits its output below the level necessary to meet the entire world's needs for it, so that other countries will still require to produce a certain amount.

In general, then, it can be said that countries tend to specialise in the production of those commodities which they produce most efficiently in relation to other commodities, but that in many cases the full working-out of this tendency is prevented by artificial obstacles to trade, by transport costs and by the possibility of variation in average costs of production in different countries as their levels of output change. Insofar as it operates, the law of comparative costs indicates how free trade between nations, in precisely the same way as division of labour within a nation, can raise the overall real income of society.

2. THE TERMS OF TRADE

The analysis must now return to a question which was left on one side during the discussion of the law of comparative costs. Assuming that countries specialise in producing certain commodities and then carry on trade with one another, how are the relative values of the traded commodities determined?

If the first example quoted in the previous section is considered once again, it will be apparent that the conditions of production in the countries involved will impose definite limits on the rate at which they will exchange goods with one another. The productive abilities of the two hypothetical countries were such that with each unit of labour

Country A could produce 20 loaves or 5 machines
and Country B could produce 18 loaves or 6 machines.

If A now specialises in producing loaves, it is obvious that it will not exchange more than 4 loaves for a machine, since it could itself produce a machine with the labour it uses in making 4 loaves. If it were to give *more* than 4 loaves for a machine, then, it would be losing rather than gaining from international trade. Similarly B

will not be willing to exchange a machine for less than 3 loaves, since it can itself produce 3 loaves with the same amount of labour as it needs to make a machine; rather than accept, say, 12 loaves for 6 machines, it would prefer to employ the labour necessary to make 6 machines in the production of loaves, making 18 loaves as a result. The upshot is, then, that the rate of exchange between loaves and machines cannot be more than 4 to 1 (or A will not trade) and cannot be less than 3 to 1 (or B will not trade).

The rate of exchange must therefore lie somewhere between 3 and 4 loaves to one machine. What decides the exact point at which the relative values of the traded commodities, usually known as the *terms of trade*, will settle? Essentially it is a matter of demand and supply. If, in this example, the supply of loaves from country A is large relative to the demand for them in B, then A will have to offer something close to 4 loaves in order to obtain a machine. On the other hand, if B's demand for loaves should suddenly rise, or if the supply should be reduced for some reason, B will have to offer more machines for any given quantity of loaves, and the exchange rate will fall closer to 3 to 1.

In this example, the limits imposed by production conditions are so narrow that the terms of trade between A and B (i.e. the value of machines in terms of loaves) cannot vary much. In the world of reality, however, the major imports of the United Kingdom are foodstuffs and raw materials, and the cost of producing British requirements of these products in overseas countries is vastly different from what it would be if Britain had to produce them herself. Similarly, if Britain's major buyers had to produce for themselves the engineering products they now import from Britain, their cost would be much greater than it is at present. The limits to the rates of exchange between British imports and exports set by the production possibilities, therefore, are far apart, with the result that Britain's terms of trade can vary considerably.

Changes in the terms of trade can most easily be measured by a comparison between the average price of exports and the average price of imports. If a country's exports become cheap relatively to its imports, then more goods will have to be exported by that country in order to purchase a given quantity of imports, and the terms of trade are said to have worsened. Conversely a rise in the

value of exports in terms of imports—an improvement in the terms of trade—enables a country to purchase a given volume of imports with a smaller volume of exports. The degree to which the terms of trade of the United Kingdom have varied in the present century may be seen from Table XIV.

TABLE XIV: *Terms of Trade of the U.K.* 1913–55

	Average value of exports	*Average value of imports*	*Terms of trade* (2 ÷ 1)
	1	2	3
1913 - -	68	98	143
1923 - -	130	146	113
1938 - -	100	100	100
1946 - -	196	213	109
1951 - -	319	443	139
1955 - -	323	394	122

Source: *London & Cambridge Economic Service Bulletin*, No. 18, June 1956.

The 'terms of trade' index in column 3 is obtained by dividing the index number of the average value of imports by that of the average value of exports; a rise in the terms of trade index, therefore, indicates a *worsening* of the terms of trade—i.e. that the prices of the country's imports have risen in relation to those of its exports—while a fall indicates an improvement. As can be seen, between 1913 and 1938 Britain benefited enormously from the change in the relative values of exports and imports; the average value of her exports rose by nearly 50 per cent while import prices remained almost unchanged. A good many causes underlay this development, the major ones being the large increase in world agricultural production stimulated by the disruption of trade channels during the first World War, the fall in the cost of primary products (i.e. the products of agriculture, mining, forestry and fishing) resulting from rapid technical developments during the 1920's and 1930's, and the collapse of commodity prices during the depression of the 1930's—primary products, being in highly inelastic supply, are liable to marked fluctuations in price in response to shifts in demand.

Since 1938, however, the tables have been turned, largely be-

cause of the fact that the production of agricultural commodities has risen substantially less than that of manufactured goods. To some extent this has been due to political disturbances in major sources of agricultural supplies (e.g. Malaya and Burma); more important, however, has been the tendency for countries which were previously almost wholly agricultural to shift their resources into manufacturing to an increasing degree (e.g. Argentina and Australia). The sharp rise in import prices during 1950 and 1951 was a result of the sudden desire of many countries to accumulate stocks of primary products, such as tin and rubber, after the outbreak of the Korean war; the inelasticity of supply of these products resulted on this occasion in a violent increase in their prices. Since 1951, Britain's terms of trade have improved somewhat, although in 1955 they were still significantly less favourable than at any time since the first World War.

The direct effect of a worsening of a country's terms of trade is bound to be unfavourable, of course. In 1955, Britain's expenditure on imported goods amounted to about £3,500 million, so that a rise of 10 per cent in the price of imports relative to that of exports would add some £350 million to the value of exports needed to purchase a constant volume of imports. Obviously this would be a considerable increase in the real burden falling on the economy. However, certain indirect effects of changes in the terms of trade would tend to offset this. First, a rise in import prices represents a rise in the incomes of foreign suppliers, and this will tend to raise their demand for goods exported by other countries. A country whose terms of trade have worsened, therefore, will often find that after a short time-lag the volume, and probably the price, of its exports begin to rise, easing the burden of the increased import bill. Secondly, a country such as Britain has invested a great deal of capital in primary producing countries over the last two centuries, so that some of the increased profits resulting from a rise in the price of primary products will come back to it in the form of increased profit earnings on its foreign assets. These compensating effects are unlikely to offset the direct effect entirely, however, and instability in the terms of trade must always remain a potential source of danger to countries heavily dependent on foreign trade.

3. INTERNATIONAL TRADE AND EMPLOYMENT

In the discussion of the components of aggregate demand in Chapter 25, the question of foreign trade was left on one side. It is necessary now to fit exports and imports of goods and services into the pattern of income-determination. Exports share many of the features of investment expenditure; like investment, the production of exports results in the receipt of incomes by the factors involved in their manufacture, so that the overall levels of incomes and demand are raised. Again, the production of exports, like that of capital goods, adds nothing directly to the supply of consumer goods available in the home market, so that there is no increase in supply to offset the increased demand generated by the incomes paid out in their manufacture. Like investment expenditure, then, the production of exports is inflationary.

Imports, on the other hand, are the opposite in both respects. The incomes generated by their production do not enter into a country's income flow; they are received by factors of production in some other country. On the other hand, imports *do* add to the supply of goods available in the home market, so that they increase aggregate supply without resulting in any increase in aggregate demand. Imports, then, have a deflationary effect on the economy.

One way to incorporate foreign trade into the analysis of the determination of income and employment is to treat the net excess of exports over imports as an addition to investment expenditure (or as a deduction from investment expenditure if imports exceed exports). The income-determination equation therefore becomes

$$Y=(I+E')\frac{1}{1-C'}.$$

Where E' represents the excess of the value of exports over that of imports ('net exports') and the other symbols represent the same elements as before.

Alternatively, the fraction of income which is spent on imported goods (*the propensity to import*) can be treated as an income 'leakage' exactly comparable with savings, since expenditure on imports prevents the passing-on of income to other members of the eco-

nomy, just as saving does. The income-determination equation will then be

$$Y=(I+E)\frac{1}{1-C'+m},$$

where E represents the total value of exports and m the fraction of income which people spend on imported goods. Both expressions will give the same answer, of course. For example, assume that in an economy the rate of investment is £100 million, the level of exports is £80 million and imports £60 million and the overall average propensity to consume is $\frac{4}{5}$. The first type of formulation would give (in £ million):

$$Y=(100+20)\frac{1}{1-\frac{4}{5}}=600.$$

Since imports are £60 million out of a total income of £600 million the propensity to import can be taken as $\frac{1}{10}$ and the second type of formulation gives:

$$Y=(100+80)\frac{1}{1-\frac{4}{5}+\frac{1}{10}}$$

$$=180\times\frac{1}{\frac{3}{10}}=600.$$

The second method has the advantage that by it the tendency to purchase imports is automatically allowed for, whereas with the first method the total expenditure on imports has either to be known or to be estimated separately beforehand.

The importance of foreign trade in the determination of the level of incomes and employment is therefore fairly clear. An export boom will have precisely the same internal effects as a rise in the rate of domestic investment, while a switch in the expenditure of consumers from foreign to home-produced goods will have the same internal effect as a fall in the propensity to save. The interrelationship of exports and imports must never be forgotten, however. Obviously, for the world as a whole exports and imports must be equal in value, so that the inflationary effects in one country of a rise in exports (or fall in imports) must be precisely offset by deflationary effects somewhere else in the world. As later discussion will

show, this fact has considerable importance in assessing the relative advantages of imposing restrictions on foreign trade.

4. FREE TRADE AND PROTECTION

The law of comparative costs points clearly to the conclusion that complete freedom of trade between countries will encourage international specialisation and secure all the possible benefits which can arise from it. Yet in real life foreign trade is distinguished by the multiplicity and extent of the restrictions imposed on it by almost every country. Imports are discouraged by tariffs (taxes which raise the cost of imports) or quotas, i.e. regulations limiting the amount of a particular commodity which may be imported into a country in a given period. Occasionally, exports are encouraged or discouraged by the grant of subsidies or imposition of special taxes. As will be seen later in the discussion of the monetary aspects of international trade, the exchange of currencies against one another is frequently subject to restrictions imposed with a view to limiting a country's imports from certain areas. What are the reasons for all these restrictions and interferences?

To a large extent the reasons are of a political or social, rather than economic, nature. Agriculture is frequently protected by restrictions on imports, for example, mainly on strategic grounds; it is feared that home agriculture could never survive if it had to face the full competition of imported products, and the extinction of agriculture might prove to be a serious business if a country found itself at war and cut off from its foreign suppliers. Similarly, industries may be protected because it is felt that they play an important role in the social life of the community, or for no better reason than that the employers or workers of an industry are in a position to bring political pressure to bear on the government to relieve them of uncomfortable competition from abroad.

Can interference with free trade ever be justified on strictly economic grounds? In general the logic of analysis leads firmly to the conclusion that it cannot. If foreign trade develops, it can only mean that one country has at least a relative advantage over another in the production of a given commodity, and interference with the free flow of goods can result only in a sacrifice of some of the benefits which would follow from specialisation. The country restricting

its imports is reducing its real income, because it is using resources to produce goods at an unnecessarily high cost.

One exception of an economic nature is frequently quoted, however. This is the case of an *infant industry*. It sometimes happens that a country seeks to develop an industry for which it is well suited, and in which it could effectively compete with foreign producers once the industry was well established. In the early stages of its development, however, the industry is unable to meet competition from abroad; it has of necessity to commence on a small scale, so that it cannot yet reap the full economies of large-scale output, and it has had insufficient time to build up a skilled labour force, specialised marketing and maintenance organisations, or the accumulated experience of well-established entrepreneurs. Hence protection in the early stages of its development would permit it to acquire all these things and to grow into a mature and fully-competitive industry requiring no further protection, just as a small child must be protected from the full force of the elements if it is to grow eventually into a healthy and virile adult. In principle, the argument is sound enough; it simply takes account of the fact that in the real world knowledge is not perfect and economic development is not instantaneous. In practice it leaves a good deal to be desired. Industries protected from competition have a habit of never growing up; the sheltering of an industry is not always conducive to the energetic pursuit of maximum efficiency. Furthermore, the removal of protection would tend to injure the employers and workers of the industry concerned, and these, being a relatively small and homogeneous group, would be quick to express organised opposition to the change. Those who stand to benefit by the removal of protection, on the other hand, are the general body of consumers, who are unorganised and frequently unaware of the loss they suffer through protection and the resulting high prices. Changes which arouse fierce opposition from the potential losers but no acclaim from the beneficiaries are seldom regarded as practical politics.

