THE COMPETITIVE PROCESS

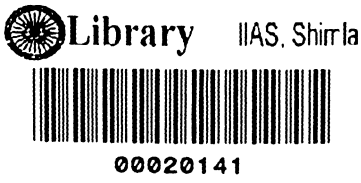
THE COMPETITIVE PROCESS

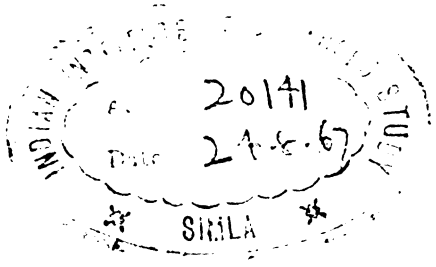
by

JACK DOWNIE

GERALD DUCKWORTH & CO. LTD
3 Henrietta Street, London, W.C.2

First published 1958





Printed in Great Britain by
The Camelot Press Ltd., London and Southampton

CONTENTS

PREFACE

I HOPE that my argument in this book is sufficiently self-explanatory to need no preface. But some explanation is perhaps needed of why I chose to write it.

The theory of competition is at once the pride and the shame of economics, a logical structure of the greatest elegance which has only the most tenuous connections with the reality it is supposed to interpret. But it is a dangerous structure to meddle with, providing a ready grave for the reputations of reformers. I hoped that it would be possible to investigate the nature and size of the observable effects of "lapses from competition"—my original interest—without commitment to anything very much in the way of theory. But I soon was forced to recognise that the reason so little work had been done on such observable effects was that a theoretical basis is needed for it, and does not exist. Reluctantly, therefore, I was driven to think and write about the theory of competition. This is my general apologia. The more important weaknesses of the theory which has emerged are, I hope, adequately indicated at the appropriate points. I should make it plain here, however, that I have confined both my theorising and my empirical investigations to what is usually described as manufacturing industry. Distribution and other service activities require an approach of their own.

The ground I have worked over has been well tilled, and I cannot hope to acknowledge, or even identify, the sources of all the ideas I have appropriated. Marshall's trees in the forest will immediately leap to mind. Indeed, in a sense I have done no more than try to rescue Marshall's notion from the damage it suffered when Mrs. Robinson pointed out that pike in a pond might be a better analogy. In concentrating on growth and change rather than equilibrium I am swimming with the main stream of post-war economics and, of course, I have drawn much food for thought from Schumpeter. J. Steindl's book on *Maturity and Stagnation in American Capitalism* contains a model of the competitive process which is very like my own, although he places

more emphasis on market imperfections, and his main interest is somewhat different.

I acknowledge with gratitude a grant received from the Government under the Conditional Aid Scheme for the use of counterpart funds derived from United States economic aid. This made it possible for me to take the two years' leave during which this book was written. Since I have now returned to the Civil Service, I should perhaps make clear that the views expressed are purely personal.

The Oxford University Institute of Statistics gave me a place to work and the use of all facilities. I am very grateful to its Director, Mr. F. A. Burchardt, for his kindness in this respect, and for the help and encouragement which he gave me throughout. I thank Miss Mary Gisborne for the efficiency and speed with which she turned my drafts into typescript, Miss Jean Morris, for relieving me of all the cares of calculation, and all my colleagues at the Institute, who made my stay so pleasant and rewarding.

Miss J. M. Haigh worked for me as research assistant for six months and helped greatly, in particular in clearing my mind of a welter of misconceptions about the organisation of industries and markets. Mr. F. R. Oliver of Nuffield College carried out the work on the stability of rates of profit which is summarised in Chapter XII.

The Office of the Censuses of Production and Distribution provided me with the empirical material with which Part II is concerned. The National Institute of Economic and Social Research kindly gave me access to the working files of their enquiry into Company Income and Finance. I am most grateful to both of them, and I wish to thank in particular Mr. H. C. Stanton, the Director of the Census Office, who went to great pains in devising ways of meeting my needs.

I have profited greatly from discussion with others. Mr. W. A. B. Hopkin and Mrs. M. F. W. Hemming first suggested this topic to me and have been helpful throughout. Mr. C. B. Winsten has dug me out from many an intellectual morass. I had the good fortune to enjoy almost daily discussions with Mrs. Joan Robinson and Mr. N. Kaldor in Poona in 1955 at a critical period in my

thinking. T. Balogh, F. A. Burchardt, D. M. B. Butt, L. A. Dicks-Mireaux, C. H. P. Gifford, E. F. Jackson, K. G. J. C. Knowles, Joan Robinson and C. B. Winsten kindly read my first draft and made many helpful comments. To all these and to others not mentioned by name I give my best thanks.

For the errors in which I have obstinately persisted I take full responsibility.

J. D.

LONDON, 1957.

INTRODUCTION

Chapter I

THE OBJECT OF THE ENQUIRY

THE period since the war has seen a marked revival of political interest in Western Europe in the issues loosely denoted by the terms competition and monopoly and a sharp change in the prejudices with which they are approached. The first gleam of the awakening in Britain was the White Paper on Employment Policy of 1944. But it is significant that competition appeared there only somewhat parenthetically as a condition for counter-cyclical government spending to have its intended effect of raising employment. By 1948, the tide was running faster and monopolies and restrictive practices achieved political status in their own right in the Act of that name. By 1956, the cautious recognition in the Act of 1948 that there was a problem for examination had been replaced by widespread conviction that the problem was real and the solution known. And now, the rebirth of free trade as a live issue in international politics presages the re-entry of the international cartel problem from the obscurity in which it has long obsolesced.

It is not difficult to see the reasons for this revival of interest. If the problem of fully employing resources is solved, as many rightly or wrongly believe, it is natural that the focus of attention should shift to how those resources are employed. And it is one of the most venerable propositions in political economy that the way in which resources are used will be greatly affected by the character of the relationships between the firms who use them. The force of this tradition has been greatly strengthened by the growth of the ideological influence of the country where it is most passionately entertained. The fact of greatly superior American productivity has loomed large in our thoughts because of the severe problems which it has created. And it is only a short step from observing that Americans attribute a large role in this development to their anti-trust policy to accepting their belief as

established truth. Finally, the waning of belief in the virtues of positive government intervention in the working of the economy implies, as a logical counterpart, recognition of the need for action to prevent self-appointed bodies of industrialists from stepping in as representative government steps out.

The ultimately important function of a social science is to produce works of what I would call a metapolitical character, which provide a framework of the mind for the politician and an assembly of digested fact to clothe that framework. It is somewhat surprising, therefore, that the revival of political interest in questions of competition has not been matched by any significant revival on the academic side, at least as evidenced by published writings. One possible conclusion is, of course, that all that is necessary and possible on the metapolitics of monopoly has already been written. I do not think myself that this conclusion is right. The object of this chapter is to justify my belief. The object of the book as a whole is to examine why the academic soil has become so sterile in this field and to suggest the way in which a new metapolitics can be created.

Let us first try to establish, in a rough and ready way, the nature of the issue with which the politicians are concerned. The terms in the debate—competition, monopoly, restrictive practice and the rest—can mean all things to all men. The shades of meaning range from that of the buyer, who is tempted to brand as monopolistic any action by a seller which he finds uncomfortable, to those in the economic text-books, whose applicability to the circumstances of practical business can sometimes be seen only by the eye of faith of the professional academic. I shall have a good deal to say on concepts and definitions later in the book. For the moment, however, all that I wish to establish is the general nature of the issue with which politicians are, and economists should be, concerned. This, I suggest, is the nature of the rules of the game for a market economy. I will explain.

I mean by a market economy one in which most capital assets are privately owned and administered and in which the owners or administrators enjoy a wide measure of discretion as to how

they use those assets. Private, let it be noted, is here used, perhaps somewhat loosely, to mean "not government".

The basic assumption which makes a market economy a defensible form of economic organisation is that the rivalry—struggle to succeed—of the different owners or administrators of capital assets will yield a sufficient degree of enjoyment—income—to the society which defends their ownership to maintain its preference for this over other possible forms of organisation. There are, in fact, two articles of faith; that rivalry produces the best results; and that there is rivalry in a market economy. In these respects the working of a market economy is analogous to a sporting game.

Now in games of sport we are prepared to say, without hesitation, the keener the rivalry the better the game. But no one except the most intransigent trust-buster would characterise the economic game in this way. The fact is, of course, that for the sporting game there is a comprehensive and commonly agreed set of rules within which the game is played and which prevent rivalry from manifesting itself in forms which, as we say, would spoil the game.[1] And it is the absence of such a common book of rules—enabling us to make an implicit and common set of reservations—which makes most of us unwilling to assent to explicitly unqualified propositions about the virtues of rivalry in the economic game. Pure, perfect or workable competition, monopoly, oligopoly and the other expressions which are the common coin of discussion for those who labour in this field are then, I suggest, best regarded as the names of alternative sets of rules under which, it is believed, the economic game is or should be played.

More specifically, political dispute relates to the three broad questions which have to be settled when any book of rules is drawn up; those of authority, of eligibility and of conduct. Who shall make the rules of the game and be responsible for seeing that they are adhered to? What rules shall be made defining those allowed to play? And what rules of conduct shall be laid down for their play?

The traditional answers in Britain must be distilled from the

[1] Even so, matches are sometimes spoilt by teams breaking the spirit of the rules although abiding by their letter.

practice of Parliament and the Courts rather than from the pronouncements of statesmen or politicians. Inspection of the record shows that, in the past, the Government itself has had little ambition to become the rule-making authority on questions of business conduct, while the Courts have certainly treated any occasional move in this direction as betraying absence of mind rather than firm intent. The authority of statute or common law has, by and large, been confined to a limited number of issues—such as trade marks, patent protection, passing off, etc.—where there has been no serious dispute about the need for rules or the kind of rules needed. For the rest, the State has taken the attitude that if rules of conduct need to be made—on which its view has changed from time to time—this is a matter for firms themselves. Given this acquiescence, firms in some industries have made very elaborate sets of rules to control their mutual relations and devised extensive machinery for enforcement. Neither the State nor firms have sought to devise rules to limit the size of participants in the game.

Those whom, without prejudice, I will call conservatives adhere to the traditional answers. The reformers (again without prejudice) seek two Acts from the Government; first, an Act to rescind most of the rules which firms themselves have made; second, a positive rule—made and administered by the State—to limit the permissible size of industrial enterprises. These are the parties in the post-war debate and such are the issues.

Political disputes may be about values or about facts; about the object of an activity or the means by which that object is best secured. Disputes about values may be irreconcilable; disputes about facts can, in principle, always be resolved.

It is fairly clear, I think, that the dispute with which I am concerned is not primarily one which concerns values. In a market economy the actions of individual producers are expected to determine, in their sum, what is produced, the efficiency with which it is produced, and the distribution of titles to this production. There would be general agreement that the more is produced from a given endowment of productive resources, the faster these increase, and the closer the pattern of output to what people "really" want, the better is the result. There is, of course,

room for dispute about the proper distribution of income. Moreover, some would add further to the list of objects I have given, with particular reference to such "sociological" values as the concentration of economic power, the preservation of a career open to talent, or the need for business conduct to be defensible by the standards of private morality. Nevertheless, I think that the criteria I have suggested would be accepted by most parties as covering the more important aspects, and that differences of view as to the relative weights to be given to each are essentially peripheral to the dispute we are examining.

The argument is, therefore, one as to facts. The question is whether, as a matter of fact, the reformer's book of rules would lead to a better result in terms of economic performance than would that of the conservative.

I have said that a factual dispute, which this one is, is in principle always capable of resolution. This is not to say that in any case where there is agreement on ends it is in practice possible to collect such evidence that differences of opinion cease to exist. The very fact that we speak of opinion in connection with empirical questions implies that universal agreement is never, in logical terminology, necessary. There is a whole spectrum of situations in the empirical world, with the probability of disagreement increasing as we pass along the spectrum. At the one extreme is the carefully controlled and indefinitely repeatable laboratory experiment which demonstrates the variation of one quantity with another. It is only in some Pickwickian sense that I can remain unconvinced of the equality of the angles of incidence and reflection. At the other extreme is Cleopatra's nose,[1] whose proper significance there is now little hope of determining.

The essential difference between Cleopatra's nose and the glass prism is the uniqueness of the events which lie at the contentious end of the spectrum. In the laboratory experiment (in simple cases) all relevant factors, save those whose connection is under investigation, can be held constant over an indefinite number of

[1] Some explanation may be necessary for those who have not encountered this classic problem of philosophy. The argument is as follows. If Cleopatra's nose had been half an inch longer, Antony would have been much less besotted. He would, therefore, not have followed her panic retreat at the battle of Actium. Octavius would not then have won the battle and become Augustus. And so on, for as long as interest and historical knowledge will sustain you.

repetitions of the experiment. By contrast, the type of event with which the social sciences are concerned is always unique, in the sense that the whole complex of factors is likely to differ for each example we observe. It is this which helps to make social sciences so difficult. But it does not make them impossible. So long as the event is repeated sufficiently the influence of attendant factors can either be seen to be irrelevant or be separately established. As recorded experience lengthens and the techniques for analysing it are improved, issues tend to move from the nose towards the prism end of the spectrum, becoming, as we say loosely, less and less matters of opinion and more and more matters of fact.

The effects of abolishing the death penalty for murder, for example, were once matters of mere opinion and acrimonious dispute, and appeals to supposedly self-evident facts about human psychology were the stuff of which arguments were made. With the passage of time and the steady increase in the number of, in other respects widely different, countries where execution has been abolished, the confidence with which the uniquely deterrent character of the death penalty can be denied has steadily increased. We have not yet reached the stage when continued affirmation of this supposed quality can itself be taken to imply mental derangement, and to those who take a retributive view of punishment all factual evidence on this point is, of course, irrelevant. But the birth of a new "established fact" is well in sight.

The striking feature of the argument as to the effect of rules of the game on business performance is that, old though the dispute is, it has made virtually no progress along the spectrum of opinion. Human psychology or the mere, and usually emotionally charged, fact of the particular example still provides most of the ammunition, and Adam Smith's dictum that "People of the same trade seldom meet together, even for merriment and diversion, but the conversation ends in a conspiracy against the public or in some contrivance to raise prices" remains one of the heaviest cannons of the reformer, not as a received idea whose validity can be established by an appeal to the evidence, but as a mere expression of opinion.

A superficial reason for this arrested development is the persistent conviction of many on both sides of the fence that their

position is so self-evidently true as to require no support from formal evidence. Such conviction usually arises from an unconscious narrowing of the criteria by which business performance is to be judged, and the record is filled with arguments of the "it stands to reason" kind whose relevance depends on a narrowing of this kind.

The reformer will argue, for example, that it stands to reason that, if business-men get together to fix a price, this will be higher than the one which would have ruled if they had not so connived. If this were not so, he will argue, why should they get together? Notice that, even if the argument is acceptable, its relevance rests on the assumption that the price which would rule in the absence of agreement would be the "right" (best) one. But it is going too far to accept the argument. Its apparent self-evidence rests on the unstated assumption that costs would be the same in both situations. Our assent is, in reality, to the proposition that business-men will think it worth while to agree on a price only if it is higher in relation to costs than the one which would otherwise rule. And to give it striking force in this form we must say, which the conservative would stoutly deny, that the level of costs is either unaffected by agreements or irrelevant to the issue.

Similarly, the conservative—particularly the business-man—will affirm, from his practical knowledge, that a radical change in the rules will create such confusion amongst the players as to spoil the whole game. And a reasonable man would admit that sudden change is indeed likely to create confusion for all and hardship for some. But again the argument is inconclusive, and perhaps irrelevant, unless we take the view that short-term (and more certain) consequences are far more important than long-term.