Other arguments of an economic nature advanced in favour of protection have even less real foundation. One of these is that a high-wage country must protect itself against competition (invariably labelled 'unfair') from low-wage countries. The argument usually loses sight of the fact that it is wage-cost per unit which

matters in trade, rather than the absolute level of wages; the United States has the highest wage-level in the world, but its labour costs per unit of output are amongst the lowest. Secondly, the argument overlooks the fact that any attempt to shut out goods produced in low-wage countries can only tend to maintain those low wages, or even reduce them further; only in conditions of expanding markets and relatively full employment can wages be expected to rise. The main weakness of the argument, however, is that a country can only gain by obtaining goods at a low cost, since its real income is raised thereby; if an exporting country chooses to hold down its real income in order to export goods, the loss is the exporter's, not its customers'.

Another argument common in periods of depression is that protection should be imposed in order to stimulate home employment. Although the suggestion has a sound theoretical basis (as has been seen) its attraction is largely illusory. Protection will provoke retaliation, both automatic and deliberate. If a country reduces its imports, the incomes of its suppliers will be reduced, and this is bound to react adversely on their demand for the protecting country's exports if that country plays a major role in world trade. (If it does not, there is little to be gained from protection anyway.) Furthermore, protection is a game at which more than one can play; a country placing restrictions on its imports will almost always find that similar restrictions have been placed on its exports in retaliation. Any stimulus to employment in home industry, therefore, is likely to be partly or wholly offset by the damage done to export industries. The experience of the 1930's demonstrated this, if nothing else.

As a long-run policy, then, protection can be adopted only at the cost of a loss of real income. It may be that strategic or social gains are more than sufficient to compensate for this, but restrictions on free trade are seldom justifiable from a strictly economic point of view. At the same time, the world of reality is one in which readjustment and transference of resources, especially labour, as between different industries is a slow and painful business, and where the unaided operation of the price mechanism may be weak and wasteful when large and sudden changes are necessary. In a world of perfect competition the case for completely free trade

would be unassailable on economic grounds. In the world of reality, where competition is usually far from perfect, the issues are less clear-cut. When unrestricted imports may result in the wastage of capital and unemployment of labour, it is by no means always obvious that the balance of even economic advantage lies unambiguously with completely free trade. Economic welfare may often be augmented, rather than reduced, by temporary protective measures designed to assist and ease a process of transition. It is one of the unfortunate facts of life that such measures have an unhappy knack of establishing themselves as irremovable features of the economic and political landscape, and therefore result in a permanent loss of real income; this loss is no less regrettable because it is not readily apparent to the general public by whom it is suffered.

FURTHER READING

R. F. HARROD, *International Economics* (Rev. ed., Cambridge University Press, London, 1939), Chs. I–III, IX.

J. ROBINSON, *An Introduction to the Theory of Employment*, Ch. XI.

P. T. ELLSWORTH, *The International Economy* (Macmillan, New York, 1950), Chs. 5–8.

CHAPTER 30

Foreign Exchange

1. THE MECHANISM OF FOREIGN PAYMENTS

The conduct of international trade by barter—with producers in different countries exchanging their products directly for one another—would be subject to all the disadvantages of barter in internal trade. Since the use of money overcomes these disadvantages for transactions between residents of the same country, transactions between those of different countries are naturally also carried out through the medium of money. This permits *multilateral trade*—that is, the sale of goods in one country and the use of the proceeds in the purchase of goods from another—rather than a system in which exporting countries are forced to purchase their imports from the country in which their goods are sold. As in internal trade, the use of money in international exchanges allows producers to sell in the best market and buy in the best market, whether or not those markets happen to be found in the same country.

In international trade, however, a complication arises which is absent from exchanges within a country. This is that the various independent countries of the world each have their own separate currencies, with separate names, differently valued units and issued on different bases. Even currencies bearing the same name may be found to have different varieties. The Fiji pound, for example, is a different pound from the one circulating in the United Kingdom, while the Tongan pound is distinct from both; there are French francs, Belgian francs and Swiss francs, all with different values.

When an importer has purchased goods from a foreign producer, therefore, he is frequently (although not always) faced with the problem of securing a foreign currency in order to pay for the goods.

An importer of American wheat, for example, will find his supplier unwilling to accept payment in pounds sterling; the American cannot use pounds to pay his workers, to build a new factory, or to buy petrol for his car, in a country where dollars, and not pounds, circulate as money. In America the British pound would lack the first attribute of money—i.e. general acceptability. The importer has therefore to secure an adequate supply of the currency used in his supplier's country.

He will do this through the *foreign exchange market*, which is the market in which the different currencies of the world are exchanged for one another. The use of the word 'market' may be a little misleading here, since there is no single institution or locality in which such exchanges are carried out. The foreign exchange market comprises all those persons and institutions who, usually amongst other things, carry on the business of buying and selling foreign currencies; it includes, therefore, Central and commercial banks as well as independent brokers and agents, merchant banking houses and bullion dealers.

An ordinary importer or exporter will usually secure foreign currencies through his bank. For example, if the trader wishes to use his British bank deposit to pay a debt in French francs, his bank will instruct its Paris branch (or a French 'correspondent bank' which has agreed to act as its agent) to credit the trader with the appropriate amount of French francs, and the trader can then transfer this credit to his supplier in the same way as he would transfer a bank deposit to another British resident by means of a cheque. The British bank will reduce the trader's credit in its books, of course, by the equivalent sum in pounds calculated at the current rate of exchange between British pounds (pounds sterling) and French francs.

In the normal course of events, a bank transferring pounds into francs in this way would be receiving a more or less equal amount of business in the opposite direction; its Paris branch, or correspondent, will be exchanging francs for pounds on behalf of *its* customers. If there is a net movement in one direction—say, more pounds are being turned into francs than francs into pounds—then the British bank will find that its balance in Paris is falling, or that its indebtedness to its correspondent French bank is growing. It

may therefore need to replenish its holdings of francs by an outright sale of sterling on its own account.

It would usually do this through a specialist foreign exchange broker (an individual or group of individuals) whose business is the purchase and sale of currencies, rather than the borrowing and lending of currencies. These brokers make their living by performing three functions. First, they carry out routine purchases and sales for clients on a commission basis. Secondly, they conduct complicated *arbitrage* business if differences emerge in cross-exchange rates, and thereby even out such differences. An example may make this clear. Suppose that in both London and New York the current rate of exchange between pounds sterling and dollars is £1 = $4, that the rate of exchange between pounds and francs in both London and Paris is £1 = Frs. 1,000, and that in New York the rate of exchange between dollars and francs is $1 = Frs. 250. This is an equilibrium situation. But suppose the dollar-franc rate in New York for some reason changes to $1 = Frs. 200. The arbitrage operator in London can now buy Frs. 1,000 with £1 in London, sell the Frs. 1,000 in New York for $5 and then re-convert his $5 into pounds, ending up with £1 5s. 0d. The rush to sell francs against dollars and dollars against pounds will be so great that either the New York dollar-franc rate or the dollar-pound rate (or both) will immediately move, so as to correct the discrepancy. Finally, the foreign exchange broker will buy and sell single currencies on his own behalf, buying currencies whose exchange rate he expects to rise, and selling those whose rate he expects to fall. In this, of course, he hopes to make a profit; insofar as his expectations are based on genuine assessments of the inherent values of various currencies, he carries out the valuable economic function which can be performed by all speculators—the bringing-forward in time and consequent smoothing of future changes. There is in all these respects a close similarity between the activities of the foreign exchange broker and the Stock Exchange jobber.

A final special aspect of the foreign exchange market is that known as *forward exchange*. Many traders and investors find it desirable to have the promise of a certain amount of a specified currency (at a known rate of exchange) well in advance of the date on which they expect to be using it. By placing an order for the future delivery of

a currency at an agreed rate they protect themselves against any fluctuations which may occur in the rate of exchange in the meantime. Both banks and brokers therefore buy and sell foreign currencies for future delivery—I can arrange with my bank to transfer pounds into Dutch guilders in three months' time at a rate agreed on now. Naturally the dealers in forward exchange usually quote somewhat less favourable forward rates than the appropriate current rate in order to compensate for the risk they are taking in promising to deliver at a specified rate a currency whose price may have changed considerably when the time comes for them to deliver it.

This is an outline of the bare mechanics of foreign exchange operations. The fundamental question—on what basis is the rate of exchange between one currency and another determined?—has been ignored in giving it. Basically, the exchange rate of a currency can be fixed in one of two ways; it can be maintained in some specified relationship with a precious metal (or some other currency) or it can be determined by the day-to-day pressure of supply and demand in a free market. These two methods of establishing exchange rates, the method of fixed exchange rates and the method of variable rates, carry rather different implications for the progress of an economy, and they are best considered separately.

2. THE ROLE OF EXCHANGE RATES

In the long run, every country entering into trade must aim at securing some kind of balance between its imports and its exports. If it persistently imports more than it exports it can do so only by using up its accumulated assets or by running into debt, and neither process can continue indefinitely. There is an exact analogy here with a private individual who persistently spends more than he earns; he can do it by using up his savings or running into debt, but this state of affairs cannot go on for ever. (Less obviously, a country cannot continuously accumulate creditors; if its creditors never repay, the exports are in fact being given away, and no country will accept such a situation indefinitely.) It follows, therefore, that the prices of a country's exports and imports must be such as to at least hold out the promise of ultimate solvency.

Now export and import prices have two components; the price in

the country of origin and the rate of exchange between the currency in the country of origin and in the country of sale. For example, British motor-cars are sold in France. Their cost to the French buyer is a combination of (*a*) the British price, and (*b*) the pound-franc exchange rate, and a change in the final cost can result from a movement in *either* factor. If a car is sold at £500 to a French importer, and the current exchange rate is £1 = Frs. 800, then the cost to the French purchaser will be about Frs. 400,000. That cost could rise to Frs. 500,000 if *either* the cost of production in Britain rose to £625 *or* the pound-franc exchange rate moved to £1 = Frs. 1,000. Whichever factor was at work, the price of British cars in France would rise, and the amount of them purchased would therefore fall.

A country which finds that its exports and imports are becoming seriously out of step, therefore, can do one of two things. It can *either* adjust its internal price level in the appropriate direction *or* it can vary the exchange rate of its currency. If Britain finds that its imports of goods and services exceed its earnings from exports, it can either try to reduce costs of production at home or reduce the external value (i.e. the exchange rate) of the pound. If home costs are reduced, exports will be cheaper and normally more of them will be sold abroad; conversely, foreign goods will have become relatively expensive as a result of the cheapening of home products, and imports will therefore tend to fall. The same effects would follow from a reduction in the exchange rate of the pound. If British wireless sets cost £20 to produce, American refrigerators cost $500 to produce and the current exchange rate is £1 = $5, then British wireless sets will have to be sold in America at about $100 and American refrigerators will sell in Britain at about £100. If the pound is now *devalued* so that the exchange rate falls to say, £1 = $4, British producers of wireless sets will be able to sell in America at about $80, and still cover their costs, while American producers will have to sell refrigerators in Britain at about £125 in order to cover their costs. British exports will be cheaper, and more will be sold; imports into Britain will be dearer and fewer will be bought.