I believe, however, that the more fundamental explanation for the obstinate stagnation of the "monopoly" argument is to be found in the failure of the professional economists to meet their responsibilities. The progress of any idea from the realm of opinion to that of established fact depends, not only on the collection of information, but also on the formulation of the idea in such a way that it can be proved or disproved by the information, and that is a job for the specialist. I go more fully into the traditional

formulation in my next chapter. But even a cursory review of economic writings provides strong *prima facie* evidence for believing that economists have failed to frame this idea in a useful and fruitful way. The essential characteristic of the literature of competition and monopoly is the sharp distinction between theoretical and practical work. On the one hand is the typical study on the theory of the firm, whose purity is never sullied by the intrusion of statistical material. On the other is the empirical study of a particular industry, whose mountain of descriptive material is far too intransigent to be formed and informed by a theoretical framework. It stands to reason, if I may use the term, that there must be something wrong with a subject where theory and practice are so far apart.

In his introduction to the Cambridge Economic Handbook series, Keynes gave the following account of the nature of economics. "The Theory of Economics does not furnish a body of settled conclusions immediately applicable to policy. It is a method rather than a doctrine, an apparatus of the mind, a technique of thinking, which helps its possessor to draw correct conclusions."

What I have been suggesting in this chapter is that the apparatus of the mind which economists have provided for thinking about questions of monopoly and competition is one which does not help either politicians or judges in arriving at practical conclusions. I cannot ignore that the outline of an alternative framework of thought which I have tried to provide in this book is vague, sketchy and inconclusive by comparison with the crystalline elegance of the neo-classical theory of the firm. Nor can I claim that politicians or judges will find here the policeman's manual for which they must crave. My hope is that I have done something to show where further development is possible, as I believe it is not in the direction in which we have been accustomed to look.

PART I: THEORY

Chapter II

THE THEORY OF THE FIRM[1]

I TURN now to justify my contention that economic analysis has failed to furnish a useful framework for thinking about problems of competition and restrictions on it. I shall suggest that attention has been directed to the wrong problems, and that the concepts which have been developed are not useful when enquiry is shifted to the questions which I believe to be important.

The object of my criticism is what is usually termed the competitive model or the theory of the firm. It is the method of analysis whose coming of age in the thirties is associated with the names of Joan Robinson and E. H. Chamberlin, and which has now been systematised, most fully perhaps by F. Machlup.[2]

I am very conscious that the summary form in which I have presented both my exposition and my criticism may mislead by suggesting a cavalier attitude. I plead in extenuation that my main aim in this book is constructive rather than critical. As regards the welfare economist, I can only plead that my venture on to the needle-point where none but angels can balance is so fleeting as to warrant a charitable blind eye.

The Problems Analysed

The commonest form of the question with which this book is concerned is whether departures from competition are against the public interest. It is assumed in economic analysis that the public interest is in economic performance rather than in forms of economic organisation as values in themselves, and this assumption is very proper, since only the former is within the professional competence of the economist. But performance has many aspects,

[1] The layman is warned that this chapter, though not drafted with deliberate obscurity, is directed primarily to the professional. Understanding of the remainder of the book will not be impaired by its omission.

[2] *Economics of Imperfect Competition*, Joan Robinson; *Theory of Monopolistic Competition*, E. H. Chamberlin; *The Economics of Sellers' Competition*, F. Machlup.

and there has to be some principle of selection for separating the important from the trivial if any progress is to be made. I suggest that, as a first approximation, the best principle of selection is to define the public interest as what the public is interested in.

In the area with which we are concerned, two aspects of economic performance have attracted particular attention in the post-war period. First, the continued co-existence of firms of widely differing efficiency in the same line of business or, an alternative formulation of the same point, the extent to which average efficiency in many industries falls short of the level which the practice of the most efficient firm shows to be possible. Second, the alleged slowness with which innovations in products or productive techniques are diffused through British industry, and the resulting slow growth in efficiency. Even a cursory study of the pronouncements of governments and of those political economists, professional or lay, who have sought to influence their conduct will demonstrate the attention which these two issues have attracted.

Now I do not wish to imply, either that the preoccupations of governments always reflect those of the thoughtful members of the community, or that thoughtful members need always be right in their judgment of what is most important. But there is at least a *prima facie* indication that these two aspects of performance—the dispersion of efficiency and progress—should have an important place in an economic analysis of the effects of alternative rules of the game on economic performance.

The surprising thing is the scant attention which these two questions receive in the theory of the firm.

The problem which that theory purports to answer may be summarised as follows. Suppose that individual scales of preference and the state of technical knowledge are given, and that supplies of the different factors of production are also given, including in particular a supply of entrepreneurs with differing capacities for the organisation of factors of production. Then there will be one set of techniques and one distribution of productive resources between different entrepreneurs and different lines of production which will result in preferences being satisfied to the maximum extent possible with the resources and

techniques available. If the individual's choice of what he will buy or of where he will put his factor to work is taken to depend on the relative prices (commodity or factor) which confront him, there will be one, and one only, set of relative prices which corresponds to this optimum allocation of factors of production. The economist seeks to establish the characteristics of this optimum pattern of output and prices, to determine what rules of the game could lead to its being achieved in a market economy, and to discover what divergence from the optimum will result if various different rules be adopted.[1]

Clearly, one of the two major questions with which I have suggested that the analysis should deal is expressly excluded by this formulation. Technical knowledge is given, and the analysis is in terms of one position and not of movements from one position to another. Hence there is no place for progress or the growth of efficiency. This is, of course, not in itself a condemnation. Progress in understanding, like that in the material world, depends on the division of labour and there is no objection to treating the growth of efficiency separately, so long as the results of the two lines of enquiry can be brought together at the end. But I give reasons later for believing that this integration is not, in fact, possible.

The second of my questions, that of dispersion, is covered by the analysis, under the, at first sight curious, title of the theory of the optimum size of firm. It is inevitable that firms of differing efficiency should co-exist, it is said, because there are technical limits to the growth of any one firm; its efficiency will begin to decline after it passes a certain size. Hence, the less efficient firm is protected against absorption. Attention is then concentrated on the way in which different rules of the game may check the growth of the efficient firm before it is of optimum size.

I deal with the defects of this answer to the problem of differences in efficiency in the second part of this chapter. The important point to note at this juncture is that the question occupies a very subordinate position amongst those to which the theorists address themselves. It is reasonable to judge the importance

[1] I have chosen to concentrate on long-period equilibrium, since this has usually been regarded as throwing up more fundamental conclusions.

attached to various questions, in a rough and ready way, by the amount of space allotted to them. By this criterion the important issues in the theory of the firm are the pattern of commodity and factor prices; in other words, the composition of output and the distribution of income. Indeed, most of the analysis is conducted on the "heroic"[1] assumption that all firms are of equal efficiency.

Far be it from me to suggest that it does not matter what is produced, or that the distribution of income is an unimportant question. What I do suggest, however, is that it is a profound mistake to think that either of these patterns will be affected at all significantly by changes in the rules of the game.

Let us suppose that the firms in a particular industry agree together to fix a price higher than that which would otherwise rule and than the optimum price, to divide the market between them, and to prevent other firms from entering the industry. The effect of the higher price is to reduce the amount of the industry's product which people will buy, and this represents a decrease in welfare. The amount by which the industry's output will shrink depends, of course, on the rise in price and on the sensitivity of demand to price changes—i.e. on the elasticity of substitution. The higher the elasticity of substitution, the greater the effect on demand of a given change in price.

Now a commodity with a high elasticity of substitution has, by definition, one or more close substitutes; that is, there are other products which the purchaser regards as meeting his needs almost as well. But, if this is so, it is difficult to see that much harm is suffered from a rise in price which diverts purchasers to these close substitutes. Conversely, although it is plausible to argue that there is a considerable loss of welfare when high prices force purchasers to turn to substitutes which they regard as distinctly inferior, the elasticity of substitution for so unique a product will, by definition, be low, so that only a large rise in price will lead to any substantial diversion. I conclude, therefore, that the loss of welfare resulting from people's being driven into buying a basket of goods less satisfactory than that which they would and could buy, were the rules of the game only different, will be

[1] See Chamberlin, *loc. cit.*

small, unless there is reason to think that relative prices will vary very substantially with the rules.

This last is an empirical question, with which the theorists have not concerned themselves. For my own part, I find it difficult to believe that the variations will in fact be substantial. Since percentage profit margins on turnover are, typically, of the order of 10%, even a doubling of profit margins would lead to only a 10% increase in price. But we must conceive of the price agreement as being under constant destructive pressure, from both the individual self-interest of the participants and the firms outside it who would like to share the spoils, and the strength of this pressure will vary with the richness of the spoils which the agreement yields. Given also that the participants must always be looking over their shoulders at the political reactions to their policy, the effect on price of the typical agreement will evidently be considerably less than in the above example.[1]

It is difficult to carry discussion of the relevance of the traditional analysis of the relations between competition and income distribution very far, because of its signal failure to develop any theory of factor supply. And the broad gap between the analysis of factor incomes and that of personal incomes makes matters so much the worse. But, fortunately, brief examination of the magnitudes involved is enough to establish that this issue also is of minor importance as compared with those of "efficiency" and "progress" with which the policy-maker is concerned. Income from employment accounts for over four-fifths of the total net domestic income of the United Kingdom. This means that even if income from property were halved, income from employment would be raised by only some 10%. No systematic data on the dispersion of efficiency within industries exists. But the qualitative statements of those who have carried out industrial studies suggest that the spread between the worst firms and the best, and, therefore, the scope for raising income, is much more

[1] The above argument follows the traditional analysis in supposing a one-commodity industry. If we take, instead, a more common-sense definition of an industry, as a group of firms producing a variety of products, the practical scope for "profiteering" on any one of them will, of course, be very much greater. The descriptive literature is, in fact, rich in examples of such "discrimination" in product prices by multi-product firms. But the argument in the remainder of this chapter is directed to establishing that the traditional model cannot cope with the issues raised when we take this further step towards reality.

than this, a conclusion which is supported by the evidence presented in Chapter XIV. And the difference in income from employment (the distribution of income remaining constant) resulting from a rate of productivity increase of 4% per year as compared with one of 2% would amount to over 10% after only 5 years.[1]

To sum up. The main preoccupation of the theory of the firm has been with aspects of economic performance where the quantitative effects of different rules of the game are likely to be small. The tolerance of the economic system for differences in efficiency and the rate of growth of efficiency over time—which lay opinion expects to be considerably affected by the conditions of competition—have received very much less attention.[2]

The Method of Analysis

A method of analysis may be right, however, even if the problems to which it has been applied are unimportant. We have now to consider, therefore, why it is that the tools of thought of the theory of the firm are unsuitable for dealing with the questions which I have posed.

The basic concept in the theory of the firm is that of equilibrium. This relates to the states of mind of the individuals who make up the economy. To be in equilibrium is to be content, in the sense that you see no possibility of improving your financial position unless there is a change in the conditions which confront you. As so defined, it is neither interesting in itself nor could its

[1] It is true that even quite small changes in the distribution of income could have substantial effects on the level of employment. Indeed, the possibility that monopolistic rigidity of prices might frustrate full employment policy was discussed at some length in the White Paper on that subject in 1944. But to introduce the possibility of unemployed resources is to vitiate the fundamental assumption of scarcity of resources on which the traditional theories of welfare and the firm are built.

[2] It is worth noting why the theory of the firm took the path which it did. It may be regarded as an attempt to establish and make more precise the familiar proposition that "monopoly leads to inefficiency, and to prices being raised through restriction of supply". It is, of course, relatively easy to demonstrate the truth of this proposition by argument from plausible assumptions about human behaviour, and the pre-industrial history of England was rich in examples of commodity cornering. But the prescriptive force of such demonstrations rested on the implied proposition that the lower price and higher output which would result if monopoly were absent would be, in some sense, more "right" or socially desirable. And once this was seen the hunt was up for a standard of the "right" price and output which could be defended as independent of the special prejudices of the individual disputant.

presence or absence be determined by any means other than asking the individuals concerned. The object of the economic theorist is to define the relationships between prices and between prices and costs which, on reasonable assumptions about human psychology, may be taken to leave people with the feeling that they are doing as well for themselves as they can. These relationships are the "conditions of equilibrium". They will vary with the rules of the game in force. The unique set of prices which corresponds to the optimum allocation of factors of production provides the standard against which alternative conditions of equilibrium may be judged.

It is explicitly recognised that in certain circumstances the concept of equilibrium ceases to be useful, the method of analysis breaks down. This is when a firm knows that the outcome of any action will depend on the reactions of other firms, which themselves will depend on what its own further reaction is expected to be. Such is the situation of a firm which shares a market with only a few others, which is in an oligopolistic market. In such circumstances the firm can never *know* where its best advantage lies; and so can never be in equilibrium in the sense defined.

So long as oligopoly can be regarded as an uncommon phenomenon this blank spot in the analysis is tolerable, though irritating. But the ubiquity of oligopoly is a conclusion which has perennially threatened to follow from the assumptions of the theory of the firm, and I do not believe that the conclusion can be avoided on any plausible line of reasoning.

I have mentioned that the original answer to the question "Why does the system tolerate firms of differing efficiency?" was that the costs of any firm begin to rise once it passes some optimum and not too large size. This also was the protection built into the theory against the generalisation of oligopoly. But the assumption that diseconomies of scale began to appear at a level of output which was small, relative to the market served, became increasingly difficult to reconcile with either facts or other assumptions in the theory, and the coming of age of the theory of the firm to which I have referred resulted from the work of a reconstruction agency called in to find a substitute assumption.

This could only be the assumption that, after some point,

the return on further expanding sales would begin to fall. In technical language, once it was admitted that the long period cost curve of a firm might be flat or even negatively inclined over all relevant ranges of output, the notion of equilibrium could be preserved only if the demand curve also was given a negative slope. And people were quick to discover convincing reasons why, if technology set no limit to the growth of a firm, the job would be done just as well by the difficulties and expense it would experience in expanding its market.

It seems to have gone unnoticed that the reconstruction was undone, almost casually, as soon as it had been completed. J. M. Clark pointed out in 1939 that the steepness of the slope of the demand curve of the individual firm would depend very much on the length of the period under consideration. Unless it be operating in something like an organised commodity exchange, an unmatched price cut will have no effect on a firm's business in the five minutes immediately following. But in five days the effect will be more appreciable, and so on through five weeks, months and years. Clark suggested that in most cases the long-period demand curve would be much more nearly flat than the theorists of the firm had usually implied, and drew the comforting conclusion that the evils of imperfect competition were likely to be less than people had supposed. But, although he pointed out that these evils were accompanied in their exit by the notion of an optimum size of plant, he did not apparently make it sufficiently obvious that he had dismissed also that of a maximum size of plant and, therefore, the conditions for equilibrium to have any meaning.

The point can be clarified by considering a particular highly simplified example in more detail. Suppose there to be a group of firms of equal efficiency which produce a physically identical product; and suppose that each firm benefits from the attachment to it of a proportion of the customers by force of habit. Then any firm which seeks to expand its share of the market will have to incur some expense in breaking the tyranny of habit. This may take the form of a cut in price or of, say, an increase in selling effort. But whichever it be, the important thing to remember is that, in non-oligopolistic conditions, there will be

no reaction from the other firms in the industry. It follows, therefore, that a finite *once-for-all* expenditure will produce a *permanent* transfer of customers. A bargain sale for a limited period will add permanently to the firm's faithful customers. The additional expense may be considered then as equivalent to an investment in a non-wasting asset which yields a perpetual return. Hence, the amount which must be subtracted from the long-period demand curve to arrive at a net-receipts curve is not the total expense incurred but merely the income which that amount would yield if invested in undated securities at the going rate of interest.

If the matter is looked at in this way, the plausibility of representing long-period demand curves as significantly different from horizontal becomes very slight. If both the cost curve and the demand (net-receipts) curve are almost horizontal, no near limit is set to the size of the firm when in equilibrium. And if there is no such limit, then it ceases to be possible to regard the oligopoly case as exceptional.[1]

Equilibrium, defined as a state of being content, ceases, therefore, to be applicable even within the system of analysis which is built upon it. Moreover, even were this not so, it would still be unusable for dealing with the two issues on which I have suggested that attention should be concentrated.