But why should such adjustments of internal prices or exchange rates be necessary? Assuming that the original rates of exchange

between different currencies are established at such a level that all countries achieve a proper relationship between their exports and imports, why should not these rates be maintained indefinitely? The answer is that internal price levels in different countries never maintain a fixed relationship with one another for long. One country will increase its level of productivity more rapidly than others, for any one of many reasons, so that its level of costs will become lower in relation to that in other countries; alternatively, countries will vary in the monetary policies they pursue, so that quite apart from productivity differences the original relationship of prices in different countries will inevitably be disturbed. Periodically, then, every country will find that its exports are becoming dear or cheap relatively to those of other countries, so that the initial situation of trade balance will be replaced by one of an import or export surplus. To restore equilibrium, either the internal price level or the external value of its currency will have to be changed.

3. FIXED VERSUS VARIABLE EXCHANGE RATES

The relative advantages of fixed and variable exchange-rate policies have been the subject of a great deal of controversy. Until 1931 almost every major trading nation in the world followed the special form of fixed-rate policy known as the *gold standard*. Under this policy, the value of a currency was defined in terms of a specified weight of gold, and the Central Bank accepted the permanent responsibility for converting its currency into gold, or vice versa, on demand and at the specified rate of exchange. The system had the advantage that since the values of all major currencies were defined in terms of the common element of gold, one currency could always be converted into another at a fixed rate, thus stimulating the free flow of multilateral trade with all its benefits.

The complete stability of exchange rates was of special significance, since the possibility of fluctuations in the foreign exchange market introduces serious risks into the business of exporting and importing. Suppose an exporter of British textiles to Italy agrees to sell for 1,800,000 lire a consignment of goods which has cost him £900, at a time when the exchange rate is £1 = 1,900 lire. Under these circumstances he is assured of a reasonable profit; at the current exchange rate, the revenue of 1,800,000 lire can be

converted into roughly £950, so leaving him a surplus of £50 after paying the cost of the goods. Now suppose that, during the interval between entering into the sale agreement and actually receiving the proceeds, the exchange rate moves to £1 = 2,100 lire (i.e. the pound sterling *appreciates* in terms of Italian lire). His 1,800,000 lire will now exchange for only about £860, turning his profit of £50 into a loss of £40. International trade is therefore discouraged and hampered by the possibility of exchange-rate fluctuations, and promoted by stability on the foreign exchanges.

The advocates of the gold standard also attached much importance to the virtue of automatic balancing which was claimed for it. The obligation to convert a currency into gold on demand implied that the domestic issue of credit was strictly related to the gold reserves of the Central Bank—a point referred to in Chapter 21. If Britain, for example, imported more than it exported, more people would be changing pounds into gold (so as to repatriate their earnings from sales to Britain) than would be handing over gold earned with exports for conversion into pounds. The gold reserves of the Bank of England would therefore decline, forcing the Bank to reduce its deposits so as to maintain the conventional rate of gold to total liabilities. The whole process of credit contraction would be set in motion, with interest rates rising, the rate of investment being thereby reduced, and the level of money incomes falling as the multiplier sequence worked through. The prices of British goods would therefore fall, stimulating exports and discouraging imports until a balance was restored in British foreign trade. Theoretically, the countries earning the export surplus corresponding to the British deficit (and therefore receiving the gold which Britain was losing) would adopt the opposite policy—expanding credit, stimulating investment, raising incomes and prices, increasing their imports and reducing exports. The gold standard therefore claimed a 'built-in' stabilising mechanism which automatically maintained equilibrium.

Unfortunately, the picture was not so attractive in reality. In the first place, although countries receiving gold *could* expand credit, raise incomes and so on, there was no *obligation* on them to do so, and many countries in fact failed to do so because of the fear of causing inflation. The debtor countries were in a different position.

The loss of gold *forced* them to contract credit so long as the obligation of convertibility remained, and the failure of other countries to raise their price levels meant that debtors had to deflate all the more in order to get their prices back into line with the rest of the world. The system therefore had an inherent deflationary bias.

More important, however, was the assumption of price flexibility implicit in the gold standard system. When gold was lost and credit contracted, the resulting fall in the level of incomes was assumed to cause a similar fall in costs of production. As demand fell, workers accepted lower money wages, entrepreneurs lower prices and profits and landlords lower rents; if costs fell in the same proportion as prices, the same level of employment and output would be maintained. The level of incomes can fall in two ways, however; an unchanged number of people can receive lower money incomes, or a smaller number can receive constant money incomes. The gold standard theory assumed the first sequence, whereas the second actually occurred. Trade Unions resist wage-cuts, employers' cartels and agreements prevent cuts in prices and profit-margins, long-term contracts prevent cuts in rent and interest. Falling money demand in the face of rigidity of costs can result only in unemployment, and unemployment in fact occurred. Exports increased and imports decreased, as the theory required. But these things occurred not because costs of production fell; they occurred because the unemployed factors could not afford to consume goods, forcing both home and foreign producers to look elsewhere for markets.

What it comes to, then, is that a country engaging in international trade must reconcile itself to variations in *either* its exchange rate *or* its internal level of incomes; flexibility must exist somewhere. Given the resistance to cuts in factor-rewards which now appears to be a basic feature of a modern society equipped with powerful Trade Unions and Trade Associations, fluctuations in the level of incomes can only mean fluctuations in the level of employment. In a dynamic world where change and adjustment are constantly occurring, a country indulging in foreign trade cannot have complete stability of exchange rates *and* of the level of employment. The economic world faced this fact in the 1930's, decided that unemployment was the greater evil, and the gold standard was abandoned.

It is important to bear in mind that although the discussion so far has been conducted in terms of the gold standard, the basic conclusions follow with *any* system whereby a fixed exchange rate is maintained. For example, current British policy is to maintain a rate of £1 = $2·8 indefinitely into the future. Although the British monetary system is no longer based on gold, then, it still remains true that, if the exchange rate is to be fixed, the domestic level of incomes will have to adjust itself whenever trade conditions are unbalanced. To some extent, as later discussion shows, the problem is modified by the existence of exchange control and inconvertibility, but that underlying truth is substantially unaffected.

The argument in favour of freely fluctuating exchange rates, of course, is largely identical with that against fixed rates. Above all, a country which forgoes the attempt to maintain its currency at a fixed parity with gold or another currency is, theoretically at least, able to design its domestic policy with the primary aim of securing full employment at home. (Some of the practical limitations to this power are mentioned in the next chapter.) If imports exceed exports, it allows its exchange rate to fall, making its exports cheaper and its imports dearer; it makes no attempt to close the trade gap through a reduction in domestic incomes and prices, which would give rise to at least the threat of unemployment.

If the Central Bank does not preserve a fixed exchange rate by freely converting its currency at some specified rate, the determination of the currency's external value is left to the interplay of the forces of demand and supply in the market, just like any other price. Some attempt to give more precision to this process was made by the *purchasing power parity theory*. This stated that, left to the free play of market forces, the external value of currencies would settle at that level which made the purchasing power of a given unit of currency the same wherever it was spent. For example, if there was only one commodity, bread, and a loaf of bread cost 1s. in Britain, 20 cents in the United States and Frs. 50 in France, then the exchange rates of pounds, dollars and francs would have to be such that 1s. = 20 cents = Frs. 50—i.e. £1 = $4 = Frs. 1,000. If this were not so, the theory argued, it would be possible to purchase goods at a low price in one country and resell them at a high price in another, and this one-way trade would continue until the forces of

supply and demand brought prices into equality when expressed in the different currencies.

It is undoubtedly true that in the long run the external value of a currency will be determined primarily by the relative level of costs and prices in the country concerned. To relate exchange rates simply to overall purchasing power, however, is a gross oversimplification. The general price level takes account of a great many goods and services which cannot enter into international trade and whose prices cannot be brought into equality in different countries through the process of one-way trade mentioned above. Secondly, the existence of tariffs, subsidies, transport costs, and import restrictions may result in permanent discrepancies in the prices quoted for the same commodities in different countries, and these also will be incapable of removal through the flow of trade in one direction. Finally, the demand for and supply of a currency may be influenced by a good many causes having nothing whatever to do with price levels or current trade, such as government expenditure overseas, capital movements and speculative purchases and sales.

All that can be said, then, is that when the determination of exchange rates is left to the free play of market forces, the value of a currency will depend on all the factors underlying the purchase or sale of foreign currencies. In the long run, the cost of exports in relation to imports is bound to exercise a dominant influence; if the goods exported by a country are cheap relative to its imports, with the result that an export surplus is achieved, the demand of foreigners for its currency in order to pay for exports will be greater than sales by residents against foreign currencies, and the exchange rate will be driven up until equilibrium is restored. If a country has an import surplus, the external value of its currency will tend to fall. In the short run, however, the flow of capital, trade restrictions, or speculative movements may offset or even reverse these basic tendencies.

Although a variable exchange rate has the great advantage of freeing the domestic level of employment from complete dependence on foreign trade conditions, it must not be forgotten that fluctuations in the foreign exchange market have the severe disadvantage of disturbing and discouraging international trade for

the reasons already given. (They have other, even greater, disadvantages which will be described in the next chapter.) To some extent this damaging instability can be lessened through government intervention in the foreign exchange market, and it will be useful to examine the ways in which this may operate.

4. OFFICIAL INTERVENTION IN FOREIGN EXCHANGE

At the present time, governments intervene in the foreign exchange market in two main ways. The first is through an *Exchange Equalisation Account*, or similar institution. The aim of this official agency is to even out fluctuations in the external value of a country's currency which arise because of purely temporary causes rather than long-run trends. An exchange rate may oscillate because of seasonal selling pressures, or speculative sales made in the hope of buying back at a later date and at a lower value. The Exchange Account operates with a stock of the home currency and of the major foreign currencies. For example, if the exchange rate of the pound is tending to fall because of excess selling pressure, the British Exchange Account enters the market and buys pounds sterling in exchange for some of its stock of foreign currencies; conversely, if the pound is tending to appreciate unduly, the Exchange Account sells pounds and buys foreign currencies. By adding to the demand for or supply of pounds, it offsets what are believed to be temporary fluctuations and maintains a high degree of stability in the external value of the pound.

In the long run, of course, an Exchange Account cannot prevent a movement arising out of the fundamental realities of a country's economic position. If a country persistently runs an import surplus, an Exchange Account could prevent a fall in the exchange rate only by continuously adding to the demand for its own currency by selling foreign currencies in exchange for it. Ultimately, its stock of foreign currencies would be exhausted by this procedure; its powers would then be exhausted and the inevitable decline would occur in the external value of the country's currency. The major weakness of the Exchange Account system, indeed, lies in the difficulty of distinguishing in practice between short-term and long-term trends.

Since the United Kingdom is currently pursuing a policy of

maintaining a fixed exchange rate of £1 = $2·8, the Exchange Equalisation Account now functions as a means of maintaining this rate, rather than as an agency for offsetting purely temporary movements. It operates on 'free' markets (sometimes referred to as 'black' markets) in foreign financial centres such as New York or Zurich where sterling is bought and sold by persons not subject to the British exchange control regulations. The Account's object is to ensure that the rate of sterling quoted in these free markets does not diverge significantly from the 'official' rate of £1 = $2·80. If sterling fell to, say, $2·50 in such markets, no-one would bring dollars to the Bank of England for conversion into pounds if they could avoid it; they would prefer to buy their sterling at the 'cheap' rate. In this way Britain would suffer a loss of valuable dollar reserves.

The second form of official intervention is known as *Exchange Control.* The imposition of exchange control by a country means that residents of that country are required by law to hand over all their receipts of specified currencies to the Central Bank for conversion into the home currency at the official rate of exchange, and are permitted to buy such foreign currencies only through, and with the permission of, the Central Bank or Treasury. To the extent that exchange control exists, of course, it destroys entirely the foreign exchange market in the normal sense of the term.

The primary purpose of exchange control is the maintenance of a rate of exchange higher than that which would exist if demand and supply were allowed to work themselves out in freedom. The authorities must believe that, at the rate which they are seeking to maintain, people would wish to sell more of the currency than foreigners would wish to buy. If the United Kingdom was running an import surplus, for example, it could prevent the depreciation of the pound which would normally result by refusing to allow more pounds to be sold against foreign currencies than are currently being demanded by foreigners to pay for British exports. The natural result of this will be, of course, that imports into the United Kingdom will be cut down. If British buyers cannot obtain the foreign currencies necessary to pay for foreign goods, foreign suppliers would continue to send goods to Britain only if they were prepared to accept in exchange pounds which could be used to

purchase goods in Britain, but could not be transferred into their own or any other currency. Not many foreign producers would regard such a situation as satisfactory.