I said earlier that, although there was nothing wrong in dealing separately with inter-firm efficiency differences and progress (inter-temporal efficiency differences), the results must be integrated at the end. Indeed, if we return for guidance to the argument at the political rather than the academic level, we find that it is precisely the conflict between efficiency in the short run and progress in the long which lies at its centre. The standard reaction of the business-man to any proposal for "making conditions more competitive" is that it will kill the goose for the sake of the golden eggs. Cut-throat competition he will say—

[1] Those who reject my argument and continue to put their trust in rather steeply inclined demand curves are left with a further problem to which I see no obvious solution. Imperfect competition implies an "inefficient" distribution of factors, and is, therefore, to be deplored. But it is only because of these very imperfections that the system is preserved from monopoly or oligopoly, which are also thought to have bad effects on resource allocation. The theorist is then confronted with the problem of whether the bad effects of oligopoly are worse than the effects of the imperfections which prevent it. The rather sharp division of labour between the theorists of welfare and those of the firm may help to explain why this issue has received so little attention.

the subject is rich in highly-coloured terms—might indeed eliminate the weaker firms in an industry, but only at the expense of weakening the stronger also. Or, alternatively, unregulated competition will lead to alternate swings from excess to undercapacity, with the profits of the boom never compensating sufficiently for the losses of the slump to permit a high rate of technological advance.[1]

In fact, however, the necessary integration at the theoretical level has never taken place. Analysis of the factors which determine the rate of progress has, by and large, fallen to different hands from those engaged with the theory of the firm, and, in any case, is still but little developed. And when the streams have, on rare occasions, come together, they have run side by side instead of mingling. We find ourselves back at the Cleopatra's nose end of the spectrum, with a Schumpeter maintaining that, in his opinion, changes in the technique of production are so important as to render otiose the study of the efficiency with which any particular technique is employed, while the continuing practice of the theorists of the firm implies an exactly contrary opinion.

Nor is this lack of integration accidental. The concept of equilibrium makes it impossible The fundamental assumption in the theory of the firm is that the path by which equilibrium is reached, and the speed with which firms move along it, have no influence on the nature of the equilibrium which is achieved or tended towards. The entrepreneur is taken to act as if none of the objective conditions which confront him will change, for only on this assumption can he be taken to know what it would be like to be in equilibrium. But progress is only conceivable if it can be assumed that entrepreneurs are trying to change these conditions—techniques of production, structures of consumer preferences and the like. In other words, any attempt to marry the analyses of inter-firm and inter-temporal efficiency differences immediately dissolves the concept on which the former is founded.

There is, in short, justification for the suspicion of the concept of equilibrium which arises at the most primitive level of thought.

[1] A strict presentation of this type of argument in a deliberately extreme form has been made by Mrs. Robinson. See her paper on "The Impossibility of Competition" in *Monopoly, Competition and Their Regulation* (ed. Chamberlin).

The most fundamental characteristic of a capitalist economy is growth and change, and this implies not content but discontent. In a peasant economy the emphasis is on income; in a capitalist economy it is on capital. In the former, attention is concentrated on doing as well as possible in the given unchangeable conditions. The latter is characterised by a restless urge to do better, to change the conditions, lest, through inaction, they are changed against you. The peasant age is one of contentment, this is an age of anxiety. Hence a concept of equilibrium which emphasises contentment can scarcely have much explanatory value for a market economy.

I conclude, therefore, that we must start again, by abandoning the concept of equilibrium as defined in the theory of the firm. This is easy. Building a new set of concepts is not so easy. To this I now turn.

Chapter III

ON CONCEPTS AND DEFINITIONS

FOR the benefit of those who have eschewed the aridities of the previous chapter, I will recapitulate. The problem with which we are concerned is whether various rules of the game are or are not contrary to the public interest. I suggest that, *prima facie*, the public interest is most suitably defined as what the public is interested in. In the post-war period the public has shown an abiding interest in two aspects of business performance; the existence of wide differences in the efficiencies of firms in the same industry—the dispersion of efficiency; and the rate at which industrial efficiency increases—progress. I believe that we shall be wise to concentrate attention on these two criteria for judging whether particular rules improve the game or spoil it. The traditional theory of the firm has concerned itself more with other issues. Nor are its concepts suitable when the direction of enquiry is changed. Hence, some fairly extensive reconstruction is required. The object of this chapter is to define three concepts; the industry, efficiency, and, in a preliminary way, the rules of the game.

The Industry

Concepts are not entities with an independent existence which wait to be discovered. They are simply systems of classification created by men to help in solving specific problems. If, therefore, we are to produce a concept of an industry which is useful for analysing the problem under consideration we must look more closely at what is implied when we say that the spread of efficiency within an industry is greater than it ought to, or need, be.

What we are suggesting, I think, is that things would and could be better if some ("the more efficient") amongst a group of firms took over the business of others ("the less efficient"). It is clear, therefore, that a necessary condition for a group of firms to be classified as an industry is that one should be able to take over the business of another and to conduct it reasonably effectively.

Now we shall feel doubtful of the ability of one firm to do the job of another when, as we say, they are "in a different line of business". And in saying this we shall be looking primarily at the technological character of the productive operations which the firms perform. The thought is that a firm's ability, say, to weave cotton threads into cloth gives little or no indication of the ability it would show in turning iron ore and associated materials into pig-iron, even when it had learnt how to do so; and that, in any event, it is unlikely to know how to do so and would take a considerable time to learn. This, of course, is not all. We have in mind that knowledge of the markets in which materials are bought and products are sold is something which the firm needs in either line of activity. Such knowledge becomes the more important in our minds the more nearly do the activities of the two firms approach each other from a technological standpoint. But it remains, I think, essentially peripheral to the main distinction of technological experience. I propose, therefore, to define an industry as a group of firms whose techniques of production are sufficiently alike for it to make sense to conceive of one as being able to do the business of another.

Like all systems of classification, this will give rise to difficulties in practical application. In one sense every firm is a specialist, since no two firms will be conducting their operations in an identical fashion. There is, therefore, an inescapably arbitrary element in how we choose to define a separate technique. And, having chosen a definition, there will always be troublesome borderline cases. Should this particular firm be classified as operating this technique or that or, perhaps, both? But such difficulties are not peculiar to economics. The fact that we sometimes find it difficult to decide whether a particular colour is purple or blue does not lead us to conclude that the distinction between the two colours is useless or impossible. In fact, to say that the system of classification is arbitrary and requires judgment in its application is merely to repeat that concepts are made and not discovered.

The definition of an industry which I am adopting is, of course, very close to that used by the authors of the Standard Industrial Classification in the United Kingdom and by the compilers of

comparable official systems in most countries. The nature of its productive process is one of the cardinal criteria which determine the trade or industry to which a firm is classified.[1] The official statisticians who apply these definitions have their difficulties with marginal cases, and the decision made by one man may not be the same as that which another would make in his place. But decisions are made; because they have to be. And this is surely the essential point. The impossibility of devising and applying systems of classification in such a way as to avoid debatable cases leads to paralysis of the will only when the classifying activity is motiveless or, as we say, academic. Having found a motive, in the shape of a question for examination, and defined the concept of an industry in a way which is useful for that examination, we can push on without an interminable specification of the dubious cases, ready to deal with such difficulties *only as they arise.*

For the purpose of the empirical investigations in the later *chapters I have* acted as if the definitions and the practice of *the takers* of official statistics were, in fact, "correct". The 156 trades into which manufacturing firms have been grouped seems to me about as fine a classification as is required for the purpose to which we are addressing ourselves, and there will be occasions, which will be indicated, when a broader grouping such as the "Order" of the Standard Industrial Classification is a more appropriate way in which to think of an industry. Like the official statisticians, I find it sensible that in cases where a firm operates two different techniques it should be shown as in two industries.[2]

[1] For an account of the basis of the American system of classifying manufacturing *firms into industries,* see Conklin and Goldstein in *Business Concentration and Price Policy,* Princeton, 1955.

[2] It should be noted that the definition of industry adopted here is different from that in most versions of the theory of the firm. The industry (market) is there defined in terms of a product, which is delineated in turn by a gap in the chain of substitution. The distinction between the two definitions is usually hidden by the assumption of single-product firms and single-process products. Once this assumption is removed, however, there is the problem of translating statements in which industry is used in one sense into statements in which it is used in the other. To integrate statements about markets with statements about industries is one of the most difficult problems in economics.

Some theorists have invented a further problem for themselves, by doubting the existence of gaps in the chain of substitution, on which the concept of a market dissolves and they are stranded in the inhospitable immensities of general equilibrium theory (e.g. R. Triffin, *Monopolistic Competition and General Equilibrium Theory*). I think myself that this is very largely a pseudo-problem, which arises from the vain attempt

Efficiency

We all know what we mean by efficiency in a rough and ready sort of way. This is my excuse for having used the term without definition so far. Firms take from the stream of productive resources and are able thereby to add to the stream of products. One firm is counted as more efficient than another if the relation between what it adds in products and takes out in productive resources is more favourable; that is, if its output/input or conversion ratio is higher; in more usual terminology, if its costs (aggregate productive resources used) per unit of product output are lower. But, although the general idea is clear, difficulties begin to cluster thickly around as soon as we try to make the notion of a conversion ratio or unit cost measure more precise.

Consider the hen.[1] With some violence to reality, we can conceive of it as nourished by a single input (grain) and as producing a homogeneous output (eggs). With a further effort, we can abstract from differences in the size of eggs, in the cunning of hens in concealing or their voracity in consuming them, and from a variety of other troublesome factors in the poultryman's life. And by conceiving of all hens as being sold to a canning factory at a fixed price at the end of their useful life we can dispose of the problem of scrap value. It might then seem that by taking total output of eggs and total input of grain over the whole life of each hen we could produce a grain/egg conversion ratio which would enable us to say unambiguously that one hen was more efficient (a better layer) than another. But this is not so.

It is common knowledge that any hen which reversed the order of nature by costiveness in spring balanced by a flux in the winter would greatly profit its fortunate master. And a hen which begins

to apply perfectly sharp classifying principles to a reality which is not itself clear-cut. It must be recognised, nevertheless, that the bone of contention in the spate of learned argument which immediately follows an Anti-Trust decision, or an attempt by a rash economist to measure the "competitiveness" of the American economy, is, typically, not the facts of the case but, precisely, what grouping of firms into industries is significant for the purpose of the argument. All possible attitudes are usually found, from an anxious propensity to see a market whenever two or more are gathered together to a brash certainty that all is well by definition, since, in the last analysis, all products compete for a share of the consumer's dollar.

[1] Mrs. Robinson has found the robin a more tractable bird. But only, I think, by ignoring differences between the psychic and physiological values of different sorts of insect which the robin may well think important. See "The Production Function", *Economic Journal*, No. 257, Vol. LXV.

to lay early and thereafter produces a regular flow is greatly to be preferred to one which remains obstinately unfruitful to an advanced age, even if it should belatedly make amends by veritably cornucopial behaviour. In other words, in assessing the efficiency of hens we are concerned not only with the relation between total input and total output but also with the relative time-patterns of both those flows.

Firms are commonly more complicated organisms than hens, and the difficulty in defining their conversion ratios (unit costs) in such a way as to indicate unambiguously their relative efficiencies is correspondingly greater. There are three major difficulties; first, different firms will frequently produce different products, and even one firm will usually produce a variety of products; second, the inputs of firms are equally heterogeneous; third, one element of input—capital—carries with it all the problems connected with time which confronted us in the case of the hen, and is usually a sufficiently important element to make it impossible to ignore them.

I do not pretend to have solved these difficulties in a fully satisfactory fashion, nor, indeed, to have anything very new to say on the subject. Since this is not a book on capital theory or on welfare economics in the technical sense, I shall skate over in a rather cavalier fashion many issues which specialists will consider important. My aim is simply to show what is implied by adopting, as I have, a definition of efficiency which has the merit of being measurable in statistical practice.

Let us first examine the problem created by the heterogeneity of inputs, ignoring at this stage the special problems associated with capital. The productive resources which the firm draws from the stream consist of labour of different kinds, materials and components, energy, services and so on. We reduce these to common units, as is necessary if total input is to consist of a single figure, by measuring each factor by its money cost to the firm, and so arrive at a figure of total costs. This procedure implies that, in some sense, relative prices measure the relative values as producing agents of the different resources which the firm withdraws from the stream.[1]

[1] More precisely, the relative scarcities of the different factors, given the total supplies of each, the pattern of demand for final products and the techniques of production which can be employed.

I do not intend either to go into detail on what this sense is or to establish that relative prices "truly" measure what they are supposed to measure. For throughout this study I am working on the principle that effects which in practice cannot be detected, directly or indirectly, should be ignored. In other words, it is no use worrying about imperfections which cannot be removed. Hence, we need question the validity of using market prices to aggregate heterogeneous inputs only if there is reason to suppose that they are significantly "wrong" in some detectable sense. And, of course, the question would arise with particular force if there were reason to think that the "distortion" resulted from the rules of the game, whose influence on the dispersion of efficiency we are trying to establish. For in that event our measure of efficiency would itself vary with the rules. I do not think, however, that, so long as capital is being excluded from consideration, any such "distortions" as can be detected are likely to be significant enough to warrant serious dissatisfaction with the independence of our efficiency measure.

An example may make things clearer. Suppose there to be two firms in an industry, each of which buys only labour and a material, and let the outputs of the two firms be identical in quantity and quality.[1] Firm A employs 100 workers and buys 50 units of the material; Firm B employs 90 workers and buys 55 units. If wages are £100 per annum and materials cost £200 per unit, then the total costs of the two firms are equally £20,000, and their efficiencies identical by definition. If, however, the cost of materials were only £100 per unit, Firm B, which employs relatively fewer workers and relatively more materials, would have lower total costs and, therefore, we should say it was the more efficient.

Now if it could be shown that the two relative price situations corresponded to two different sets of rules of the game we could no longer regard the efficiency measure as independent of the rules of the game, and we should need to plunge much more deeply into the problem of factor price determination. And it is sometimes alleged that this in fact is so. It is said, for example,

[1] Strictly, we must also assume that there is no time-lag between input and output if the problems connected with capital are to be completely excluded.

that an industry in which firms band together to buy and sell is likely to extract better terms from its suppliers and customers, and that part of the resulting net revenue will be extracted by its workers in the shape of higher wages.[1] What we have to ask ourselves is whether the validity of this proposition could be established in practice. I do not think that it could.

If the proposition is to be tested, there must be some independent measure of the comparability of different units of a factor, so that it can at least be established that the same factor is being sold at different prices to different industries. If we are not too finicky in our definition of comparability we can go some way towards getting such independent measures. In the example above we can conceive of a physical characterisation of the material which would enable the price paid in this industry to be compared with that paid by firms in other industries. But I find it hard to conceive of a similar treatment for labour which would make it possible to say that workers were being paid more (or less) than workers of comparable efficiency in other industries. Indeed, it is difficult to give any very clear meaning, even in principle, to the proposition that the workers in one industry are more efficient than those in another. This being so, only a minor part of the proposition is open to verification.

If we cannot conceive of how, in practice, to improve on the material we have, this indicates, not that we should fall into neurotic inaction, but that we should act as if what we have is what we want. I shall assume, therefore, that market prices properly reflect the relative values as producing agents of the various resources which firms employ. And by market prices I mean the prices actually paid by individual firms and not some sort of industry-wide average. For there are the same obstacles to producing independent measures of identity for units of "the same" factor used by firms within an industry as there are for inter-industry comparisons. I leave it to the semantic taste of the reader (for I myself am unsure) to decide whether I am ignoring or denying the reality of the traditional problem of the influence of

[1] A large part of the traditional theory of the firm is, of course, concerned with the effects of differing degrees of competition on the distribution of factor income. For the reasons given in Chapter II, I prefer to adopt a less sophisticated presentation of the problem.

the degree of competition on the relative prices of current factors.

The problem of capital, however, cannot be assumed away. The need for a distinction between capital and current inputs arises simply and solely because time must elapse between the purchase of inputs and the sale of the products which the inputs are used to produce. The firm begins to draw on the stream of inputs at an earlier date than that at which it begins to contribute to the stream of output. What we call the capital employed in the business is, at any point of time, the value of the inputs which the firm is, so to speak, carrying within itself. In terms of the hydrodynamic analogy I have been using, it is the amount of water in the pipeline which connects the stream of resources with the stream of final products. For some inputs, the period of time which elapses between their purchase by the firm and its sale of all the products which they have helped in producing is very long. These inputs are distinguished by a special name; they are called fixed capital.

Neither society nor the individual is indifferent to the length of time which elapses between the taking in of input and the appearance of output. At the most primitive level this is because the human stomach cannot wait. Beyond this point it is because men are mortal. If our lives were eternal the time at which any event other than eating took place would be of no moment. Not being indifferent, however, we must attribute some cost to the capital employed in the business in addition to the cost of the labour, materials, etc., of which it is made up. To decide what this cost should be is the first problem associated with capital.

The second problem of capital arises because, whatever length of period we take over which to measure the input/output ratio, there will always be some elements of input for which the question "What was their cost?" cannot be answered by simple reference to the prices originally paid for them. For they will have been in existence in a part-worn form at the beginning of the period and/or will still be in existence in part-worn form at its end.[1] Hence, before we can measure either the value of inputs

[1] The problem would not arise if all the fixed capital assets of the firm were bought at the same date and had the same length of life, and if the period over which we measured the input/output ratio were equal to the length of life. But merely to describe the exception is to demonstrate its formality.

during the period or the capital employed in the business we have to decide how much of a deduction from the prices originally paid must be made to allow for their being part-worn. This is the problem of depreciation, which can be stated alternatively as the problem of how the cost of a fixed capital asset, which is paid for when it is first acquired, should be divided between different parts of the whole period during which it remains in existence.

The second problem, unlike the first, is formulated in the terms I have used by firms themselves; they make explicit decisions about depreciation. The question is, therefore, whether there is any reason for rejecting the answer which firms arrive at.

Given its cost, the depreciation charged against a fixed capital asset during any period shorter than its life will depend on its length of life and the rule employed by the firm as to how the total cost should be allocated over that lifetime. Note that we are not concerned with what its length of life *ought* to be but what in fact it is. After the event, therefore, when the asset has been scrapped, differences of view as to its depreciation during any part of its life can arise only because of differences about the allocation rule. If, however, we want to measure what depreciation should be in a less historical fashion, at some point before the end of the asset's life, we are forced into forecasting how much longer it has to live. This is not something which can be determined wholly, or even largely, from technological data. The asset will be replaced when it is thought profitable to do so, and this will depend, amongst other things, on the course of prices in the future. Hence, there is no incorrigible answer to the question, "What is the remaining life of this asset?" A unique answer is possible only when the question is in the past tense. But it seems reasonable to suppose that the answer given by the firm itself—that implied by the depreciation it charges—will, in the event, prove to have been a better forecast than that of an outsider.

There is no answer to the question, "How should the cost of a fixed capital asset (now retired) have been spread over the course of its (known) life?" The question is too unqualified, and can be answered only if we know the purpose for which the

knowledge is being sought. For our purpose it is desirable to have a rule for spreading the cost which will leave the efficiency of the firm unaffected by the particular point in the lifetime of its assets at which we happen to measure it. Since expenditure on repairs and running costs in general will usually increase as the asset gets older, a depreciation rule which allocates a larger part of the cost of the asset to the earlier years of its life will be "right" for our purpose. This is what is done under the "reducing balance" method of depreciation, and I understand that this system is followed by the majority of businesses.[1]

I conclude, therefore, that we cannot do better than accept the practice of firms, as to both length of life of assets and method of spreading their cost over their life. The "correct" figure for depreciation during any period (i.e. for the cost of the input of fixed capital assets) is then the depreciation which firms charge. It follows that the capital employed by a firm is what the firm says it is.[2]

But we cannot deal with the first problem of capital in this way. For firms do not customarily ask themselves, "What is the cost of employing this amount of capital?" Firms, however, do ask themselves, "What can or do we earn by employing such and such an amount of capital?" It is tempting, therefore, to say that the cost of employing capital is what capital earns; the rate of profit on capital. The difficulty is that capital earns different amounts in different firms. I say difficulty, because in one perfectly good sense, and that the sense which is relevant at this stage of the argument, capital is the one resource where we can say indubitably that every unit is alike. It is only when we measure capital employed as a sum of money that we can speak of a rate of profit. And the interchangeability, lack of differentiation, of different units is one of the prime characteristics of money. Nor

[1] The "reducing balance" method is termed "normal" by the Board of Inland Revenue. (See *Income Tax Wear and Tear Allowances for Machinery or Plant; List of Percentage Rates*, H.M.S.O., 1950.) Under this system depreciation in any year is a constant proportion of the *written-down value* of the asset at the beginning of the year. Under the "straight line" method a constant amount is written off each year.

[2] This refers, of course, to tangible assets. Such assets as "goodwill" have no place in our measure. Indeed, in so far as they represent a capitalisation of future earning power, their inclusion would tend to make all firms appear as equally efficient by the measure we are using.

is this disposition to regard capital as undifferentiated simply something into which we are pushed by the difficulties of measurement. I have shown that the cost of employing capital, as distinct from the cost of the commodities which constitute capital employed, is simply the price which is paid for the ability to wait from seedtime to harvest.[1] And the ability to wait is given by the possession of command over commodities in general. From the present point of view, therefore, capital is rightly regarded as a stock of money.

It is clear, therefore, that if we are to follow the temptation to define the cost of employing capital as what capital earns—and I see no defensible alternative—we have to perform some sort of averaging operation on the different rates of profit which this homogeneous resource does in fact earn. There are limits, however, to the area over which this average should be struck. By accepting the prices which firms in one industry actually pay for labour, materials, etc., as the appropriate factors for aggregating these resources into one measure of input, we are implying that the rates of profit earned in the supplying industries are the proper measure of the cost of employing capital in those industries. Consistency demands, therefore, that the rate of profit earned in the industry under study be taken as the cost of employing capital in it; the scope of the average should be limited to the industry in question. This is the smallest as well as the largest area over which the average should be struck. For we have no reason for discriminating between the firms within the industry,[2] and, since we have accepted the homogeneity of capital employed when measured in money, there are no difficulties in the way of striking an average.

The cost of employing capital in a firm during any period is given, therefore, by the product of the value of the capital employed (as measured by the firm) and the average rate of profit in the industry. And the average rate of profit in the industry

[1] If there were no time-lag between input and output there would be no capital employed, except in so far as there was a lag between production and sale. But for this last lag, the distinction between employer and employed would disappear.

[2] To define the cost of employing capital in the individual firm as its own rate of profit would, of course, defeat the whole object of the enquiry by making the efficiencies of all firms identically equal.

is an average of individual profit rates weighted by capital employed.[1] It follows from the same argument from consistency as in the previous paragraph that the rates of profit in question relate to profits before taxation.

This completes the discussion on the construction of the input side of the ratio which I use to define and measure the relative efficiencies of the firms in an industry. The total of inputs is equal to the sum of current inputs purchased by a firm, valued at the prices which it pays, depreciation as calculated by the firm, and the cost of employing capital as just defined.[2]

If all firms in the industry produced the same product our difficulties would be at an end. Output for each firm could be measured in the physical ("natural") units most appropriate and relative efficiencies could be defined by comparison of the ratios $\frac{v+\bar{r}k}{q}$, where v is the cost of current inputs and depreciation, $\bar{r}k$

[1] Let k be the value of capital employed, z be the amount of profit, and Σ the summation sign; then the average rate of profit

$$\bar{r}=\frac{\Sigma\frac{z}{k}.k}{\Sigma k}=\frac{\Sigma z}{\Sigma k}$$

and the cost of employing capital in any firm $i=\bar{r}k_i$.

[2] I have assumed in the discussion above that both net profit and capital employed have unambiguous meanings when once the life of assets and the depreciation rule are known. The assumption is unjustified if the prices of inputs change over time. For we have then to decide whether to value capital (including working capital), and therefore gross profit and depreciation, at original or replacement cost. The value of inputs will be different according to which we do. In so far as the share of capital in total inputs varies from firm to firm, relative efficiencies will vary also. And this is all the more true if the average age of capital differs from firm to firm.

It seems clear that both depreciation and capital employed should be reckoned at original cost. For what we are trying to do is simply to sum the expenditures *actually incurred* in producing the output of any given period. Depreciation is simply the proportion of expenditure on inputs which are still usable in the next period which is properly attributable to the output of the current period. And we have defined the capital employed in the firm as the original cost of the inputs which the firm is carrying within it, because of the time-lag between input and output. Since most firms use original cost in drawing up their accounts, this treatment is statistically convenient as well as conceptually appropriate. Whether it is prudent for a firm to act as if depreciation on original cost would enable it to stay in business indefinitely is another question, which need not concern us at this point.

One implication of using original costs is that two firms which were in every other respect identical would differ in efficiency by our measure if the average ages of their capital stock were different and prices had changed over the period during which their capital had been bought. This is uncomfortable but unavoidable. All we can do is to take note that if price movements are substantial and there are large differences in the average age of capital the meaning of efficiency comparisons becomes dubious.

is the cost of employing capital, and q is output. But if different firms produce different products, or if, although all firms produce the same range of products, they produce them in different proportions, we are up against the same sort of difficulties as with heterogeneous inputs.

What we were looking for in the input case was some means of answering the question, "How many men is such and such a machine worth in terms of its value as a producing agent?" We answered this by saying that a machine costing £2,000 is worth as much as 5 man-years if wages are £400 per year.[1] In the case of output what we need is some way of allowing for the relative *difficulty* of different objects as problems in production. It is clear in a common-sense fashion that it is more difficult to build a battleship than a ball-point pen, and the problem is to find some way of determining how many ball-point pens represent a production task comparable to that presented by a battleship.

This may not, indeed, be the only thing in which we are interested. It makes perfectly good sense to say that a firm which produces battleships when the public really wants ball-point pens is less efficient than one which beats swords into pens. But this is a different sense from that which we have so far been considering. So long as we are concerned with *productive* efficiency it is perfectly legitimate to compare the efficiencies of two firms, even if their respective products are totally useless. And what we need for such a comparison is a means of expressing the units of one product in terms of units of another which satisfies us that proper allowance has been made for the relative difficulty of producing the two.

Clearly, the only practicable measure of the relative difficulty of producing two different products is provided by their relative costs. If, therefore, we have two firms each producing a different product and each the sole producer of that product, it is impossible to compare their efficiencies. For the conversion ratio which we have to use to express the output of one firm in terms of that of the other is precisely the ratio of the costs which we

[1] This does not mean that if a particular firm at a particular time were asked to choose between the machine and 4 men-years as a non-transferable gift it would necessarily take the former. Individual choices will depend on individual situations.

wish to compare. If, however, each of the goods is also produced by a number of firms, the problem becomes manageable. For we can then take averages of costs (input) per unit of output over each group of firms and use this to turn the outputs of the two firms into comparable units. While the conversion ratio is not independent of the efficiencies of the two firms being compared, this need not worry us unduly so long as neither is preponderant in its own group.[1]

This device will not serve, however, in the normal case where each firm produces a number of products. For, however fine the process of cost accounting, there will always be a residue of inputs which cannot be allocated between different products. And this means that it is, in principle, not possible to measure total costs (inputs) per unit of output for the individual products. The best we can do is to measure "direct" costs per unit of output, albeit on a somewhat wider definition of "direct" than that customarily used by business-men.

We are faced, therefore, with the need to choose a practical *pis-aller* for forming the conversion ratios. The choice is between direct costs per unit of output on the one hand and prices on the other. And, if we choose prices, we can use either industry wide average prices or reduce the heterogeneous output of each firm to a common sum by using the prices which the firm itself charges for its various products.

On grounds of convenience the use of firms' own prices is obviously greatly to be preferred. For the measure of efficiency which I am proposing would then be:

$$\frac{v+\bar{r}k}{q_1p_1+q_2p_2 \cdots q_np_n}=\epsilon\text{—efficiency index,}$$

where $q_1 \ldots q_n$ are the outputs (in physical units) of different products and $p_1 \ldots p_n$ are the prices of those products.[2] The sum of the constituents of the denominator of this expression is then the value of output or turnover(s). We may therefore write:

$$\epsilon=\frac{v}{s}+\frac{\bar{r}k}{s}=\frac{v}{s}+\bar{r}\beta$$

[1] What we are doing in effect is to compare the efficiencies of the two firms relative to their respective groups.

[2] The smaller is ϵ the greater the efficiency.

where $\beta=\frac{k}{s}$ is the familiar capital/output coefficient.

By definition $r=\frac{s-v}{k}=\frac{1}{\beta}-\frac{v}{k}$

and $\frac{v}{k}=\frac{v}{s}.\frac{1}{\beta}$

so $\frac{v}{s}=1-r\beta$

and $\epsilon=1-\beta(r-\bar{r})$

and β, r and $\bar{r}$ can all be derived by fairly simple processes from the business accounts of firms.[1]

Let us consider, therefore, what error is involved in following the dictates of convenience.

By using prices instead of costs, we are mingling the two aspects of efficiency which I earlier distinguished; "technical" efficiency in producing whatever it is that the firm chooses to produce and skill in choosing what the customer will be willing to buy ("market efficiency"). The firm which has mistaken the fashion in hats and is having to dispose of its models to charladies at give-away prices will appear as inefficient overall, even if its productive efficiency is peerless. Both these aspects are important, but it is very desirable that they should be separated.

Moreover, and this is particularly important, particular rules of the game may be one reason for the efficiency of some firms appearing to be very high. Suppose that certain firms only within an industry succeed in banding together to fix a price which yields an abnormally large margin over direct costs for a product which only they produce. Then their efficiency, as measured by ϵ, will tend to be higher than other firms outside the band, although there may be no superiority in "technical" or "market" efficiency which corresponds to this.

All these defects of our index become the greater when the prices of individual firms are used instead of industry averages.

It is usually possible, in economics as in other fields, to invent moderately plausible reasons for doing what one wants, when it is not what one ought. Thus it is not difficult to produce special

[1] British firms do not usually publish the value of their turnover, so that β cannot be calculated from published accounts.

cases where the use of firms' own prices would give a better measure of efficiency than industry averages. For example, two articles which are to all physical appearances identical may yet differ as problems in production and serve a different market; a council house on a big estate may be cheaper to build, and sell for a lower price, than an identical house alone on a moor. But in one perfectly good sense the large firm which produces the one may be no more efficient than the small builder who takes on the other.

After all such special pleading, however, the fact remains that to use a firm's own prices will, in general, be less satisfactory than to use average prices, and to use prices will be less satisfactory than to use direct costs. It is reasonable to hope that the truth will not be overmuch obscured by following the path of convenience. But I must admit that my choice of the measure defined above rests ultimately on the belief that it is the closest approach to the ideal which is likely to be accessible in practice.

Some Characteristics of the Efficiency Index

It is evident that efficiency as I have defined it is not necessarily identical with profitability, in the sense of the rate of return on capital which a firm earns. For the term β, the ratio of the value of capital employed and turnover, appears in the definition. If two firms have identical capital/output ratios, it is certainly true that the more profitable is the more efficient. But if their capital/output ratios differ, this no longer always follows.