Like all impediments to foreign trade, exchange control is bound to cause some loss of real income because of the artificial restraint it imposes on international specialisation. Nevertheless, in certain circumstances exchange control can be justified, and may even be unavoidable. A country at war, for example, obviously cannot allow its residents to transfer scarce foreign exchange reserves to its enemies, or even to use them for relatively frivolous purchases. In view of the dislocation of trade which war involves, foreign currencies are scarce and valuable assets which countries will wish to use only in the interests of the war effort; only by the imposition of exchange control can the authorities ensure that this actually happens.

Another set of circumstances in which exchange control may be considered necessary is that in which the United Kingdom currently finds itself. In this instance, the demand for one particular type of currency (the U.S. dollar)—or, rather, the goods which can be bought only with that currency—is in excess of what can be paid for from Britain's current earnings of dollars. If exchange control was not adopted, the only solution would be a drastic devaluation of the pound *vis-à-vis* the dollar, and (for reasons discussed in the next chapter) the authorities prefer not to resort to this. Exchange control enables the expenditure of dollars to be restricted to what are considered to be essential uses.

Finally, exchange control may be used to prevent undesirable capital movements. In the 1930's many currencies suffered instability because of large-scale movements of capital motivated by the hope of speculative gains. Suppose I believe that the pound is going to fall from $5 to $4 in the near future, and have funds of, say, £10 million which I am prepared to gamble on my belief. I can sell the £10 million for $50 million, leave my capital as a bank deposit in New York until the devaluation occurs, and then sell the $50 million in exchange for £12·5 million at the new exchange rate. This does my bank balance a lot of good, but may do the United Kingdom a lot of harm; the devaluation may have occurred *because* people like myself expected (or even intended) it to happen and

acted accordingly. If this type of activity continues, currencies may develop a harmful and entirely unnecessary instability. Exchange control can at least reduce its extent.

5. INTERNATIONAL MONETARY INSTITUTIONS

Before concluding this brief survey of the foreign exchange market, mention should be made of two international institutions established after the end of the Second World War. The first of these is the *International Monetary Fund*, established in 1947. The aim of this organisation was to secure an orderly system of exchange rates which would maintain stable rates but would also make provision for currency devaluation or appreciation, if the Fund considered it necessary and consented to the change proposed by the country concerned. In this way it was hoped to avoid the excessive exchange-rate instability of the 1930's when devaluations tended to become competitive—a country would devalue in order to improve its trade balance, and other countries would promptly devalue *their* currencies in order to prevent the worsening in their trade position which would result if the first country succeeded in its aim. The result of a chain of devaluations of this kind, of course, would be that every country ended in the same relative position as it began, but international trade as a whole would have suffered as a consequence of the disturbance to the foreign exchange market.

The provisions for agreed adjustments in exchange rates by individual members of the Fund, however, ensured that a country facing a fundamental dis-equilibrium on its foreign trade would not be forced to resort solely to internal deflation. The experience of the gold standard has shown clearly that in the modern world the latter is simply not a workable policy. Amongst other things, the rules of the Fund also provided that member countries faced with temporary trade difficulties could borrow foreign currencies from a common pool, so ensuring that alterations in exchange rates need be used only as infrequent adjustments to meet long-run trends rather than for every difficulty of a temporary nature. In short, the Fund was designed to combine the advantages of the gold standard (free convertibility of currencies and general stability of exchange rates) with the advantage of flexible exchange rates (the ability to aim at full employment), avoiding the disadvantages of each.

Unfortunately, the provisions of the I.M.F. agreement have never come into full operation, primarily because Western Europe and most other non-dollar areas have not yet found themselves in a strong enough trade position to permit free convertibility of their currencies into all other currencies, and especially into the U.S. dollar. (In recent years, however, many countries—including the United Kingdom—have found the I.M.F. a valuable source of temporary supplies of gold and dollars with which to meet emergency drains on their own gold reserves.) When Britain and other non-dollar countries are able to maintain a stable exchange rate without the imposition of limitations on the use of their currencies, it is to be hoped that a full trial will be given to the imaginative and courageous provisions of the I.M.F. charter.

A second international institution, of a more limited scope, is the *European Payments Union* established in 1950 after a process of evolution in the preceding three years. Unlike the I.M.F., it *has* come into operation and has worked with considerable success. Membership of the E.P.U. is confined to countries of Western Europe and affiliated territories, such as the British Dominions (except Canada) and the colonies. The E.P.U. agreement provides for the monthly settlement of all transactions taking place between the members, completely free convertibility of member countries' currencies with one another, and for the extension of credit by members enjoying trade surpluses to those running trade deficits. The provisions for free currency convertibility and for temporary loans to finance trade deficits have immensely stimulated trade between the various members of the E.P.U., and have therefore helped to reduce Europe's dependence on supplies from countries in the dollar area.

The provisions governing credit have been adjusted from time to time so as to encourage a gradual movement on the part of E.P.U. members towards a full convertibility of their currencies into dollars as well as E.P.U. currencies. Each member has a *quota* which defines the total deficit it can run under the Union provisions, or, conversely, the total surplus it can be required to run under those provisions. The quotas are divided into two parts: one financed by borrowing from or lending to E.P.U., the other financed by the payment of gold to, or receipts of gold from, E.P.U. (Gold is little

more than a convenient form of the U.S. dollar nowadays, since one ounce of gold is convertible into $35, and is desired solely for that reason.) For example, if Britain imports £100 million more from other E.P.U. countries in a given month than it exports to the rest of E.P.U., it would be required to pay *part* of this deficit in gold but would be granted credit by the Union for the rest. A country exporting more than it imported would receive part of the balance in gold and would be required to take the rest in the form of a credit balance in E.P.U.

Since 1950, the gold element in the quotas has been progressively increased. At the commencement of the Union, a country was required to pay in gold only 40 per cent of its deficits in E.P.U., up to the extent of its quota. (If a country used up all its quota, further deficits theoretically had to be paid 100 per cent in gold; in practice its quota was extended and 50 per cent of further deficits were paid in gold, credit being extended to cover the other 50 per cent.) By 1956 the gold element had increased to 75 per cent, so that a debtor country still within its quota can cover only 25 per cent of its deficits by borrowing from the E.P.U. This progressive stiffening in the terms of settlement has discouraged member countries from living beyond their means at the expense of other E.P.U. members, and they have been stimulated towards an approach to complete convertibility of their currencies.

Ultimately, most people look forward to the day when Europe and the rest of the non-dollar world is able to pay its own way in foreign trade without having to resort to controls and restrictions on the use to which its currencies may be put. Full convertibility is one requirement for the maximum exploitation of the potentialities of free international trade; another is the removal of fiscal and quantitative obstacles to trade—tariffs and quotas of all kinds. Yet freedom in international trade carries risks and difficulties as well as benefits. No country will nowadays allow its domestic prosperity to be completely at the mercy of its external trade, and in an imperfect world the most that can be hoped for is a movement towards a situation in which the restrictions on foreign trade are at the minimum level genuinely necessary for the preservation of a fully-employed and dynamic internal economy.

FURTHER READING

R. S. SAYERS, *Modern Banking*, Ch. 6.
B. TEW, *International Monetary Co-operation, 1945–52*, (Hutchinson, London, 1952), Chs. I–II, VI–VIII.
R. F. HARROD, *International Economics*, Ch. V.
N. CRUMP, *The ABC of the Foreign Exchanges* (11th ed., Macmillan, London, 1951), Chs. VI–XIII.

CHAPTER 31

The Balance of Payments

1. THE BALANCE OF PAYMENTS ON CURRENT ACCOUNT

Just as a company draws up every year a statement of its income and expenditure, so a country draws up an annual statement of its current transactions with foreign countries; this is known as the *balance of payments*. The word 'balance' is to be taken here in the sense of a relationship rather than as an exact equality, since it is most unlikely that the goods and services purchased by residents of a country from abroad will be exactly offset by their sales of goods and services to foreigners.

The balance of payments can be drawn up on the basis of *current* transactions (the receipts and expenditure of income which go on year after year), or it can include *capital* transactions (the lending of sums to other countries, borrowing from other countries, or transfers of assets accumulated in the past). The degree to which a country is living within or beyond its current means is best observed from the balance of payments on current account. This includes what are known as *visible* and *invisible* trade. By visible trade is meant the tangible goods which are exported from or imported into a country. The major imports of the United Kingdom, for example, are foodstuffs, such as meat, grains, fruits, coffee and tea, and raw materials—especially petroleum, cotton, wool and metal ores. Britain's major exports are machinery, cars and lorries, chemical products and textiles of all kinds. The difference between a country's expenditure on its visible imports and its earnings from visible exports is called the *balance of trade*, which is said to be 'favourable' if the value of exports exceeds that of imports and 'adverse' if the opposite is true. (The *balance* of trade should never be confused with the *terms* of trade. The former expresses the relationship between the total *amounts* spent on imports and received

from exports; the latter refers to the relationship between the *prices* of exports and imports.)

The 'invisible' items included in the current balance of payments are the services paid for by residents of a country or performed on behalf of foreigners, and the property incomes flowing into or out of a country; they are items of income or expenditure which cannot be physically seen or handled at Customs barriers as can the commodities comprising visible trade. The main items involved are earnings from shipping and insurance services, expenditure of tourists outside their home country, government expenditure in foreign countries on embassies, maintenance of troops etc., and the interest and dividends on investments in foreign countries. Under each heading, of course, a country may *receive* payments from abroad or *make* payments abroad; the United Kingdom, for example, earns a considerable income from the carriage of goods for foreigners by British ships, but it also pays out substantial sums for the transport of British goods in foreign ships. Under each heading, therefore, the net balance of earnings of a country may be positive or negative; it will be the balance of both receipts from other countries and payments to other countries.

If the balance of trade is added to the net receipts from invisible exports and imports, the resulting total constitutes the balance of payments on current account; this is said to be *favourable* if total receipts exceed total payments, and *adverse* if payments exceed receipts. It follows that the balance of trade has no special significance when taken in isolation. A country may continuously export more goods than it imports and yet have a deficit on its overall current balance of payments, if its net expenditure on invisible account is more than sufficient to offset the trade surplus. By contrast, the United Kingdom has run an adverse balance of trade since the middle of the nineteenth century; until recently, its invisible earnings have been consistently more than sufficient to compensate for the excess of physical imports over physical exports.

2. INTERNATIONAL CAPITAL MOVEMENTS

If a country has an overall deficit on its current balance of payments, naturally that deficit must be financed in some way. Foreign suppliers do not send goods without being paid for them. Conversely,

a country with a current payments surplus must have used its excess revenues in some way. A deficit or surplus on the current balance of payments, in other words, must be financed by means of transfers of *capital*. Every purchase and every sale must be paid for in some way; if foreign purchases cannot be fully paid for out of the proceeds of exports, then the difference must be paid for from capital. Put in another way, if the amount of a currency being transferred into foreign currencies in order to pay for imports exceeds the amount of foreign currencies being earned from exports, then on balance the deficit country must be running down its holdings of foreign currencies unless such currencies are being obtained by some means other than trade.

This last sentence indicates one form of capital movement which may be adopted to finance a surplus or deficit on the overall balance of payments—the transfer of foreign currencies (or gold, with which any foreign currency can be bought) from the reserves of the debtor country to those of the creditor country. If the deficit is thought to be a purely temporary one (it may be nothing more than the outcome of a regular seasonal trend) then there is nothing especially alarming about the loss of foreign exchange reserves in this fashion. If the deficit persists, however, other remedies will have to be sought; a country cannot go on losing its exchange reserves because eventually they will all be used up. (Britain's stock of gold and dollars—the form in which reserves of foreign exchange are always held, because of the convertibility of gold and dollars into any other currency—were equivalent at mid-1955 to £957 million, enough to pay for less than four months' visible imports.)