Suppose that both firms are earning rates of profit on capital above the average for the industry, firm A earning 20% and firm B 15%, while the average for the industry is 10%. B will nevertheless be of greater efficiency than A (have a lower value for ϵ) if its capital/output ratio is more than twice that of A. Conversely, if both are earning less than the average rate of profit, the one with the lower rate of profit may nevertheless be the more efficient if its capital/output ratio falls sufficiently below that of its rival.

The common-sense meaning of all this is fairly clear. We have defined the cost of employing capital to be the average rate of profit which capital earns in the industry. In a firm whose rate

of profit exceeds the average, capital is then earning more than its keep, and conversely in a firm whose profit rate is subnormal. Hence, the greater the importance of capital in total inputs, the better, in the first case, and the worse, in the second.

It is unlikely, however, that capital/output ratios will differ in practice by the amounts which the above example shows to be necessary to reverse the efficiency ranking indicated by rates of profit on capital. The latter will, therefore, usually provide a fair guide.

The second characteristic of the efficiency index is that it will be influenced by changes in output prices, and would continue to be so affected even if the outputs of different firms were valued at common prices instead of those of the firms themselves. The general logic of this is again fairly self-evident, from inspection of the initial formulation of the index as $\epsilon=\frac{v+\bar{r}k}{s}$. If the price of output rises, the costs of current inputs remaining unchanged, the denominator(s) of the index will change by the same proportion for all firms and v will be unaffected. But the average rate of profit in the industry ($\bar{r}$) will now be greater, so that ϵ will change by different amounts for different firms unless the ratio of k to v is the same for each. In other words, since we have defined the cost of employing capital as the average rate of profit, input/output ratios will be affected in different ways by a rise in the cost of employing capital if firms differ in the degree to which they employ capital.

The general effects of price (and rate of profit) changes can be defined more precisely by reverting to the alternative formulation $\epsilon=1-\beta(r-\bar{r})$.

The average of ϵ for the industry ($\bar{\epsilon}$) is evidently equal to 1 in all circumstances, which is simply to say that ϵ measures the divergence of the input/output ratio for any firm from the average for the industry.[1]

Suppose now that output prices change by a percentage x (which may be positive or negative), all input prices remaining unchanged, and let us denote the new values of the different variables by primes.

[1] $\bar{\epsilon}$ is, of course, arrived at by weighting the individual ϵ's by turnover(s).

Then

$$\beta' = \frac{\beta}{1+x}$$

$$r' = \frac{s'-v}{k} = \frac{s(1+x)}{k} - \frac{v}{k} = r + \frac{x}{\beta}$$

$$\bar{r}' = \bar{r} + \frac{x}{\bar{\beta}}$$

$$\beta'(r'-\bar{r}') = \frac{1}{1+x}\left[\beta(r-\bar{r}) + x - x\frac{\beta}{\bar{\beta}}\right]$$

$$\epsilon' = \frac{\epsilon}{1+x} + \frac{x}{1+x}\frac{\beta}{\bar{\beta}}$$

It is immediately obvious that the only circumstance in which ϵ' can be equal to ϵ is when ϵ itself is equal to $\frac{\beta}{\bar{\beta}}$. Hence, it will be the exception rather than the rule for ϵ to be unaffected by price changes.

In general, if x is positive, $\epsilon' \lesseqgtr \epsilon$ according as $\frac{\beta}{\bar{\beta}} \gtreqless \epsilon$.

And, if x is negative, $\epsilon' \lesseqgtr \epsilon$ according as $\frac{\beta}{\bar{\beta}} \lesseqgtr \epsilon$.

The extent to which price and rate of profit changes will on the average increase or decrease *differences* in efficiency between firms will clearly depend on the relationship between capital intensity and efficiency, and no generalisation is possible unless this relationship is known. However, one special case is worth clarifying, for it arises in a later chapter.

Suppose that inter-firm differences in the capital/output ratio are small enough to be neglected. Then for each and every firm $\frac{\beta}{\bar{\beta}} = 1$, and it follows that for all firms where efficiency is below the average $\frac{\beta}{\bar{\beta}}$ will be less than ϵ, and conversely for those above the average. It follows further, from the conditions summarised above, that a rise in price will reduce ϵ for all firms of below average efficiency and increase ϵ for the more efficient. That is to say, a rise in output prices will, on the average, reduce efficiency differences. It follows from similar reasoning that a fall in prices will increase the differences.

These conclusions may, indeed, be perceived intuitively. If one element in the efficiency index ($\beta\bar{r}$) has an identical value for all firms, anything which increases (reduces) its magnitude in the total index, as does a rise (fall) in $\bar{r}$, will reduce (increase) the differences between the different values of ϵ.

The Rules of the Game

It is not possible to analyse in detail the concept which I have called the rules of the game without a fuller discussion of the nature of the selling process than is convenient at this point. For the moment, therefore, I shall indicate the nature of the concept only briefly, and in addition I shall continue to assume that the term "price" is self-evident in meaning and free from ambiguity.

I exclude from the concept rules of prudent conduct which are so automatically accepted and obviously "reasonable" as to require no codification and to occasion no discussion. These I regard as implicit presuppositions rather than rules of the game. For example, the normal business-man does not need to have it impressed upon him that to carry his efforts to break a competitor to the point of breaking himself is foolish conduct, just as it is unnecessary for the laws of cricket to lay down that balls should not be stopped with the head. Firms sometimes ruin themselves in ruining another, just as cricketers are sometimes struck on the head. But in both cases we presuppose miscalculation rather than deliberate policy. The intent being contrary to the instincts of a reasonable man, there is no need to legislate against it.

On the other hand, I exclude also some rules which have been codified even in legislation or common law, if these are so generally accepted as also to have come to be taken for granted. The laws against passing off, or for protection of trade marks, are of this kind. So also is Limited Liability, which is akin to the *de facto* protection against trial for manslaughter enjoyed by the participants in a sporting game. Codification is neither a necessary nor a sufficient condition for something to be termed a rule of the game, although, as a matter of fact, most of the rules which have figured in recent discussion in the United Kingdom have been written rather than informal.

Coming now, after these negative preliminaries, to the positive content of the concept, I wish it to be understood to cover the following:

(1) The typical issue on which public discussion centres is an agreement, written or informal, between the firms in an industry that each will forswear certain courses of action which, in appropriate circumstances, could benefit the one which practised them. The implicit preamble to such agreements is that if one firm so acts all will do likewise, and, in that event, none will benefit. The reports of the Monopolies and Restrictive Practices Commission are filled with examples, the most common item of agreement being on restrictions on the freedom of action of the individual firm to vary its prices. Such agreements are analogous to the rule against dangerous tackling in a game of football. The team which indulged in dangerous tackling might win the game, not on its merits, but by injuring or intimidating the opposing side. But if the other side retaliated the result of the game might then be unaffected, leaving injury to the players and an unedifying scene for the spectators as the sole result of the action resorted to. As I have already said, spectators and players may take different views of where the line should be drawn between vigorous and dangerous tackling, and it is a dispute of precisely this kind with which we are concerned.

(2) It is customary, but not inevitable, that there should be some apparatus of coercion to ensure adherence to the agreements which are concluded. In so far as this is the sole function for which they are used, they stand or fall with the agreements themselves and raise no new issues of principle. But they may be—and it is suggested that they often are—used for another purpose; to prevent new firms coming into the industry. This is as though the contestants in a game were to say that anyone coming new to the game should not only abide by the rules of conduct followed by the others but also accept new obligations, from which the veteran players were excused merely on account of their long service.

(3) The final type of rule with which I am concerned is one which has not yet been applied in the United Kingdom and which is never likely to be established successfully by firms

themselves, even if some of them think it desirable. This is the rule, which is pressed for by those who see danger in "monopoly" as well as "restrictive practices", to limit the size (or alternatively regulate the conduct) of firms. The argument here is if that one player is too large he may win, not because he is good at the game, but because his mere size is sufficient to crush or daunt all others. It is against this sort of situation that Section 2 of the Sherman Act is directed. And it is important to remember that the reformers are concerned primarily, not with the means by which such size is achieved, but by the consequences for the game of its existence.

The meaning of the basic concepts having been established, we are now in a position to proceed with our main enquiry concerning the factors which determine the spread of efficiency within an industry and the growth of efficiency over time, and the influence on their operation of rules of the game of types 1 and 2 and of the absence of any rule of type 3.

Chapter IV

RESTATEMENT OF THE PROBLEM

I HAVE so far spoken as if the meanings of the dispersion of efficiency and of progress were self-evident. It is now time to be more precise.

In the preceding chapter I defined:

$$\epsilon = 1 - \beta(r - \bar{r}) = \frac{v + \bar{r}k}{s}$$

We may write, alternatively:

$$\epsilon = \left(\frac{v + \bar{r}k}{u}\right) \cdot \frac{1}{p_1} = a \cdot \frac{1}{p_1}$$

where p_1 is the price of any single product of any firm in the industry and $u\left(=\frac{s}{p_1}\right)$, therefore, measures the output of that or any other firm in terms of physical units of the chosen product. (In the cotton industry, for example, we might divide the turnover of each firm by the price of yarn per pound and so get a measure of the (variegated) output of each in terms of pounds of "yarn-equivalent".) a, the expression within the bracket, then has the familiar form of cost per unit of physical output,[1] and, since p_1 is a common factor, it can be ignored when we compare the efficiencies of different firms at any moment of time.

The idea which lies behind the perennial complaint that efficiency is too widely dispersed in many industries is that if the more efficient firm took over the business of the less efficient we should all be better off, since a given value of inputs would then result in a greater amount of output. More precisely, the implication is that, if this were to happen, the *average* cost per unit of physical output for the industry as a whole would be smaller. And the sort of average which we have in mind is evidently a weighted arithmetic average, with the respective outputs of different firms providing the weights; that is:

[1] The cost is, of course, more widely defined than that of which we are accustomed to think, since it includes the cost of employing capital.

$$p_1\bar{\epsilon}=\bar{a}=\frac{\Sigma au}{\Sigma u}=\frac{\Sigma v+\bar{r}\Sigma k}{\Sigma u}$$

where the summation is over all firms in the industry.

Clearly, therefore, there are two issues which we have in mind when we speak of the dispersion of efficiency; first, the extent to which the efficiencies (costs per unit of output) of the various firms differ; second, the relative sizes, in terms of output, of more and less efficient firms. We are interested, not only in what differences in efficiency there are, but also in the amounts of output which are produced at the different levels of efficiency. An industry in which the least efficient firms were a fringe of minor enterprises accounting for very little in the way of output (and absorbing correspondingly little from the stream of inputs) would cause very much less concern than one in which the relative efficiencies were the same but the least efficient produced a major part of the total output.

All information on these two issues of relative costs and relative size can be summarised in the form of a two-dimensional frequency curve of the kind illustrated by Chart I. The horizontal axis is a scale of cost per unit of output and the vertical axis a scale of quantities of output or, alternatively, proportions of total output. Against each level of costs we plot the quantity (or proportion) of output produced by the firms whose costs are at that level. The total area of the curve then represents the total output of the industry.

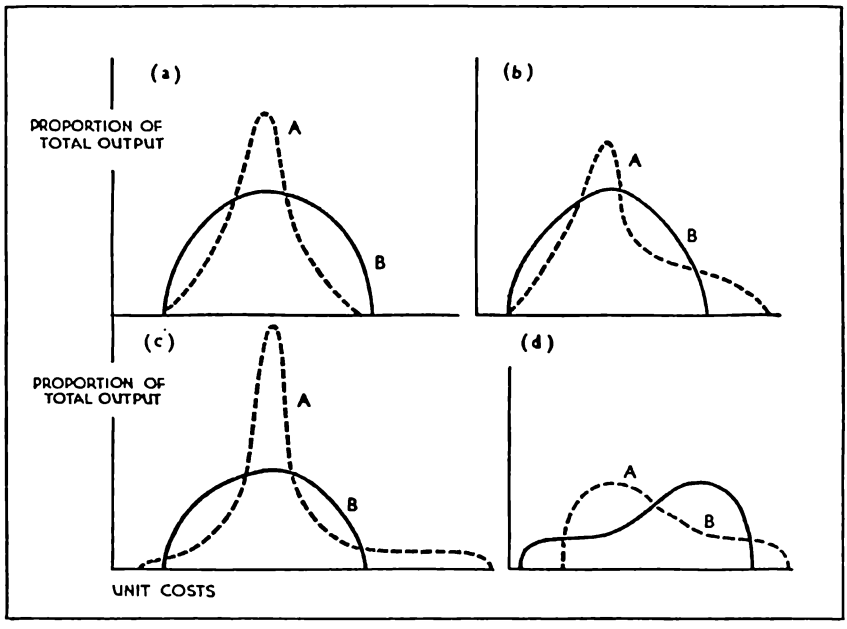
In Chart I, I have given a small selection from the infinite variety of curves which might be found in practice. What we have to decide is what characteristics of these curves it is that we take into account when we say that the spread of efficiency is greater in one industry than another. Should we say that, of the two situations represented in, say, Chart I (*a*), that yielding the more peaked curve denotes the greater spread? And if so, in virtue of what characteristics?

The development of expressions to summarise characteristics of curves is a statistical or mathematical problem. But the selection of those characteristics which are *relevant* to answering the type of question just posed is an economic problem, which can be answered only by examining more closely what it is we have

in mind when we make statements about the spread of efficiency. Moreover, if we find on examination that what we have in mind is itself not entirely precise, is tinged with vagueness, then the development of a completely satisfactory mathematical criterion for distinguishing two curves, with respect to the degree of spread which they denote, is impossible. This in itself is no reason for distress. It is in the nature of reality that some acts of choice involve more judgment than others, and the only way to deal with reality is to take it as you find it unless you know how to change it.

Chart I

HYPOTHETICAL FREQUENCY DISTRIBUTIONS OF UNIT COSTS



What we have in mind when we say that efficiency is more dispersed in one industry than another (or than it ought to be) is that the scope for improvement is greater in one than the other. The scope for such improvement depends on the scope which we (as reasonable men) think there is, either for inefficient firms to improve their efficiency, or for more efficient to take over business from less efficient firms. In seeking a measure of dispersion,

therefore, we are looking for a minimum level of efficiency which we think it reasonable that all firms should achieve and/or an idea of the scale of the transfer of business from less to more efficient firms which it is reasonable to contemplate.

Were we perfectionists, we might say that no firm in the industry should be below the standard of the best or, alternatively, that if any were, their business should be taken over by the best. But as reasonable men we should feel some hesitation in saying this if the position were (as in Chart I (*a*)) that the lowest cost firms produced only a very small proportion of the total output of the industry. Genius or superlative good luck are, by definition, confined to a few, and we should think it unrealistic, therefore, to take the standards achieved by only a few as the norm for the industry. Moreover, we should be unwilling to presume that if the firms responsible for, say, the 5% of output produced at lowest cost took over the business of all others in the industry their costs would remain unchanged. Ability to run a small business efficiently is not necessarily a good guide to ability to manage a big business.

Most of us would, I think, feel happier with the standards implied in statements like, "If only all firms in the industry were up to the (present) average in efficiency . . .". Indeed, statements of this kind are the common currency of the post-war reports on industrial efficiency. If this be accepted, then what we are referring to when we speak of the scope for improvement is something to do with the extent to which the cost of production of part of the industry's output exceeds the average.

One possibility for crystallising what we have in mind would be to take, say, the mean excess of the unit costs of the firms in the right-hand segment of our curves over average unit costs for the whole industry.[1] We should then say that, of two industries, that in which this excess was larger had the greater dispersion of efficiency. There are, however, two drawbacks to this criterion. It would lead to our saying that the two situations pictured in Chart I (*b*) are identical with respect to dispersion, whereas

[1] $\frac{\Sigma a'u'}{\Sigma u'} - \frac{\Sigma au}{\Sigma u}$, where primes indicate that only firms whose unit costs exceed the average have been included in the summation.

common sense suggests that the scope for improvement is greater in that producing the high-peaked long-tailed distribution (Industry A). In the second place, we shall not be indifferent in our practical judgments to the character of the left-hand segment of the distribution. We shall feel, for example, that an industry in which a significant proportion of output is being produced at very low costs contains more scope for improvement than does one in which most production in this segment is at costs not very far below the average for the industry, even though the averages may be the same in both cases. The criterion of dispersion should, therefore, take account of the whole distribution.