A second form of financing is the use of outright grants by foreign countries. These may extend over a number of years, so that there are some grounds for treating them as income rather than capital receipts; they are seldom permanent features of a country's payments or receipts, however, so that they are best treated as capital items. In the period since the Second World War, the government of the United States has given considerable sums of money to Western Europe and other regions in order to assist them in their post-war reconstruction (the Marshall Plan of 1948–51) and, more recently, to assist in re-armament programmes under N.A.T.O. and similar arrangements. These grants were made by means of

credit balances opened up in the Federal Reserve Bank in the names of the foreign governments; the countries concerned could then use these balances to meet their deficits of foreign exchange.

If a deficit cannot be financed by means of grants, and is not of a temporary nature, than a third means of financing it must be adopted—the raising of loans abroad. If a government or company can borrow from a foreign government or in a foreign capital market, its country will come into possession of foreign currencies, which can be used to meet the payments deficit. At first sight it may seem dangerous to finance current consumption of imports out of borrowing just as an individual is ill-advised to finance his current expenditures by borrowing from the bank. This is not necessarily so, however. If the foreign borrowing enables a country to increase its productive powers through capital investment, then the loan can be repaid from the extra income generated by the investment. Almost all creditor countries—the United States and Canada, for example—were at some time debtor countries, running import surpluses financed by foreign borrowing and using the resources thus placed at their disposal to increase their productive capacity. A country which borrows abroad to finance current consumption *without* carrying out compensating investment, however, is heading for disaster; its balance of payments problem in the future will be made even more intractable because of the interest payments and capital repayment which it will have to make to the country from which it borrowed.

As has already been observed, capital may move from one country to another in the hope of speculative gains from movements in exchange rates, while short-term capital may also be shifted from country to country in order to take advantage of higher interest rates in different financial centres. Movements of this kind have no real significance from the balance of payments point of view, because against the loss of reserves as capital moves out there will be a corresponding rise in foreign assets when the funds are used to increase the bank deposits or securities held in foreign centres by residents of the country. The primary significance of short-term capital movements of this kind lies in their possible effect on the exchange-rate of a country's currency; in this context, as has been seen, they may have most undesirable consequences.

3. THE BALANCE OF PAYMENTS OF THE UNITED KINGDOM

It will be useful to relate the foregoing general discussion of the balance of payments to the actual position in which the United Kingdom has found itself in recent years. The main features of the overall balance may be seen from Table XV. The British balance of trade has been consistently adverse for many years, and it did not

TABLE XV: *The Overall Balance of Payments of the U.K.*, 1946–1955
(*£ million*)

	1946	1954	1955
1. *Current balance*			
Imports (visible)	1082	3009	3413
Exports (visible)	917	2817	3061
Balance of trade	−165	−192	−352
Net invisible earnings:			
Shipping	+ 28	+150	+120
Interest and profits	+ 62	+ 73	+ 59
Tourism	− 29	− 6	− 12
Government expenditure abroad	−323	−174	−181
Other	+129	+304	+219
Total invisible earnings	−133	+347	+205
Total current balance	−298	+155	−147
2. *Capital movements*			
Loss of reserves (− = increase)	− 54	− 87	+229
U.S. Defence Aid	—	+ 50	+ 44
Long-term borrowing (+) or lending (−)	+324	−201	+ 40
Short-term borrowing (+) or lending (−)	+ 28	+ 83	−166
Total	+298	−155	+147

Source: *Annual Abstract of Statistics, 1956* (London, H.M.S.O., 1956), Tables 266–7.

vary greatly between 1946 and 1954. The absolute value of Britain's physical trade was trebled over this period, however, partly because of the marked rise in world prices and partly because of a substantial expansion in the volume of British foreign trade. The

balance on invisible account, however, has shown a vast improvement since the War. In 1946, Britain's invisible trade was still adversely affected by the high rate of British military expenditure abroad and by the sadly depleted condition of her merchant marine. By 1954, however, government overseas expenditure had been substantially reduced (although it remains a considerable burden on the balance of payments), while the merchant marine had regained some of its pre-war position as a major international carrier. Another source of significant improvement was the rise in the royalty earnings of British oil companies operating abroad; in the table these are included in the item 'other', which also covers the foreign earnings of British banks, insurance companies and air lines.

The capital account naturally reflects this switch in the overall state of the balance of payments between 1946 and 1954. In 1946 the deficit of some £300 million in the current balance was financed mainly by borrowing at long-term, especially from the United States and Canada, and by the sale of British assets abroad. In 1954 Britain was able to use its current surplus to increase its reserves of foreign exchange and its long-term assets. This emphasises the fact that to invest in overseas countries, including under-developed territories in the Commonwealth and elsewhere, Britain must succeed in exporting more goods and services than it imports. To assist the development of other countries, in other words, the British public must accept the fact that it will need to forgo the use of some of its income in current consumption or investment at home.

The deterioration in Britain's trading position during 1955 is clearly brought out by the figures shown in Table XV. The basic explanation of this deterioration is to be found in the excessive pressure exerted by home demand on current output; too many resources were being absorbed into consumption and domestic capital formation. The external situation was substantially improved in 1956, and there are good reasons for believing that 1954 was more representative of Britain's current trading position than the year 1955 in which inflationary pressure developed within the economy to an exceptional extent. Nevertheless, the fact that the British current payments balance could swing from a surplus of £155 million in 1954 to a deficit of £147 million in 1955 clearly indicates that a stable position of external solvency has not yet been estab-

lished. (Some of the reasons for this are discussed in the next chapter.) The question of suitable remedial measures to meet a payments deficit therefore remains one of considerable importance.

4. DEVALUATION, DEFLATION AND IMPORT CONTROLS

The relative advantages and disadvantages of fixed and variable exchange rates were discussed at some length in the previous chapter. At first sight the case against fixed exchange rates is overwhelming in the modern world; if a deficit in the balance of payments has to be corrected only by a policy of deflation, this is likely to impose burdens which are nowadays considered intolerable. The experience of 1949 and after, however, has brought home the weakness and dangers involved in the alternative policy of currency devaluation. In that year, Britain and other countries found themselves in particularly severe payments difficulties, and resorted to the expedient of devaluation; the pound sterling and other currencies were reduced in value by some 30 per cent *vis-à-vis* the U.S. dollar.

That this type of policy introduces an undesirable instability into the foreign exchange market has already been emphasised. The experience of 1949 and after emphasised other weaknesses of the policy of varying exchange rates. The first of these is the implicit assumptions it makes concerning elasticities of demand for exports and imports. Devaluation will obviously make a country's exports cheaper. But it will increase the country's total export earnings (the important thing from the balance of payments point of view) only if the quantity of goods exported increases *in a greater proportion*—i.e. only if the elasticity of demand for its exports is greater than unity. If the price of exports falls 30 per cent and the quantity exported rises only 20 per cent, total export earnings will fall by 16 per cent; devaluation will have *reduced* the country's total export earnings. Similarly, devaluation will raise the price of imports, but the quantity of goods imported will fall substantially only if the elasticity of demand is significantly greater than unity. For a country like Britain, at any rate, it is by no means certain that the relevant elasticities are such as to result in any marked improvement as a result of devaluation.

A second difficulty with devaluation is that by raising the cost of

imports to home producers it may result in a rise in the cost of exports—the imports may be raw materials which are subsequently embodied into manufactured exports. The rise in export prices, of course, will automatically lessen the advantage bestowed by devaluation. Finally, and most important, the rise in the price of imported foodstuffs resulting from devaluation may exert such a marked effect on the general cost of living that demands for wage increases are immediately inspired. This may set off the familiar wage-price spiral and end in a general rise in the level of costs and prices which offsets, or more than offsets, the initial advantage in foreign markets secured by devaluation.

The post-war experience of the United Kingdom, and other countries of Western Europe, has shown clearly these weaknesses of a policy of devaluation. Despite the devaluation of 1949, Britain was again experiencing a balance of payments crisis of great severity in 1951; this was due to a large extent to the rise in wages and prices between the two years which substantially wiped out the effects of devaluation and was unquestionably stimulated by it. If adjustments in income levels through deflation are ruled out, and devaluation suffers such pronounced disadvantages, how are balance of payments deficits to be corrected?

It must be admitted that in general terms such a problem is insoluble. Difficult and unpleasant though it may be, countries will *have* to find some way of adjusting their internal price level when faced with a deficit on their balance of payments if they are not to run the risks inherent in the manipulation of exchange rates. This is not quite the verdict of despair that it might seem at first sight. So long as the general upward trend in money incomes and prices continues throughout the world—and it shows little sign of ending at the moment—a debtor country will be able to bring its costs back into line by simply holding its price level constant while that of the creditor countries goes on increasing, without needing to attempt an absolute *reduction* in the level of money incomes. The same effect of a relative cheapening of exports will be secured. For example, suppose a country succeeds in maintaining a stable level of incomes and prices at a time when world prices in general rise 10 per cent. Its exports would then be 10 per cent cheaper, in comparison with the products of the rest of the world, without any actual fall in the

price of its exports. Similarly, if productivity is increasing—and, again, the upward trend is fairly constant in most countries—costs can be reduced by merely preventing a rise in factor-incomes, resulting in a fall in unit costs of production without the attempt to force a reduction in money wages and profits in absolute terms. But even the task of achieving stability at a time of rising prices in the outside world, or of increasing productivity, is an immensely difficult one; it must be admitted that no country yet provides convincing evidence of having mastered it. Nevertheless it is a good deal nearer practical possibility than the task of reducing incomes in absolute terms, and if international trade is ever to regain a reasonable degree of freedom and stability it is a task which will have to be accomplished.

One difficulty with policies of this kind—securing adjustment of a deficit through price stability at a time of external inflation or rising productivity—is that it must necessarily take some considerable time to achieve success. If world prices are rising at a rate of 4 or 5 per cent a year, or productivity at 2 or 3 per cent a year, the costs of a country pursuing price stability will fall in relation to those of the rest of the world at a pretty slow rate. Several years may elapse before the movement has progressed far enough to close a substantial payments deficit. If a country's exchange reserves are too small to sustain the outflow which will continue in the meanwhile, import controls—either direct quantitative restrictions on specified imports, or the use of exchange control measures —may be resorted to.

The advantage of such controls is that they enable a country to ensure that imports are restricted to the level it can currently afford, and, furthermore, that the imports allowed in are those giving the highest degree of benefit to society as a whole. Just as in a war foreign currencies are mobilised in the interests of the war effort, so in peace direct controls can ensure that foreign currencies in short supply are not being employed to purchase luxury goods rather than essential foodstuffs and raw materials. Exchange control is still employed in Britain and elsewhere to restrict expenditure on goods from the dollar area to essential purchases which assist the economy as a whole, and which cannot be satisfactorily replaced by goods produced at home or in non-dollar areas.

As a short-term measure to permit readjustment within the economy, import controls have much to be said for them. There is no other way of ensuring that scarce currencies are not squandered on Cadillacs and nylon stockings while bread has to be rationed because of a shortage of wheat, or industry is hampered by a lack of raw materials. As a permanent feature of the economy, however, they suffer from all the disadvantages mentioned earlier in connection with protection. In particular, they may foster inefficiency by allowing home industries to continue and expand without the foreign competition which would force them to hold down costs to the absolute minimum. In any case, controls can never be a substitute for a policy aimed at the correction of the fundamental causes of a deficit. Insofar as they succeed in reducing imports, the supply of goods available in the home market at a time of full employment is thereby reduced, so that the pressure of demand on prices is accentuated. There are at least some reasonable grounds for arguing that in the last resort there is no satisfactory substitute for a policy which aims directly and honestly at restraining internal demand within the limits which the country can afford.

FURTHER READING

R. F. HARROD, *International Economics*, Chs. VI–VIII.
B. TEW, *International Monetary Co-operation*, Chs. III–IV.
J. E. MEADE, *Planning and the Price Mechanism*, Ch. V.
W. A. LEWIS, *The Principles of Economic Planning* (Dobson, London, 1948), Ch. V.