Both these difficulties can be overcome in part by the use of a very familiar statistical measure—the standard deviation. This is defined as

$$p_1\sigma_\epsilon=\sigma_a=\sqrt{\frac{\Sigma(a-\bar{a})^2u}{\Sigma u}}$$

It may also be written as:

$$\sigma_a{}^2=\frac{1}{\Sigma u}\cdot\left[\frac{\Sigma(a'-\bar{a})^2u'}{\Sigma u'}\cdot\Sigma u'+\frac{\Sigma(a''-\bar{a})^2u''}{\Sigma u''}\cdot\Sigma u''\right]$$

where primes and double primes indicate production at costs respectively higher than and lower than the average for the whole industry. In this form it can be seen to be a weighted average of the mean divergences (from the average for the whole industry) of unit costs in the two segments. The primary purpose of using squares of the divergences in forming the mean divergence is, of course, to prevent its being identically zero.[1] But this procedure has the added advantage of partially meeting the difficulty mentioned above. The process of squaring the deviations results in very high or low costs per unit being given greater weight in the sum than they would have in the normal arithmetic average first considered.[2] By this measure, therefore, industry A in Chart I (*b*) would have a greater dispersion of efficiency than would industry B, which is agreeable to common sense.

[1] The arithmetic mean is by definition that value of unit costs for which the sum of positive and negative deviations is zero.

[2] If the weights are equal, the mean of 1, 2 and 3 is 2. The square root of the mean of the sum of their squares is $\sqrt{\frac{14}{3}}$, which is approximately 2·15.

The value of the standard deviation of costs per unit of output will, of course, vary with the units in which we have chosen to measure u. $\sigma_\epsilon = \frac{\sigma_a}{p_1}$ is, of course, independent of the units in which output is measured, and may be used, therefore, for inter-industry comparisons of dispersion. $\frac{\sigma_a}{p_1}$ is termed the coefficient of variation of unit costs.

I shall assume henceforth that what we mean by saying that efficiency is more dispersed in one industry than another is that the value of σ_ϵ is greater for the first industry. It is not difficult to invent hypothetical cases where adherence to this criterion would lead to results which are plainly contrary to common sense. For example, if we were confronted by the two cases illustrated in Chart I (*c*), uncritical use of the coefficient of variation of costs would result in industry A's being regarded as that with the greater dispersion. In practice we should be loath to accept this conclusion. We should suspect rather that the high values of a, whose squaring is largely responsible for the high value of σ_a, were the result of some special factors and should wish first to investigate their causation more closely. Moreover, the two distributions in Chart I (*d*) differ in a way of which we should wish to take note, although their coefficients of variation are identical. All this is only to say, however, that since there is something eclectic in what we mean by the dispersion of efficiency it is not possible to devise any criterion which can be applied in a mechanical way, nor even to define, except at inordinate and unusable length, *precisely* what we mean by it. Tools of thought can be dangerous to the workman like any other tools. The moral is to use them with care and not at all in situations for which they are unsuitable. I shall try to keep this moral in mind.

There are difficulties also in reducing the idea of technical progress, self-evident in a general way, to more precise and measurable terms. When we say that an individual firm has made technical progress between two dates we mean that the amount of output which it produces per unit of aggregate input has increased between the two dates. Remember, however, that the

means we are using to aggregate the heterogeneous mixtures of which output and input are composed are relative prices. Clearly, we need to use the same relative prices at the two dates we are comparing. But if relative prices have changed between the two dates we then have a choice between two sets, and our measure of the change in efficiency will differ according to which set we use. Neither measure is more or less correct than the other. If the firm has changed the composition of its output during the period, so that some of the goods produced at the earlier date have been replaced by new goods at the later date, we are in an even worse state. For neither set of relative prices provides us with a means of reducing the outputs of the two dates to common units. Supposing that we have solved these problems in some way and produced measures of the changes in efficiency of individual firms, we face a difficulty of a similar kind when we try to average these changes to yield a measure of the change in the efficiency of the industry. It would be natural to weight the changes for individual firms by their outputs in striking such an average. But, if the relative outputs of firms have altered over the period we shall, again, get two different averages according to whether we use the outputs of the earlier or the later period.

Since the notion of progress is fundamentally imprecise, it is impossible to provide a definition and measure of it which is both precise and expresses all that we mean. The sets of measures most suitable for each of the different shades of meaning are set out in the literature of index-number theory, to which I have neither the desire nor the competence to add. For the purpose of this book I propose to act henceforward as if the following two measures expressed all that we mean by progress and had each a unique value.

Change in efficiency of firm i =

$$\frac{\epsilon_{i_1}}{\epsilon_{i_0}} = \frac{v_1' + \tilde{r}_0 k_1}{v_0 + r_0 k_0} \cdot \frac{s_0}{s_1'} = \lambda \epsilon_i$$

where 0 and 1 indicate two dates and primes indicate revaluation at the prices of year 0.

Change in the efficiency of the industry =

$$\lambda\bar{\epsilon}=\frac{\Sigma\lambda\epsilon_i s_{i0}}{\Sigma s_{i0}}$$

where s_{i0} is the output of a firm in period 0.

Subject to the imprecisions inherent in the concepts, I can now restate the object of this book in precise terms. It is to examine the nature of the evidence (known or knowable) for the proposition that the dispersion of efficiency σ_ϵ in any given industry at some point of time, or its progress $\lambda\bar{\epsilon}$ over any period of time, will both be affected by the sort of rules of the game under which the industry is operating. From Chapters V to X the argument is conducted in what are usually called theoretical terms; in plain words, I do not verify most of the premises from which I argue. In the remaining chapters I examine what support is lent to the conclusions of the analysis by some of the more accessible empirical material.

Chapter V

THE NATURE OF THE THEORY

BELIEF in the influence of rules of the game on dispersion and progress implies knowledge of the mechanism by which they are determined and with whose operation the rules are believed to interfere.

An industry has a continuous history, in the sense that it persists through time. Any measurement of the dispersion of efficiency refers to a snapshot from that history, and our measure of progress is of the difference between two snapshots at different points of time. When we ask questions about the mechanism through which it comes about that dispersion at some point of time is such and such and not something different, or that progress between two points of time is what it is, what we seek is an understanding of the mechanism through which the present has grown out of the past or the future will grow out of the present. The present being timeless by definition, nothing can happen in it; the happenings which have produced it are in the past. We can ask, "Given two snapshots, tell me what was the nature of the process by which the later in time developed from the earlier"; or alternatively, "Given the present situation tell me what the future will be." By asking such questions we are implying that historical continuity means more than mere succession; that there is causality also. And by causality we mean that any given historical situation will contain within itself the seeds of its own transformation into a different situation later in time. If we define "situation" widely enough, this proposition is identically true. On narrower and more convenient definitions, we are forced to distinguish between internal and external[1] causes of the transformation, between seeds of transformation and "shocks" from outside. The progress of knowledge consists in the definition of situations in such a way that the seeds of transformation present in them can be detected and the manner of their operation known.

[1] Endogenous and exogenous in the jargon of economists.

In the situation with which we are concerned there are two characteristics in which we are interested—efficiency and output. When, therefore, we ask questions about the process by which dispersion or progress has been determined we are enquiring about the nature of the forces tending to transform them, which have been generated by (were present within) particular constellations of efficiencies and outputs in the past. More precisely, we are enquiring about the process of interaction between efficiencies and outputs within an industry which leads to both being transformed.

The most generally recognised force making for such transformations is a tendency for more efficient firms to grow (increase their outputs) at the expense of less efficient firms. Indeed, I have defined an industry in the preceding chapter as a group of firms within which such transfers of business can take place. Without some such tendency it is impossible to regard a market economy, as in fact we do, as a means of selecting, with whatever degree of imperfection, the fit from the unfit. This being so, we must regard the relative outputs of today—which constitute one element of our dispersion measure—as the product, *via* what I propose to call the transfer mechanism, of the relative efficiencies of the past. If the transfer mechanism continued to be operated by an unchanged set of relative efficiencies the ultimate result could only be the concentration of the whole output of an industry in the hands of one, the most efficient, firm. The dispersion of efficiency would be eliminated by the elimination of all but one firm. The transfer mechanism would be self-destructive and the end result would be the disappearance of anything which could plausibly be described as a market economy. And, if the data presented in later chapters are any guide at all to the magnitude of the differences in efficiency which exist, the process of concentration could take place with great speed.

Monuments to the self-immolation of the transfer mechanism are, in fact, not uncommon. Some industries are very largely dominated by a single firm. There are a considerable number where what is technically described as a high degree of concentration has been reached; that is a state in which either a small number or a small proportion of the firms in an industry produce

a high proportion of the output. And it may well be true that there is an inexorable trend towards greater concentration throughout the economy. But concentration is by no means universal, and there is lively and continuing dispute about whether the average degree of concentration has or has not been increasing in the United States and the United Kingdom over the last twenty years or so. There must, therefore, be some counter-force to the transfer mechanism.

It is possible to argue that it is certain rules of the game which prevent the market economy from destroying itself, by inhibiting the working of the transfer mechanism. Indeed, one of my main purposes is to examine how far such rules perturb the mechanism. But to assume that the mechanism would quickly destroy itself without such rules is to rob ourselves of any standard for judging the rules. For we are judging the rules in the context of a market economy, and if the rules are necessary for the very existence of such an economy the distinction between the rules and the economy virtually disappears. Indeed, since separating sheep from goats is the prime function of the market mechanism, it is difficult to see that much is left if that function has to be suspended to prevent its self-destruction.

Nor is it necessary to accept this nihilistic conclusion. For it is an evident fact that relative efficiencies do not remain indefinitely the same, and because there are such changes the transfer mechanism finds itself, so to speak, with its work constantly undone to a greater or lesser degree. Today's most efficient and fastest growing firm is tomorrow supplanted by another, formerly less efficient, which now takes over the lead in growth. It is this, I suggest, which provides the main counter-force to the transfer mechanism. It will be seen later that the necessary conditions for this counter-force to be fully effective are so stringent that they are unlikely often to be satisfied. This implies that there is a long-term trend towards greater concentration. Nevertheless, if changes in relative efficiency are reasonably frequent and substantial, this is sufficient to postpone the nightmare of universal high concentration to a fairly remote future. And this is all we need for present purposes.

These changes in relative efficiency can be regarded as the

product of chance, without destroying their effectiveness as a counter-force to the transfer mechanism. But it is natural first to seek a force which provokes them. And common sense suggests that this is the transfer mechanism itself. This is the mechanism which threatens the inefficient firm with destruction, and I suggest that it is the efforts of such firms to avoid destruction which result in changes in the constellation of efficiencies. This implies that relative efficiencies change through improvements in the less efficient rather than deterioration of the better firms, and this, in turn, means that the average efficiency of the industry will increase. For this reason, I shall refer to the process whereby relative efficiencies are changed as the innovation mechanism.

To sum up, I suggest that there are two fundamental transformation mechanisms which are necessary to and characteristic of a market economy—the transfer and innovation mechanisms. It is their interaction which determines the dispersion of efficiency observable at any moment of time and the change in efficiency over any period of time. The answers to questions about the virtue or vice of particular rules of the game depend on the way in which they influence these interactions.

The immediate task defines itself as a closer analysis of the nature of these two transformation mechanisms. This occupies the next two chapters. I have found it necessary to cut down a considerable number of trees in order to see the wood, and many of the complexities of reality are assumed away. In particular, I assume that the industry examined in Chapters VI and VII maintains with other industries only a buyer-seller relationship. The consequences of relaxing this assumption are examined in Chapter VIII.

Chapter VI

THE TRANSFER MECHANISM

FOR a firm to grow, it must have both the will and the means. Hence, the proposition that efficient firms will grow faster than, or at the expense of, inefficient, implies that the former have a greater desire and/or more ample means to grow.

Motives and Means for Growth

It is by no means *necessary* that the management of a firm should wish it to grow in size indefinitely. Growth involves thought, effort and worry, and there is ample historical evidence that willingness to undertake these is by no means an immutable "instinctive" characteristic of human nature. Empire-builders are made, not born. It is perfectly possible to conceive of a situation in which those coming into ownership of capital assets regarded them as a fund available to finance consumption during their lifetime. Indeed, the fabrication of such nightmares to frighten tax-inspectors has been a popular sport of the post-war period. It is not merely possible but even plausible to imagine that firms were content to achieve a certain size and, thereafter, were concerned only to maintain that position. That certain rules of the game tend to produce this situation is one of the allegations we shall need to examine.

I shall for the moment, however, assume that all firms in an industry are consumed with an equal desire to grow without limit. This assumption can be justified on three grounds. The first is that, if you suspect or fear that other firms are trying to grow, the only way to be sure of not shrinking below what you would otherwise regard as the right size is to try to grow also. Surrender of further ambition once the right size is reached means that any subsequent change can only be in the downward direction. To surpass one's ambitions is less painful than to fall short of them, and the uncertainty of the future makes it doubtful whether a short-fall will subsequently be made good so easily as a present excess can be achieved. Even one who is totally

free from megalomania will, therefore, continue to press on when the going is good.

The second point is that great size brings security in a manner additional to that just discussed. The bankruptcy of a small firm will, like the fall of the sparrow, in fact go unnoticed, except by the widows and orphans dependent on it and the creditors who have brought in the bailiffs. The death of a large firm, however, is so shocking an event as to be almost impossible. Someone, be it bank, insurance company, erstwhile rival, or government, will always be found to set it on its feet, and though the former owners or controllers may well be displaced in the process, they are likely to receive some financial consolation for their deprivation. This relatively invulnerable position must be attractive to even the most slothful and security-minded of business-men, although its force diminishes when once the invulnerable size is reached.

Both the preceding points apply irrespective of the form of business organisation. The third acquires its peculiar force from the divorce between ownership and control of firms which has resulted from the development of the joint-stock system. Outside the rather special enclaves of politics, the civil service and the learned professions, general prestige and influence are closely dependent on business standing, and the business standing of an individual depends in turn on the volume of assets which he controls and the amount by which they have grown under his stewardship.[1] Since the desire for high standing has considerable claims to be considered an instinct, this situation creates a powerful incentive for those who administer firms to set their sights ever higher.

If we are to regard the desire to grow as common to all firms in an industry, the explanation of different rates of growth must lie in differences in command over the means to growth. The means to growth are customers and capacity.

I have defined capital employed in Chapter III in an unambiguous way. The concept of capacity is closely related but is much less precise. We are accustomed to think of any particular volume of output (its composition and the technique of production

[1] For a penetrating and entertaining analysis of this aspect of the sociology of modern capitalism see *The Great Crash* by J. K. Galbraith.

being given) as requiring a minimum amount of productive equipment and as involving a definite amount of stocks and work in progress. Since equipment and stocks and work in progress constitute capital employed, we imply, therefore, a unique relationship between output and minimum capital employed. Putting the matter the other way round, we say that a particular (maximum) level of output is possible with a particular value of capital employed; this is the capacity of the plant whose cost is measured by our figure of capital employed. But we know very well that, in practice, the amount of output which can be obtained from a plant of given size depends on how it is worked. Fixed equipment will produce more per week with multiple shifts, and work in progress will not rise proportionately to output. So that in speaking of *the* capacity we are making a whole set of implicit assumptions as to a particular manner of working. When we speak of a firm working at 120% of capacity we mean that a different manner of working is being followed from that assumed in our definition of capacity.