CHAPTER 32

Britain in the World Economy

1. THE INTER-WAR YEARS

Before the outbreak of the first World War, the United Kingdom enjoyed a dominant position in the international economic scene. It was the leading industrial and commercial power of the world, and was able to invest abroad to the extent of some £100–£150 million each year (equivalent to, say, £350–£500 million at present-day prices). Signs of a decline in the relative importance of the United Kingdom were not lacking at the beginning of the century, but the leadership of London in international economic and financial affairs was unquestionable. Britain's international economic position was transformed, however, by the war of 1914–18. First, the concentration on the war, and the dislocation of transport, necessarily involved the loss of many foreign markets; these markets consequently became dependent on supplies from other sources, including their own industries which were stimulated by the interruption of the flow of imports from Britain. Secondly, a substantial number of British foreign investments were sold in the interests of the war effort, and the invisible account of the balance of payments was depleted by the income which was thereby permanently lost. Thirdly, Britain suffered the inflation almost inevitably experienced by countries at war. Finally, and perhaps of greatest significance, the United States emerged from the war with a vastly expanded industrial and agricultural capacity and became the world's major creditor.

In the decade following the first World War, then, the United Kingdom was struggling to regain external solvency in a world in which its trading position had worsened drastically in comparison with 1913. The problem was rendered even more difficult by the return to the gold standard in 1925 at the pre-war parity between

the pound and the dollar. Given the inflation of British prices and the enormous increase in American productivity, this was an error of the first magnitude. Coupled with the fact that Britain's traditional exports of coal and cotton goods were undergoing a secular decline—oil and new European coal deposits were replacing British coal, while Japanese and Indian textiles were invading such markets for foreign textiles as remained—this unduly high valuation of the pound sterling created an almost insoluble problem for British export industries. The departure from gold in 1931, and the development of new export industries—especially vehicles and high-quality engineering and chemical products—eased the situation to some extent during the 1930's. Had it not been for the tremendous improvement in the British terms of trade during the 1930's, however—the improvement was one of 20 per cent between 1929 and 1938—the British balance of payments would have been in a desperately weak condition throughout the inter-war years.

As it was, during the 1930's Britain had ceased to be a lender to the outside world on any significant scale. The loss of traditional export markets, together with the increasing rigidity in the economy resulting from associations amongst both workers and employers, maintained a chronic element of unemployment. Even in the 'boom' year of 1937, about 11 per cent of the insured labour force in the United Kingdom was unemployed.

2. THE POST-WAR PROBLEMS

So far as its permanent economic effects were concerned, the results of the war of 1939–45 can be said to have been similar in kind to those of the war of 1914–18, and perhaps greater in degree. Once again established export markets were lost; once again the industrial production of countries previously content to import British goods was stimulated; once again the economy of the United States, so far from suffering the damage and neglect experienced by British and European industries, was stimulated by the war; once again British investments abroad were depleted and invisible earnings lost.

Three special features of the post-1945 economic situation of Britain are worth note. The first was the vitally important change in the British terms of trade. In 1947 they were 16 per cent less

favourable than in 1938, so that to pay for its current imports Britain had to export in the later year some £250 million more exports than would have been necessary if the terms of trade had remained at their pre-war level. And, as was noted earlier, the terms of trade are slightly less favourable to the United Kingdom now than they were in 1947.

The second feature of the 1940's and 1950's has been the persistence of a substantial 'dollar gap'—i.e. the balance of payments disequilibrium has been especially marked *vis-à-vis* the dollar area (mainly North and Central America). As Table XVI shows, until 1955, and with the single exception of 1951 (when the terms of trade deteriorated violently), Britain has secured a substantial surplus in trade with non-dollar areas; with the dollar area, however, and especially the United States, it has consistently run a substantial deficit.

The causes of this dollar deficit, which has been experienced in varying degrees of intensity by almost every non-dollar country, have already been indicated. The United States economy has been

TABLE XVI: *Regional Balances of Payments on Current Account of the U.K.**

	With dollar area	*With all other areas*	*Overall*
1949	−296	+327	+ 31
1950	− 88	+388	+300
1951	−440	+ 33	−407
1952	−294	+420	+126
1953	−116	+191	+ 75
1954	−137	+292	+155
1955	−251	+104	−147

* Defence Aid excluded throughout.

Source: *Annual Abstract of Statistics*, No. 93, 1956 (H.M.S.O., London, 1956), Table 268.

stimulated by two wars, in contrast to the economies of other countries which suffered severe shocks from them in the form of lost markets, large-scale transfer to war production, and capital wastage by physical damage and a lack of proper maintenance and replacement. The wars, however, have merely emphasised the productive

superiority of the American economy. That superiority, which arises from the possession of vast natural resources, to some extent the existence of a domestic market great enough to permit the exploitation of all the potential economies of large-scale operation, and the high degree of mobility and flexibility amongst both its labour force and entrepreneurs, would have asserted itself in world trade even if the two great wars had never occurred.

The existence of a deficit with the dollar area is not a new phenomenon in the British balance of payments; it existed before the war. Previously, however, Britain was able to offset this dollar deficit with the surplus earned in other regions; the latter could be converted into dollars because of the surplus which those other areas, in turn, earned with the dollar area. The dollar gap was closed, in other words, by triangular trade. An example may help to clarify this point. Suppose that Britain has an overall deficit of £20 million in its trading with the United States but a surplus of £20 million in its trade with Belgium. *Provided* that Belgium runs a surplus with the United States (or for some other reason has a comfortable supply of dollars coming in) Britain could convert the £20 million surplus earnings of Belgian francs into dollars, and use the dollars to settle its deficit with the United States. In pre-war years, the position was actually that Britain earned a surplus with the rest of the Sterling Area (discussed in a later section); since the rest of the Sterling area managed to earn a surplus with the dollar area, the dollar surplus of the rest of the Sterling Area could be used to settle Britain's dollar deficit. The difficulties of the post-war years have arisen because the rest of the world no longer earns a dollar surplus (for the reasons already given) which can be used to offset their deficit with Britain and therefore to meet Britain's dollar deficit.

The dollar gap is the fundamental cause of Britain's continuing inability to permit the full convertibility of the pound sterling. Because most of the world would like to buy more from the United States than it does at present, free convertibility would probably result in a rush to convert pounds into dollars, and in the purchase from the dollar area of many commodities at present bought within the Sterling Area. This would naturally result in a continuing drain of British dollar reserves, and these would be exhausted before

long. Convertibility of the pound can be safely embarked upon, then, only when Britain and the other countries using the pound as their currency are able to earn a consistent surplus in their trade with non-dollar areas which can be used to offset any deficit which they still run with the dollar area.

The third, and perhaps most fundamental, difference between Britain's post-war position and that of earlier days is that British policy must now grapple with the immense task of attaining solvency in foreign trade *without endangering full employment at home.* This dual aim of economic policy has never before existed and it raises issues which have recurred throughout the discussion of prices and foreign trade in previous chapters. A balance of payments deficit cannot now be corrected by deflation, since this would endanger full employment. On the other hand, the alternative policy of devaluation loses much of its effectiveness if it results in the wage-price spiral discussed at the end of the preceding chapter. The existence of full employment, of course, bring with it dangers of recurring inflation, if Trade Unions and employers use advantages conferred on them by full employment to maintain a continuous rise in wages and profits without regard to the level of productivity. The excess demand associated with an inflationary situation inevitably spills over into increased imports and the reduction of exports as goods are drawn into the buoyant home market.

So the discussion returns to the fundamental problem of reconciling full employment with relative stability of prices, a problem which has cropped up more than once in the preceding pages. A country which is able to provide almost all its needs from within its own borders is at least free to choose between the evils of inflation and the evils of unemployment, if its population is unable or unwilling to discover some means of self-restraint by which both may be avoided. A country like Britain, however, dependent as it is on imports for essential foodstuffs and raw materials, has no such clear choice. A chronic inflation (in comparison with the rest of the world) is a course which is simply not practicable for a country highly dependent on export earnings. Ultimately, then, the problem is one of convincing the population as a whole that the exploitation of an assured market for products, or labour, in order to attain

sectional advantages is a process which is as anti-social as burglary or arson; given a general awareness of that fact, the problem will then be one of translating that awareness into effective policy.

3. THE STERLING AREA

In the course of the previous section several references have been made to the Sterling Area. As is evident from the expression, it refers to those countries which use one form or other of the pound sterling as their currency. Until the collapse of the gold standard in 1931, the Sterling Area as such could hardly be said to have existed. All currencies were freely convertible into one another at exchange rates fixed within narrow limits, so that it made little difference to an exporter what currency his exports were sold for.

With the ending of the gold standard, however, countries had to decide whether to allow their currencies to fluctuate in value on the exchange markets, or to maintain them in some stable relationship with gold or another major currency. A large number decided to maintain a fixed relationship with the pound sterling—especially, of course, the countries of the Commonwealth. Why should they do this? The reason was primarily that they had developed especially close economic relationships with the United Kingdom over the course of their history; hence the maintenance of a fixed parity between their currency and that of the United Kingdom—the pound sterling—would bring to at least part of their foreign trade (usually the most important part) all the advantages of a stable exchange rate, which the world as a whole had lost through the collapse of the gold standard.

The economic relationships involved were not merely those of a market for exports and a source of imports, although the United Kingdom was both their major market and their major supplier. Many of the territories concerned had been developed by British colonists who had naturally remitted their profits to, and transacted most of their business through, British banks, finance houses and insurance companies. London had therefore become the banker for most of these countries, and their main foreign assets were held in the form of sterling securities in London. Further, the major commodity markets were in the United Kingdom—the Liverpool Cotton Exchange, the London tin and tea exchanges etc.—so that

the exports of these countries were frequently sold through London, and in terms of sterling, even though the final market was elsewhere. Finally, these countries had been developed initially with the aid of British capital, and had come to regard Britain as the major source of capital for further development. Since the major parts of their trade, assets and debts were therefore expressed in terms of sterling, it was advantageous to these countries to have their currencies maintained in a stable relationship with the pound sterling.

At the outbreak of the war in 1939, it was necessary for the purposes of the United Kingdom exchange control, which was then instituted, to define the precise extent of the Sterling Area; its membership has not changed substantially since. Countries which belong to the Sterling Area currently include all members of the British Commonwealth except Canada, and, in addition, Iceland, Iraq, the Irish Republic, Jordan and Libya. In 1955 they accounted for more than a quarter of total world trade.

In the post-war world of trade restrictions and exchange control, membership of the Sterling Area has proved of considerable benefit to the countries belonging to it. As all the members have currencies which are directly connected with sterling, exchange-rate fluctuations have been absent within the area; when the United Kingdom devalued the pound in 1949, all the other members of the Area (with the exception of Pakistan) devalued their currencies more or less equally, so that their relative positions were unaffected. Further, there have been no restrictions at any time on the conversion of one sterling currency into another, so that payments for current or capital transactions within the Sterling Area have been completely free. These two facts—complete stability of relative exchange rates and free convertibility—have immensely stimulated multilateral trade between members of the Sterling Area at a time when international trade generally was hampered and distorted by restraints on trade and the conversion of currencies. Another feature of the Sterling Area is the practice amongst the members of holding the bulk of their gold and dollar reserves in a central pool in London, rather than holding separate reserves in their own central banks. This has enabled the reserves of the Area as a whole to be used to the best advantage, just as a bank with

many branches can economise on its cash reserves by offsetting unusual outgoings in one region by receipts from other regions. Each member has been able to call on the accumulated reserves of the entire Sterling Area at moments of stress.

Britain's position as the centre of the Sterling Area has been subject to a good deal of strain in recent years, however. The inconvertibility of the pound sterling into dollars has naturally reduced the attractiveness of London balances as an internationally-accepted asset; some of the Sterling Area member countries which have managed to attain a surplus in their trade with the dollar area have therefore questioned the advantages they derive from contributing gold and dollars to the central pool in exchange for sterling which can be put to limited uses only. (Unhappily, the biggest dollar-*earners* have been the colonies, while the biggest dollar-*spenders* have been certain Dominions, so that the natural order of things whereby the richer countries assist the poorer has been reversed.) Further, Britain's leadership in the past has owed much to its ability to supply capital for development; this presupposes a surplus on the overall balance of payments and, as has been seen, the United Kingdom is still far from being a consistent creditor on a scale sufficient to supply the capital needed by under-developed members of the Commonwealth. Sterling Area members in need of large and regular inflows of capital, therefore, have had some cause to wonder whether it might not be more to their advantage to link their currencies to the dollar, because the United States is now the world's major international lender.

The stress has not been entirely one-sided. Britain, too, has undergone considerable strain as the centre of the Sterling Area. Although its supply of capital to the rest of the Area has been subject to instability, and in any case has been inadequate in the post-war years, it has nevertheless lent large sums to other members in the form of exports which have been paid for, not with imports in exchange, but simply with a running-down of the sterling balances held in London by the countries concerned. It has been estimated that, during the years of the Marshall Aid programme, the United Kingdom provided more capital through such 'unrequited exports' to other sterling countries than it received itself in the form of Marshall Aid. Naturally this capital outflow has added considerably

to the internal and external difficulties of the United Kingdom economy.

Further, the continued inconvertibility of the pound sterling owes much to Britain's position as banker for the entire Sterling Area. On balance, as has been noted, Britain has secured an overall surplus in its trade during most recent years. If its surplus with the non-dollar world (and especially the rest of the Sterling Area) had been settled in gold and dollars, the British dollar gap could have been filled easily enough, and the convertibility of its currency could have been maintained. The gold reserves in London, however, have to meet the drain arising not only from the United Kingdom purchases of dollar goods but also that caused by dollar outlays by *all* members of the Sterling Area. The pound can become convertible when not only the United Kingdom but the Sterling Area as a whole is earning enough dollars to cover its dollar expenditures. Hence, unsound policies on the part of important members of the Sterling Area can prevent full convertibility of the pound, despite a position of strength and solvency attained by the United Kingdom itself. Is it really worthwhile for Britain to incur the losses and hardships implied by continued inconvertibility for the sake of one or two Sterling Area members which persist in overspending? Can the British economy, weakened by two wars and undergoing severe internal stresses, bear the strains which leadership of the Sterling Area imposes on it?

It is to be hoped that it can and will. The entire Sterling Area has benefited in the past from the stimulus to trade between members provided by its existence, and by the continual exchange of views and co-ordination of policy which have occurred—even if only imperfectly—within its framework. At bottom, the Sterling Area is the economic expression of the strong political ties which have been shared by its members, and which give every sign of having been an influence of great value in the modern world. If Britain abandoned her leadership of the Sterling Area, no member has the historical, economic or constitutional background which would make it acceptable to the rest of the Commonwealth as an alternative leader. Britain itself, on the other hand, could never hope to wield as a single unit the power it now possesses in international economic affairs by virtue of its position as the centre of the Sterling Area.

Some might regard the decline of British influence in the international economic (and therefore the political) scene as a desirable development from the point of view of the world as a whole and of Britain itself. The majority would hardly share this view, and in this it would be supported by a good deal more than mere sentiment.

FURTHER READING

W. A. LEWIS, *Economic Survey, 1919–1939* (Allen & Unwin, London, 1949), Chs. V, XI–XIII.

SIR DENNIS ROBERTSON, *Britain in the World Economy* (Allen & Unwin, London, 1954).

H. G. JOHNSON, *The Overloaded Economy* (Univ. of Toronto Press, Toronto, 1952).

B. TEW, *International Monetary Co-operation*, Chs. X–XI.

INDEX

Acceptance houses, 281
advertisement, 148, 162, 165, 172, 176–9, 182, 184
aggregate demand:
  and employment, 313–15
  components of, 315–17
amortisation, 98
appreciation, 385, 392
arbitrage, 381
arbitration, 212–13
Argentina, 372
Associated Electrical Industries, 120
Australia, 372
average revenue, 149–52, 157–9

backwardation, 286
balance of payments:
  nature of, 396–405
  of United Kingdom, 400–2, 406–11, 413–14
balance of trade, 382–5, 392, 396–7, 400
Balogh, T., 275, 288
Bank Charter Act, 264–6
bank advances, 109, 271–3, 285, 288, 301–8
bank credit:
  as money, 260–2
  control of, 288, 299–309, 247–8, 385
  creation of, 256–60, 274–5
bank deposits, 273–4
Bank of England, 259, 261, 263–71, 274–6, 282, 299–306, 385, 390
Bank of England (Nationalisation) Act, 268, 301
Bank rate, 299–300, 302–4
banks:
  cash ratio of, 271, 275, 294, 300–2, 304–6
  historical development of, 263–5
  operations of, 260–1, 263–76, 280–2, 285, 294, 300–9, 401, 411
  secondary liquidity ratio of, 272–3, 305
barter, 7, 251–4, 379
B.B.C., 116, 234
Beacham, A., 93, 123, 170
bears, 287
bills of exchange, 269, 271–3, 276–82, 299–300, 302–5
black markets, 23, 390
Board of Trade, 169, 308
bond, 283, 291–2, 296–7
boom, 307–8, 341–2, 374, 407
Bowen, I., 82
British Celanese, 122
British Electricity Authority, 128, 163
British Motor Corporation, 121, 297
British Overseas Airways Corporation, 234
brokers, 280–1, 285–6, 300, 381–2
B.T.H., 120
budget, 347–50
budget line, 48–9
building societies, 285, 303
bulls, 287
Burma, 372
Burton, 92

Campion, Sir H., 95
Canada, 80, 180, 393, 399, 401, 412
capital:
  consumption, 99–103
  definition of, 16–18
  fixed, 17, 95
  formation, 99–103, 245, 253–4, 287, 306, 322–3, 342, 401
  international movements of, 388, 391–2, 396–401
  personal, 95
  social, 95
  types of, 94–6
  working, 17, 95
Capital Issues Committee, 308
capital market, 109–10, 283–8, 304
cartels, 386
catering industry, 212
Census of Distribution, 231
Census of Production, 231–3, 361
Central Electricity Board, 116, 161
central government, 80, 117, 234, 237, 264, 267, 316, 321, 356, 388
Central Statistical Office, 247
Chamberlin, E. H., 159, 179
Chancellor of the Exchequer, 308, 349
chemical industry, 407

climate, 71, 83, 85, 92
closed shop, 162
coal industry, 83–4, 89–90, 116, 118–119, 163, 200, 208, 407
Cohen, R., 93
commodity, 132, 147–8, 160
comparative costs, law of, 365–9, 375
competition:
  imperfect, 165, 171–9, 227
    nature of, 171–2
    equilibrium under, 173–9
  perfect, 153–9, 164–5, 171–2, 181, 183, 227, 240, 313, 377
    definition of, 154–5
    demand curves, 155–6
    market equilibrium, 156–9
competitive goods, 41
complementary goods, 41
composite demand, 41
Connor, L. R., 361
constant returns, 111, 125–6
contango, 286
convertibility, 384–7, 392–4, 398, 409–414
Co-operative Societies, 116
cost curves:
  average, 125–7, 151–3, 157–9, 183
  average and marginal, 128–31, 138–140, 174–6
  marginal, 132–4, 141–2, 151–2, 157–159, 163–4, 182–4, 190, 240, 314
cost of living, 355–9, 403
Courtaulds, 122
Crump, N., 395
Crusoe, R., 97, 99
Currency and Bank Notes Act, 266
current accounts, 273

death duties, 244
debentures, 114–15
deduction, 8
Defence Aid, 398, 400, 408
deflation, 337–41, 348–50, 374, 385–6, 392, 402–3, 410
demand:
  and utility, 27–42
  contraction of, 37
  curves, 34–40, 51, 148–50, 155–6, 196–8, 215
  elasticity of, 54–63, 150, 164, 167, 190
  extension of, 37, 62
  inter-relationships of, 40–2
  law of, 22–3
  perverse curves, 38, 198–200
  shift of, 37, 62
  theory of, 27–63
deposit accounts, 273
depreciation, 17, 97–8, 221, 233–4, 238, 321, 342–3
derived demand, 41
devaluation, 383, 391–2, 402–3, 410, 412
Development Areas, 92–3
Dillard, D., 298, 324, 336, 351
diminishing returns, 85–9, 107–8, 125–6, 183, 190, 196–7, 215, 368
diminishing utility, law of, 22, 30–1, 50, 186, 297
discount houses, 281, 300, 302–4
discount rate, 277–8, 300, 302–3
discriminating monopoly, 163–4
Distribution of Industry Act, 92
distributive trades, 80, 116, 231, 361
dollar, 96, 381–3, 387, 390–1, 393–4, 398, 408–10, 413–14
dollar area, 408–9, 413
dollar gap, 390, 408–10, 414
economic analysis:
  methods of, 7–10
  objectives of, 10–11

Economic implications of full employment, The, 351
economic laws, 22–3, 318
economic problem, 4, 143, 313
economic rent, 217–20, 228, 240
economics:
  definition of, 5–7
  nature and scope of, 3–13
  relation to other sciences, 12
economies of scale, 108–11, 165, 176, 179, 183, 376, 409
economy of high wages, 202
Edey, H. C., 247
efficiency, 11
elasticity of demand:
  and wages, 211
  concept of, 54–7
  exports and imports, 402
  factors governing, 57–61
  income elasticity, 61
  measurement of, 55–6
  significance of, 61–3, 150, 190
elasticity of substitution, 210–11
elasticity of supply, 140, 211, 371–2
electrical goods, 120–1
Ellsworth, P. T., 378
employment:
  and aggregate demand, 313–15
  and international trade, 373–5, 376–8
  theory of, 313–51
Employment Policy, 351
engineering industry, 80, 116, 120–1, 407
entrepreneur:
  functions of, 104–7
  identity of, 113–16

enterprise:
  definition of, 18
  factor of, 104–123
equilibrium:
  definition of, 21–2
  industry and firm, 150–3, 158–9
  market:
    definition of, 147–53
    under imperfect competition, 173–9
    under monopoly, 162–6
    under perfect competition, 156–9
  stable and unstable, 21–2
equi-marginal returns, 31–3, 181–2, 191, 227
equity, 11
ethics, 12–13
Europe, 3
European Payments Union, 393–4
exchange, 5–6
Exchange Equalisation Account, 389–390
exchange control, 390–2, 412
expectations, 222–3, 287–8, 297, 306–7, 319–21, 334, 341–2, 348, 381
exports, 93, 118–19, 121–2, 232, 236, 308, 315–16, 340, 344, 346, 360–361, 366–415
external economies, 108, 110–11

factor indivisibility, 110–11
factors of production, 18–19
Federal Reserve Bank, 399
fiduciary issue, 266, 269
Finance Corporation for Industry, 285
financial economies, 109
fiscal policy, see budget
fixed costs, 128–30, 137, 139–40
Ford Company, 121
foreign exchange:
  brokers, 381–2
  market, 380–2
  rates, 379–95
  reserves, 398, 401, 409, 412–13
  state intervention in, 389–92, 412
Forth bridge, 352
forward exchange, 381–2
France, 80
francs, 380–1, 387
full employment, 143, 343–7, 387, 392, 394, 410

Gas Board, 116–18
gas industry, 168
G.E.C., 120
General Post Office, 161
Germany, 254, 340
gold, 96, 254–7, 263–4, 266–7, 269, 385–7, 393–4, 398, 407
goldsmiths, 256–60
gold standard, 384–7, 392, 406–7, 411
gross domestic product, 11
gross output, 97–8
guilders, 382

Hansen, A. H., 298, 324, 336, 351
Harrod, R. F., 378, 395, 405
Hayek, F. A., 103
Henderson, Sir H., 13, 23, 42, 103, 123, 143, 159
Hicks, J. R., 53, 213
High Wycombe, 91
hire purchase, 303, 308
Hubback, E., 82
Hungary, 340