There is, therefore, a subjective element in the notion of capacity. There is no unique physical relationship between output and capital employed. But in speaking as if there were we are following the practice of business-men. It is they who determine methods of working and they, as their speech reveals, have ideas (which may change over time) as to "normal" methods of work. "Normal" conditions are, by definition, those which business-men will seek to establish or maintain. Hence, if our interest is in trends and not in short-term fluctuations, we are justified in conducting our analysis as if capacity had a unique and unalterable value.

Growth, therefore, requires more customers and more capital, and I shall assume, for the moment, that a given growth in output requires a definite increase in capital employed.

A firm can get more capital (money) by earning it or borrowing[1] it. It earns money by making profits and not paying them out to shareholders. If growth were financed wholly from earnings, therefore, the firm could not grow (increase its capital employed) at a rate greater than the rate of (retained) profit on capital

[1] I use this term to cover all sources of outside finance, including share issues.

employed.[1] Borrowing makes a greater rate possible. But there is a fairly definite ceiling to the amount which a firm can (or would wish to) borrow during any period. Detailed analysis of the factors which determine this ceiling would require a book in itself. But the general nature of the determination is simple enough. The potential lender will be guided in his judgment of the probable success of the expansion he is being invited to help finance by the evidence of the firm's past success provided by its rate of profit. And the bigger the proportionate expansion of capacity which the firm is proposing to finance by borrowing the greater the past success which will be needed to persuade him. Large proportionate expansions will be more suspect than small, first, because they reduce the cover for the loan more and, second, because success at one size is not indubitable evidence of ability to succeed at another; we like our debtors to prove themselves at each level before going on to the next.

There will be some rate of profit below which there is no practical possibility of a firm's borrowing. Above this level, borrowing becomes possible, and the greater the rate of profit earned the more can be borrowed. But the scope for securing outside capital will certainly not increase in proportion to the rate of profit, and there will be some ceiling in any period to the amount which a firm can secure. As a first approximation, therefore, we can regard the potential rate of growth of capacity as related in a definite if complex fashion to the rate of profit.

But the rate of profit which a firm earns will, given its costs and capital employed, depend on the prices which it charges. And the rate at which its market will be growing and, therefore, the rate at which it wants to add to capacity will also depend on its prices. Both the means to finance new capacity and the market to employ it will, therefore, vary, in opposite directions, with the prices it charges. And there will be some level of prices at which the rates of growth of capacity and the market are equal. If the firm is not producing to capacity but is still earning the means to further growth, the appropriate corrective action will be to find more customers by offering more attractive terms than it

[1] Henceforward the rate of profit unqualified is to be taken to mean the rate of net profit on capital employed.

believes some other producers are willing or able to. Conversely, if demand persistently exceeds capacity the appropriate action is to raise the rate of growth of capacity which it can finance, by earning more money, by raising prices. I shall call the level of prices at which the firm's capacity and market are equal and increasing in step the equilibrium level; the equilibrium price for short.

What this equilibrium price is for any particular firm will clearly depend, however, on the prices which other firms in the industry are charging, on the general level of prices in the industry. We must now consider how this is determined.

The General Level of Prices and Profits

An industry is simply a collection of firms. Hence, if we exclude for the moment the possibility of firms entering or leaving the industry and assume that each firm is in one industry only,[1] what is true of the individual firm will be true of the industry also. On these assumptions, there will be some definite and fairly stable relationship between the rates of profit and investment for an industry as there is for a firm.

The rate of increase in capacity which is required to prevent either excess or deficient capacity arising in the industry (to keep it in equilibrium[2] as I shall term it) will depend on the rate of growth of demand for the products of the industry. If we abstract for the moment from changes in technique,[3] the rate of investment in the industry will depend only on the rate at which capacity is being added to. The rate of investment which keeps the industry in equilibrium (equilibrium rate of investment) will depend, therefore, on the rate of growth of demand for the products of the industry.

If this rate of growth of demand itself depended on the prices of the industry's products (and so on its rate of profit) we should be forced into another cycle of the regress which drove us from the prices of the firm to those of the industry, and there would

[1] The consequence of relaxing these assumptions is considered in Chapter VIII and their validity in Chapter XI.

[2] Equilibrium is henceforth to be understood as I define it in this chapter, and is not to be confused with the senses of equilibrium discussed in Chapter II.

[3] The complications introduced by dropping the assumption that all investment is for adding to capacity with an unchanging technique are examined in Chapter VII.

be no point of rest in the analysis until it embraced the whole economy.[1] But we can find salvation at this point, I believe, in the doctrine of *de minimis*.

What we have to consider is whether a small (absolute) change in the average rate of profit in an industry (say from 7% to 10%) would, through raising prices, have a significant effect on the rate of growth of demand for its products. The first point to note is that profits represent a rather small proportion of the ex-factory price of most products. As I pointed out in Chapter II, the average percentage of gross profit to turnover in the United Kingdom is typically of the order of 10%. Even a doubling of the rate of gross profit, which probably implies trebling the rate of net profit, would therefore add only 10% to prices. Since the order of the variations in profit rates which we need to consider will normally be very much smaller than this, it is clear that the price changes under consideration will be small also. Hence, it is only if the sensitivity of the rate of growth of demand to price changes is enormous that we shall fall into sensible error by thinking as if the rate of growth of demand for the industry were independent of its rate of profit. And there is no reason to suppose that the average sensitivity of demand for the whole (variegated) output of an industry will in fact be enormous. Indeed, I have given reasons in Chapter II for believing that it will usually be rather small.

I conclude, in short, that the rate of growth of demand for the products of an industry can be regarded as depending on the rate of growth of the national income (on which the rate of growth of one industry exercises only a small influence) and as being little affected by the sort of variations in the industry's prices which we need to contemplate.[2] At any time, therefore,

[1] Indeed, for a fully open economy, we should be driven to include the whole world.

[2] An exception has, I think, to be made for "new" industries. They will typically begin by selling a new product to a rather restricted ("luxury") market at what, from the point of view of later history, is a very high price, and large profits will be earned. These profits (together with money and firms attracted from other industries) will finance rapid expansions of capacity and improvements in productive techniques. Falling prices and expansion of the market will go hand in hand, until such time as the product is being supplied to a mass market. After this point the industry pursues the more sedate course assumed as general in this chapter. So many of the assumptions I have made are falsified for a new industry that the solution is to produce a special theory for such cases rather than to over-complicate the "normal" theory to accommodate it.

the general level of prices which will keep the industry in equilibrium will depend, the average of costs in the industry being given, on the rate of growth of demand, the value of the capital output ratio and the relation between the rate of profit and the ability to borrow on the market.[1] The equilibrium price for any individual firm is, therefore, determinate, and so its rate of growth. We must now examine this determination more carefully.

The Mechanism at Work

The essential characteristics of the transfer mechanism can be seen most clearly by exploring some highly simplified examples. I suppose an industry containing two firms only, which are initially of the same size as measured by output. Each produces the same single product and sells it at the same price. Neither distributes any of its profits to its shareholders, but invests in additional capacity any profit which the tax gatherer leaves it, technique being unchanging. No capital is received from outside the industry. I assume that the industry is initially in equilibrium (capacity is equal to demand) and that equilibrium is continuously maintained. This requires, *inter alia*, that it should be possible to increase or diminish capacity by quite small amounts and, further, that a firm which has an operating loss (negative net profit) should respond by selling some of its capacity to enable it to continue in business on a reduced scale. The important consequences of dropping these last assumptions are examined

[1] Assume for simplicity that the industry produces only one product and that all firms sell it at the same price. Let z, k, i, u stand respectively for net profit, capital employed, net investment and output.

For any firm, $i=f(z)$, where f expresses the relation between profit earning and borrowing ability.

For the industry, $\sum i=f(\sum z)$, if f is identical for all firms.

Also $\sum i=\alpha \sum k$, where α is the rate of growth of demand for the industry's product.

So $$\alpha\frac{\sum k}{\sum u}=f\left(\frac{\sum z}{\sum u}\right)$$

or $\alpha\bar{\gamma}=f(\bar{z})$, where $\bar{\gamma}$ is average capital employed per unit of output and $\bar{z}$ is the average profit margin per unit of output.

If f has the form λ, then

$$\bar{z}=\frac{\alpha\bar{\gamma}}{\lambda}$$

and the product price is

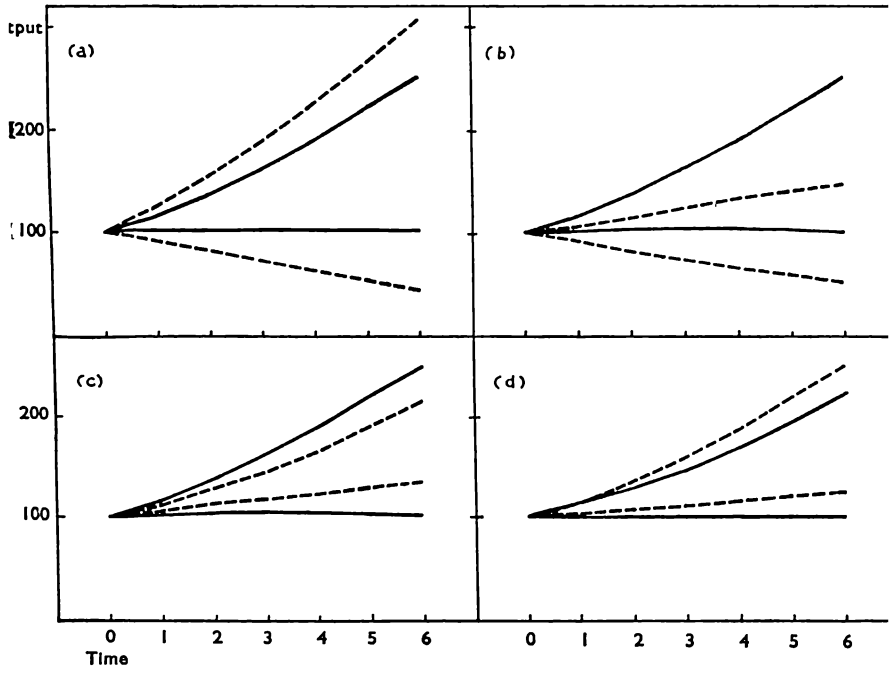
$$p=\frac{\sum v}{\sum u}+\frac{\alpha\bar{\gamma}}{\lambda}$$

in Chapter IX. I assume, finally, that one firm has lower cost per unit of output or turnover $\left(\frac{v}{s}=\frac{v}{up}\right)$ than the other; firm A' costs are 10*s*. per unit and firm B's are 9*s*. This assumption i common to all the cases considered, and costs per unit are take to be constant over time.

Chart II shows how the industry will develop in various cir cumstances.

Chart II

RATES OF GROWTH UNDER ALTERNATIVE CONDITIONS



In the first two examples [(*a*) and (*b*)], there is no taxatio of profits. It follows, then, from the assumptions above and th reasoning of the previous section, that the price ruling at an time will be given by: $p=\bar{c}+\alpha\gamma$, where c is cost per unit of out put $\left(\frac{v}{u}\right)$, γ is capital employed per unit of output $\left(\frac{k}{u}\right)$, α is th percentage rate at which demand for the industry's product i

increasing, and a bar above a symbol indicates an average, weighted by output, for the industry.[1]

Whatever be the price, B will make greater profits (or smaller losses) per unit of output than will A. Unless there be very large differences in the amounts of capital which they employ per unit of output, it will also earn a higher *rate* of profit. On the assumptions of the model, it will, therefore, grow faster than A. Since the cost component of the price equation is a *weighted* average of the unit costs of the two firms, this will fall over time, because of the increasing weight of the lower cost firm. Hence, unless α grows steadily (the rate of increase of demand accelerates), the price also will fall steadily. If both firms are growing, but B is growing the faster, a fall in price will reduce both their rates of growth. But, since the fall in price, which is the same for each, represents a smaller proportion of the (larger) profit margin of B, the rate of growth of B will fall by less than that of A. If A be shrinking (making a loss) a fall in price will, for the same reason, normally accelerate its rate of decline by more than it reduces the rate of growth of B.

Chart II (*a*) illustrates the effect of differing degrees of capital intensity. The unbroken lines (Case 1) show the respective growths of A and B when total demand for the industry's product is growing at the rate of 10% a year and the value of capital employed per unit of output is 6*s*. for each firm. The broken lines (Case 2) show what the growths would be if capital employed were only 3*s*. per unit.

Variations in capital intensity are evidently very important. It can be seen from the price equation above that, because of the smaller value of capital employed, the price is lower throughout for Case 2 than for Case 1. But the reduction in the profit margin is proportionately greater for A than for B, and is, in any case, proportionately smaller than the reduction in capital employed. Hence B increases much more rapidly than in Case 1, and A, which formerly just hung on at the margin of profitability, now makes losses and shrinks from the start. By year 6, B is, in fact, nearly seven times the size of A as compared with only 2½ times in Case 1.[2]

[1] See footnote on p. 69.

[2] In Case 1 price falls from 10·1 shillings in year 0 to 9·88 in year 6, and in Case 2 from 9·8 to 9·42.

The most striking feature in both cases is, however, the tremendous power of the mechanism. Developments have been traced over only six years. But, even in Case 1, the face of the industry has changed out of all recognition. Had we taken higher figures for γ, the changes would obviously have been smaller. But there is no reason to think that the values taken are abnormally low. Most investigations have suggested that the ratio of fixed capital employed to *net* output in manufacturing industry is of the order of unity.[1] But the ratio of net output to gross output is typically of the order of 40%, and is often as low as one-fifth. Hence the ratio of fixed capital to turnover must usually be considerably less than one-half, and, since stocks and work in progress are small in relation to turnover in most manufacturing industries, the ratio of total capital employed to turnover also. We must accept, therefore, that, in the isolated conditions we are examining, the transfer mechanism is a very potent force.

It is all the more so if demand for the industry's products is not growing very fast. This is illustrated by Chart II (*b*). Case 1 (the unbroken lines) is as in the first example, with $\alpha = 10\%$ and $\gamma = 6s$. Case 3 (broken lines) has γ also equal to 6*s*., but there is now no increase in the total output of the industry ($\alpha = 0$). Price is now constantly equal to the average costs of the industry ($\alpha\bar{\gamma} = 0$), and the rate of growth of B is thereby much reduced. But A now makes losses from the start and shrinks fast, so that the *relative* growth of B, on which our interest centres, is now greater. By the end of the period B is 2·8 times the size of A compared with only 2·5 times in Case 1, and A is heading fast for extinction. Evidently, under the conditions postulated, a rapidly growing industry will nourish high cost firms much longer than will one where demand is stagnant.

Case 4, in Chart II (*c*), shows that a similar protective mantle can be provided by taxation of profits. Case 1 is here compared with one in which all else is identical but net profits are now being taxed at a flat rate of 50%. Under the assumptions of the model, the price has now to rise sufficiently to accommodate the

[1] See, for example, T. Barna, *The Replacement Costs of Fixed Assets in British Manufacturing Industry in 1955*, Journal of the Royal Statistical Society, Vol. 120, Part I, 1957. Net output is equal to gross output (turnover) *minus* current expenditure on materials, components, fuel, etc. See Chapter XIII below.

tax, to provide sufficient profits to cover, not only the industry's need for capital, but also the tax which the Government demands. The imposition of the 50% tax is, in fact, precisely equivalent in its effect to a rise in the rate of growth of the industry from 10% to 20%. The high cost firm is now able not merely to hang on at the margin of profitability but actually to grow, while the rate of growth of B is correspondingly reduced. It is a far cry from the primitive simplicity of the model to the complexity of the real world, and the extent to which taxes are passed on in prices is a subject of perennial controversy. I have little doubt, however, that the damping effect of taxation which I have illustrated is an important reason why the transfer mechanism in reality operates in a less brutal fashion than the earlier examples suggest.