Iceland, 412
I.C.I., 122, 297
imports, 74–5, 167, 189, 232, 236, 315–16, 360–1, 365–415
incentives, 245–6
income:
  definition of, 16–17, 262
  distribution of, 185, 192, 234, 239–246, 338–40, 344, 346
  formation, 325–36, 341–3, 373–4
  national, 230–47, 314–15, 323, 333–334, 350
income effect, 38, 51–3, 60–1, 199–200
increasing returns, 22, 107–11, 125–6, 190
index numbers, 352–61, 370–1
indifference curves, 43–53
indifference maps, 47–8
induction, 9
Industrial and Commercial Finance Corporation, 285
industrial organisation, 113–16
industrial production, index of, 354, 360–1
industry:
  and firm, 148–53, 161
  definition of, 131–2
infant industries, 376
inferior goods, 38–40, 52, 135
inflation, 337–41, 346, 348–50, 374, 406–7
inheritance, 244
insurance, 224–5
insurance companies, 224–5, 401, 411
integration, 121–2
interest:
  long-term and short-term, 295–7
  theory of, 289–98
internal economies, 108–10
International Monetary Fund, 392–3
international trade:
  and employment, 373–5, 377–8

level of, 231–2, 315, 350–1, 365–415
state intervention in, 375–8, 404–5
invisible trade, 396–7
investment:
and employment, 315–16
determinants of, 319–22
domestic and foreign, 102
rate of, 74–5, 102–3, 306–7, 313–24, 331–6, 341–5, 347–9, 372–4
Iraq, 412
Irish Republic, 412
iron and steel, see metal manufacture
Iron and Steel Board, 120
issuing houses, 284, 303
Italy, 80

Japan, 407
jobbers, 286, 381
Johnson, H. G., 415
joint demand, 40–1
joint-stock companies, 113–16, 221
joint supply, 141–2
Jordan, 412

Keirstead, B. S., 229
Keynes, J. M. (Lord Keynes), 262, 298, 324, 336, 351
Keynes, J. N., 13, 23
Knight, F. H., 229
Korea, 372

labour:
as factor of production, 67–82
definition of, 18
demand for, 195–8
direction of, 82, 347
division of, 68–70, 108, 110, 254, 366
marginal product of, 196–8, 201–3, 209–10
mobility of, 80–2, 313–4, 341, 346, 377, 409
supply of, 198–200
land:
and diminishing returns, 85–9
and industrial location, 89–93
cost of production of, 216–17
definition of, 18
demand for, 215
rent of, 214–16
supply of, 83–5
legal tender, 261
Lewis, W. A., 405, 415
Libya, 412
linking of processes, 109
liquidity:
definition of, 254
preference, 289–95, 306
ratio, 272–3, 305
Liverpool Cotton Exchange, 91, 411
local government authorities, 7, 80, 168, 234, 237, 307, 321, 348
location of industry, 89–93
London and Cambridge Economic Service, 371
London Clearing House, 270, 274–5
London tin and tea exchanges, 411
long period, 137–40

Macrae, N., 288, 309
MacGregor, D. H., 143
macro-economics, 10
Malaya, 372
Malthus, T. R., 72
manufacturing industry, 116
margin, definition of, 19–21
marginal efficiency of capital, 320
marginal rate of substitution, 49–50
marginal revenue, 149–52, 156–9, 163–4, 182–4, 190, 197–8, 314
market, definition of, 148
marketing economies, 109
Marshall, A., 7, 13, 23, 42, 63, 93, 103, 123, 143, 159, 170, 192, 213, 220, 229, 246
Marshall Plan, 398, 413
Meade, J. E., 192, 246, 351, 405
mechanisation, 69, 108–9
metal manufacture, 80, 90, 116, 119–120, 162
micro-economics, 10
Ministry of Labour, 211–12
monetary policy, 299–309, 347–8, 384
money:
functions of, 251–4, 289–90
market, 269, 271–2, 276–82, 302
requirements for, 254–6
use of, 6–7, 37, 96, 124, 142, 147, 238, 251–90, 294–5
Monopolies Commission, 169
monopoly:
and public policy, 166–70
discriminating, 163–4
effects of, 154, 160–72, 184, 187, 240
equilibrium under, 162–6
nature of, 160–1
profits, 227–9
sources of, 161–2
moral suasion, 301
multilateral trade, 379, 412
multiplier, 328–34, 341–2, 350, 385

National Coal Board, 119, 160–1
national debt, 267, 279

national income:
  accounts, 235–8
  and welfare, 238–40
  Blue Book, 98, 118, 231, 235–6, 242
  distribution of, 239–46
  level of, 230–47, 314–15, 323, 333–334, 350
  measurement of, 230–5
nationalised industries, 116, 234, 272, 348
N.A.T.O., 398
net output, 97–8
Nevin, E., 298, 309
new issue market, 283–5, 303
New York, 390–1
New Zealand, 232

one-man businesses, 113
open-market operations, 300–1
opportunity cost, 142–3, 217
optimum allocation, 183–6
optimum factor-combination, 106–7, 110–11, 126
optimum output, 153, 176–9, 183
optimum population, 72–3
optimum size of firms, 112–13
ordinary shares, 115
overhead costs, see fixed costs

Pakistan, 101, 412
partnerships, 113
patents, 161
Peacock, A. T., 247.
pension funds, 284
piece-rates, see wages
Pigou, A. C., 13, 246, 361
politics, 12
population:
  declining, 73–5, 345
  optimum, 72–3
  Royal Commission on, 77, 82
  size of, 71–5, 198, 239
United Kingdom, 76–80
pound sterling, 379–81, 384–5, 387, 389–91, 409–15
precautionary motive, 293
preference shares, 114–15
price:
  definition of, 147
  control of, 23
  mechanism, 180–92, 225–7, 377
price effect, 50–1
price level:
  fluctuations in, 337–51, 384, 403–4
  measurement of, 352–3, 355–60
  stability of, 346–50, 410
prime costs, see variable costs
product differentiation, 165, 172–3, 176–9
production, theory of, 67–143
profit maximisation, 8, 106, 132–3, 151–3, 158–9, 162–3, 166–7, 174, 176, 189, 198, 229
profits:
  and allocation of resources, 225–7
  monopoly, 163–6, 227–9
  normal, 125, 151–2, 163, 174, 176, 186–7, 225, 227
  role of, 188–9
  theory of, 221–9
  undistributed, 318, 349
propensity to consume, 318–19, 330–1, 333–4, 342–6
propensity to import, 373–4
propensity to save, 317–19, 329–31, 334–6, 344–6, 374
property, ownership of, 236, 243–6
public corporations, 116
Purchasing Power Parity theory, 387–388

quasi-rent, 219–20, 223, 228, 236
quotas, 368, 375, 388, 393–4

railways, 116
rationing, 23, 101, 189, 191–2
rent, 85, 214–20, 223, 236–7
rent of ability, 217–18
Rent Restriction Acts, 241
Restrictive Trade Practices Act, 169
Restrictive Practices Court, 169
retail price index, 355–9
risk, see uncertainty
Robbins, L., 13, 23, 192
Robertson, Sir D. H., 192, 262, 415
Robinson, E. A. G., 79, 93, 123, 170
Robinson, J., 153, 159, 179, 220, 324, 336, 378
Rootes Group, 121
Royal Commission on Population, 77, 82
Royal Mint, 256

salaries, see wages
sales curve, 155, 168, 173–5, 184
Sargant Florence, P., 82, 213
saving:
  and investment, 313–24, 325–36
  definition of, 322–3
  equilibrium rate of, 329–30
  level of, 100–1, 237, 293–4, 317–19, 335, 343
Sayers, R. S., 262, 275, 288, 309, 395
Scandinavia, 180
scarcity, 3, 5–6
Schultz, H., 63

securities, 114–15, 266–9, 271–3, 282–288, 291–8, 300–1, 303–6, 315–17, 319, 411
selective credit controls, 307–9
settlement day, 286
shipbuilding, 116, 120–1
short period, 137–40
silver, 254–6, 341
site values, 87, 89–93
sliding scales, 208
slump, 307
small firms, 111–13
social services, 71, 74, 245, 316, 344, 348, 358
Soviet Union, 100
speculation, 287–8, 293, 307–8, 381, 389–92, 399
speculative motive, 291–4
stability, 11
Standard Motors, 121
state intervention, 185–6, 200–2, 204, 211–13, 241, 375–8, 389–92, 404–405
Stephenson, R., 218
Sterling Area, 409, 411–15
stock appreciation, 118, 236
Stock Exchange, 39, 272, 283–8, 303, 381
stocks, 136, 302, 342
Stone, R., 246
subsidiary trades, 91, 110
subsidies, 117, 185, 189–90, 232, 236, 245, 375, 388
substitution:
  elasticity of, 210–11
  marginal rate of, 49–50
supply:
  and price, 133–4
  and time, 136–40, 142
  curves, 124, 131–6
  elasticity of, 140, 211, 371–2
  joint, 141–2
  operation of, 124–43
sweated trades, 212

tariffs, 161, 189, 368, 375, 388, 394
taxation, 12, 61–2, 185, 189–92, 219, 232, 234–7, 244–6, 334, 339, 349, 358, 375
Tennessee Valley, 85
terms of trade, 369–72, 396–7, 407–8
Tew, B., 395, 405, 415
textile industry, 80, 116, 122–3, 407
time rates, see wages
Town and Country Planning Acts, 93
Trade Boards Act, 211–12
trade cycle, 340–3
Trade Unions, 81, 162, 200, 204, 206–213, 240, 339, 346–7, 386, 410
transactions motive, 290, 293
transfer cost, 217–20, 228
transfer payments, 234–5, 237, 245, 349, 356, 359
Treasury, 266, 278, 301, 308, 347, 390
Treasury bills, 278–9, 300, 305
Triffin, R., 179
Truptil, R. J., 275, 288

Uap, 256
uncertainty, 105–6, 223–5, 228, 297–298, 320
underwriters, 284, 303
unemployment, 11, 211, 313–15, 318, 343–4, 347, 350, 386, 407, 410
United Kingdom:
  balance of payments of, 400–2, 406–411, 413–14
  banking in, 256–8, 260–2, 263–75
  capital stock of, 95–6
  domestic investment in, 98, 102, 236, 344
  economy of, 4, 181, 232, 234–47, 299, 308, 316
  exchange rate policy, 389–92, 402–403
  foreign investment by, 372, 401, 406–407, 412
  foreign trade of, 340, 350–1, 370–2, 393–4, 396–8
  industrial structure of, 79–80, 116–123
  international position of, 406–15
  national product of, 98, 195, 235–6, 238–9, 262
  population of, 76–80
  property in, 243
  retail prices in, 355–9
  size of firms in, 111, 119–23
  terms of trade of, 370–2, 407–8, 412–13
  unemployment in, 344, 407
  wages in, 203, 211–13, 346–7
United States, 3, 4, 80, 85, 168–9, 202, 206, 288, 377, 380, 398–9, 401, 406–9
utility:
  definition of, 14–15
  inter-personal comparisons, 29
  maximisation of, 31–2, 34–5
  measurement of, 28–9
  preference versus, 43–4, 52–3
  total and marginal, 27–9

value, definition of, 16
value judgements, 12
variable costs, 128–30, 137, 139–40

Vauxhall Motors, 121
vehicle industry, 116,121–2, 407
visible trade, 396–7

wages:
  determination of, 200–2
  index of, 360
  real and money, 207–8
  relative, 202–5
  state intervention in, 200, 202, 204, 211–13
  theory of, 195–213
  time and piece rates, 205–7
Wages Councils Acts, 212
Wall Street, 288
wealth:
  definition of, 15–16
  distribution of, 11
welfare:
  and national income, 238–40
  definition of, 15–16
  level of, 29
Western Europe, 344, 393–4, 398, 403
wholesale prices, index of, 360
Wicksteed, P. H., 23, 42
Wilson, T., 309
Wootton, B., 213

Zurich, 390

PRINTED IN GREAT BRITAIN BY ROBERT MACLEHOSE AND CO LTD
THE UNIVERSITY PRESS, GLASGOW