In each of the preceding examples the two firms have employed identical amounts of capital per unit of output. Chart II (*d*) illustrates the effect of relaxing this assumption. Case 1 (unbroken lines) is as before, with demand growing at 10% per year, γ equal to 6*s*. for each firm, and no taxation of profits. But in Case 5 (broken lines) the higher cost firm employs only 4*s*. of capital per unit of output, while B employs 8*s*.

The effect is to free A for some considerable time from the spectre of bankruptcy and to check substantially the growth of B. The average capital intensity of the industry is the same in both cases in the initial year. But in Case 5 B now needs to spend more to achieve any given increase in output than it did before. Since its rate of growth is still the higher, the average amount of capital required to provide any given increase in the output of the industry is greater than in Case 1. This means a higher average profit margin and price, and it is this which gives protection to A.

It is surprising, nevertheless, that the effect should be so small. The difference in capital intensity assumed is very large compared with the difference in unit costs—10% in costs and 100% in capital employed per unit. Even so, B has succeeded in growing so much faster than A that it is 75% larger by the end of this short period, and A, although still growing in year 6, will not go on doing so for many more years.

Indeed, the most general conclusion from all the cases examined is that the transfer mechanism retains great power, whatever the circumstances. So long as profits are the main determinant of potential growth the system will always show a strong intolerance of cost differences. The possibility of borrowing for expansion has not been explored, since there is too great a variety of alternative assumptions to choose from. But it is evident on general grounds that its introduction would tend to increase the power of the mechanism. If the industry as a whole is taking in, by borrowing or the sale of new shares, more money than it pays out in interest and dividends, as is likely if demand is growing very fast, this reduces the average profit margin which is needed. And, as has been seen, the lower the average profit margin the more swiftly does the transfer mechanism work. Moreover, it will normally be the most prosperous firms who are able to borrow most, and this adds still further to the power of the mechanism. We may conclude, therefore, that strong counter forces are needed if the transfer mechanism is not to rush industries incontinently towards high degrees of concentration.

The Transfer Mechanism and the Dispersion of ϵ

The standard deviation of two variables is a maximum when their weights are equal. It is obvious, therefore, that the effect of the transfer mechanism in the examples considered is to reduce the dispersion of costs. We cannot, however, argue directly from this to the dispersion of efficiency as we have defined it.

In the first place, it does not necessarily follow that, of two firms, the more efficient will always have the lower costs and higher rate of profit. This was argued in Chapter III. But it is the firm with the higher rate of profit which grows more quickly under our assumptions. There is a possibility, therefore, that the more efficient firm will grow at a slower rate than the less efficient. This may be accounted an inherent defect either of the market mechanism or of the definition of efficiency I have adopted. For my own part, I see no justification in logic for amending the definition. Fortunately, somewhat unusual values of the variables are needed to make the anomaly important in practice.

It will further be recalled from Chapter III that the efficiency

index is influenced by changes in prices and the average rate of profit, and, in particular, that falling prices will increase dispersion in the special case where capital output ratios are identical in all firms. This equality is present in all but the last of our examples. There are, therefore, two opposing forces at work. The changes in relative size which the transfer mechanism brings about tend to reduce the dispersion, while the falling price which it also produces tends to increase the dispersion.

Table 1

VALUES OF ϵ AND σ_ϵ

		Year 0		*Year 6*	
Case	*Firm*	ϵ	σ_ϵ	ϵ	σ_ϵ
1	A	1·0495	·0495	1·0731	·0456
	B	·9505		·9722	
2	A	1·0510	·0510	1·0930	·0353
	B	·9490		·9871	
3	A	1·0527	·0527	1·0799	·0477
	B	·9473		·9719	
4	A	1·0467	·0467	1·0641	·0466
	B	·9533		·9689	
5	A	1·0262	·0262	1·0342	·0259
	B	·9738		·9804	

It will be seen from Table 1 that the first influence is in all cases the more powerful. While the index for firm A is in each case higher in the closing than in the opening year, that of B moves closer to unity and the change in weights results in a fall in the average dispersion.

The results in the examples are in accordance with intuitive expectations; it is congenial to common sense that a mechanism which favours the efficient and takes from the inefficient should progressively reduce the dispersion of efficiency. The effects of differences in the rate of growth of demand, in capital intensity, and in the level of taxation are also in line with *prima facie* expectations. Since general proofs of the validity of these propositions are difficult to provide, most readers may be content to let the issue rest at this. But, for those who can stomach some clumsy,

though basically simple, algebra, the working of the mechanism can be analysed for an industry with more than two firms. Those to whom the next section is unattractive will not find their subsequent understanding impaired by not reading it.

A More General Model

The basic assumptions are the same as in the preceding section. The composition of output for each firm does not change over time and *relative* prices for different products are unchanging. Hence, the meanings of a change in the volume of output for the firm or the industry or a change in price are unambiguous. Costs of production $\left(\frac{v}{u}=c\right)$ are constant over time for any firm. All net profits are invested in additional capacity, capital per unit of output $\left(\frac{k}{u}=\gamma\right)$ being identical for all firms and constant over time. There is no taxation, no inflow of capital from outside the industry, and no movement of firms into or out of the industry. The industry is continuously in equilibrium.

In Chapter IV, I defined

$$p_x\epsilon=a=\frac{v}{u}+\bar{r}\frac{k}{u}=c+\bar{r}\gamma$$

where p is the price of any product (x) taken at random and u is output in "x-equivalent". Evidently, when u is used for weighting, as is appropriate,

$$p_x\bar{\epsilon}=\bar{c}+\bar{r}\bar{\gamma}$$

The standard deviation of ϵ is formed from the squares of deviations from the mean. But γ is identical for all firms in the industry.

Hence

$$p_x(\epsilon-\bar{\epsilon})=(c-\bar{c}),$$

and $p_x\sigma_\epsilon=\sigma_c$

Let us now write p and σ without subscript to mean p_x and σ_c, and denote different periods by numerical subscripts.

Since $\bar{\epsilon}$ is always unity,

$$p=\bar{c}+\bar{r}\bar{\gamma}$$

And, in the conditions postulated above, $\bar{r}=\alpha$, where α is the rate of growth of demand for the products of the industry.

For any firm, by assumption,

$$k_1=k_o(1+r_o),$$

or $$u_1=u_o(1+r_o)$$

$$r_o=\frac{p_o-c}{\gamma}=\alpha-\frac{c-\bar{c}_o}{\gamma}$$

So $$u_1=u_o(1+\alpha)-u_o\frac{(c-\bar{c}_o)}{\gamma}.$$

$$\Sigma u_1=\Sigma u_o(1+\alpha)$$

since γ is a constant and $\Sigma(c-\bar{c}_o)u_o=0$.

$$cu_1=cu_o(1+\alpha)-cu_o\frac{(c-\bar{c}_o)}{\gamma}$$

$$=cu_o(1+\alpha)-u_o\frac{(c-\bar{c}_o)^2}{\gamma}+c_ou_o\frac{(c-\bar{c}_o)}{\gamma}$$

So $$\bar{c}_1=\frac{\Sigma cu_1}{\Sigma u_1}=\frac{(1+\alpha)\Sigma cu_o}{(1+\alpha)\Sigma u_o}-\frac{\Sigma(c-\bar{c}_o)^2u_o+\bar{c}_o\Sigma(c-\bar{c}_o)u_o}{\gamma(1+\alpha)\Sigma u_o}$$

$$=\bar{c}_o-\frac{{\sigma_o}^2}{\gamma(1+\alpha)}$$

And $$p_1=\bar{c}_o-\frac{{\sigma_o}^2}{\gamma(1+\alpha)}+\alpha\gamma=p_o+\frac{{\sigma_o}^2}{\gamma(1+\alpha)}$$

$${\sigma_1}^2=\frac{\Sigma(c-\bar{c}_1)^2u_1}{\Sigma u_1}$$

Writing $\delta=\frac{{\sigma_o}^2}{\gamma(1+\alpha)}$, and substituting for $\bar{c}_1$, gives:

$${\sigma_1}^2=\frac{\Sigma(c-\bar{c}_o+\delta)^2u_1}{\Sigma u_1}$$

$$=\frac{\Sigma[(c-\bar{c}_o)^2+2\delta(c-\bar{c}_o)+\delta^2]u_o}{\Sigma u_o}-\frac{1}{\gamma(1+\alpha)}\cdot\frac{\Sigma[(c-\bar{c}_o)^2+2\delta(c-\bar{c}_o)+\delta^2](c-\bar{c}_o)u_o}{\Sigma u_o}$$

Bearing in mind that
$\Sigma(c-\bar{c}_o)u_o=0$
by definition,

$$\sigma_1^2=\sigma_o^2+\delta^2-\frac{1}{\gamma(1+\alpha)}\cdot\frac{\Sigma(c-\bar{c}_o)^3u_o}{\Sigma u_o}-\frac{2\delta}{\gamma(1+\alpha)}\cdot\frac{\Sigma(c-\bar{c}_o)^2u_o}{\Sigma u_o}.$$

Writing $\mu_{3\cdot o}$ for $\frac{\Sigma(c-\bar{c}_o)^3u_o}{\Sigma u_o}$, and remembering that

$$\frac{\Sigma(c-\bar{c}_o)^2u_o}{\Sigma u_o}\cdot\frac{1}{\gamma(1+\alpha)}=\delta, \text{ we can write this as,}$$

$$\sigma_1^2=\sigma_o^2-\delta^2-\frac{1}{\gamma(1+\alpha)}\cdot\mu_{3\cdot o}$$

Substituting back for δ gives the final expression.

$$\sigma_1^2=\sigma_o^2\left[1-\frac{\sigma_o^2}{\gamma^2(1+\alpha)^2}\right]-\frac{1}{\gamma(1+\alpha)}\cdot\mu_{3\cdot o}$$

Hence the standard deviation:

$$\sigma_1=\sigma_o\sqrt{1-\frac{\sigma_o^2}{\gamma^2(1+\alpha)^2}-\frac{1}{\gamma(1+\alpha)}\cdot\frac{\mu_{3\cdot o}}{\sigma_o^2}}$$

I have shown that $p_1=p_o\left[1+\frac{\sigma_o^2}{p_o\gamma(1+\alpha)}\right]$

So

$$\frac{\sigma_1}{p_1}=\frac{\sigma_o}{p_o}\cdot\frac{\sqrt{1-\frac{\sigma_o^2}{\gamma^2(1+\alpha)^2}-\frac{1}{\gamma(1+\alpha)}\cdot\frac{\mu_{3\cdot o}}{\sigma_o^2}}}{1+\frac{\sigma_o^2}{p_o\gamma(1+\alpha)}}$$

$$=\frac{\sigma_o}{p_o}\cdot\frac{\sqrt{\gamma^2(1+\alpha)^2-\sigma_o^2-\gamma(1+\alpha)\frac{\mu_{3\cdot o}}{\sigma_o^2}}}{\gamma(1+\alpha)+\frac{\sigma_o^2}{p_o}}$$

Since $\sigma_\epsilon=\frac{\sigma}{p}$, this last expression gives the dispersion of ϵ in year 1 in terms of its dispersion in year 0.

The meaning of these equations is consistent with the conclusions drawn from the analysis of the two-firm industry. Other

things being equal, output prices will fall period by period so long as there is any dispersion of efficiency in the industry. With a given dispersion, the rate at which prices fall will be the greater the smaller the capital intensity of the industry and the rate at which demand for its products is growing. The dispersion of efficiency $\left(\sigma_\epsilon=\frac{\sigma}{p}\right)$ will also shrink over time in normal circumstances, and the rate at which it shrinks will again be greater the lower are α and γ.

The expression $\mu_{3\cdot 0}$ is known as the third moment of the distribution about the mean. It indicates, roughly speaking, the extent to which the distribution of costs is not symmetrical about its mean. It can be positive or negative. It is obvious from the signs of the equation that dispersion will shrink more rapidly if $\mu_{3\cdot 0}$ has a positive value than if it is zero or negative. Reference to Chart I will make clear the common sense of this. Curve A in Chart I (*c*) has a positive value for $\mu_{3\cdot 0}$ (is positively skewed) and curve A in Chart I (*d*) has a negative value. The concentrated block of low-cost producers in Chart I (*c*) would have no difficulty in rapidly gobbling up the tail of high-cost firms. But the relatively thin tail of efficient firms in Chart I (*d*) would find the fat body of the less efficient a more formidable mouthful, which could be digested only slowly. We may note in passing that a positively skewed distribution like that in Chart I (*c*) represents an inherently highly unstable situation, and its persistence in any practical case will raise serious doubts as to whether the transfer mechanism is operating as it should.

Finally, it is obvious from inspection of the equations that prices and dispersion will fall by greater amounts over any given period the greater is dispersion at the beginning of the period. The system reacts more violently against large dispersions than against small. This means that as time goes on the power of the transfer mechanism diminishes; the more of its job has been done the more slowly does it do what remains.

Summary

In this chapter I have examined the forces which dispersed efficiency in an industry will provoke and the way in which they

will tend to remove the provocation. The examination has suggested that these forces—the transfer mechanism—are exceedingly powerful, and capable of eliminating dispersion quite quickly, through the absorption by the most efficient firms of most of the business of the industry. Efficiency differences, in the circumstances assumed, would quickly lead to a high degree of concentration in the industry.

Various factors have been discovered which will weaken the power of the mechanism and slow down the pace of concentration; in particular, capital-intensive methods of production, high levels of taxation of profits, and a rapid rate of growth in the demand for the products of the industry.[1] But their effects are only moderate. If the tendency to concentration is to be checked to the degree which it appears to be in reality, the counter-force of the innovation mechanism must be very strong. This is examined in the next chapter.

[1] This last factor is admissible only if it is right to assume, as in this chapter, that the average rate of profit in an industry will be the higher the greater is its rate of growth. This in turn requires that borrowing should be small in relation to total capital expenditure and that the movement of firms between industries should not be substantial. The latter question is examined in Chapter VIII. The former has *prima facie* plausibility in view of the well-known fact that a very high proportion of investment in most industries is financed from retained profits. Full justification would require a separate study, but the observed differences in average rates of profit in different industries are consistent with the conclusion that borrowing operates to bring about inter-industry equality of profit rates in only a very qualified way.

Chapter VII

EFFICIENCY DIFFERENCES AND THE INNOVATION MECHANISM

I CONCLUDED in the previous chapter that efficiency differences can persist (and concentration be avoided) only if today's most efficient firm is not the best tomorrow. So long as the transfer mechanism is operating, persistent differences in rank (with respect to efficiency) imply that the ranks of individual firms should change over time. In this chapter I consider why it is reasonable to expect such changes to occur.

Let us first consider why the efficiencies of different firms within an industry should differ at all. If, begging certain questions, we assume that all firms have equal access to the market in which productive factors are hired, we can consider them as differing, not in the quality of the factors they hire, but in the way in which they organise labour, capital and bought-out materials and fuel into a productive whole. We can then speak indifferently of the efficiency of a firm or the efficiency of its management, and our question is why some managers should be less efficient than others.

The task of management is not fundamentally different from that of leadership in any social activity, and the two-fold classification of function which is applied in appraising the efficiency of, say, a military commander will fit the managerial case very satisfactorily also. We must distinguish, in fact, between what the management is trying to do—its objective—and the method it chooses to achieve that objective—its technique. For completeness, we should take account also of the effectiveness with which the method is applied. But this is a second order question by comparison with the first two, and I shall not have much to say about it.

The objective is to produce a particular set of products in certain proportions, and the method is a technique of production, which can be specified in terms of the proportions and types of

men, machines and materials and the way in which these are organised together. How we choose to define a technique of production is, of course, an open question. We may, if we like, be very nice in our distinctions and choose to say that the technique has altered if there is the slightest change in the proportions, types or manner of organisation of the different factors. But it is both convenient and consonant with familiar usage to define techniques in such a way as to give them a certain amount of endurance. There are, for example, certain differences between hand-mills and continuous strip-mills in the steel industry which survive the continuous stream of smaller variations to which each process is subject, and we need some way of describing these abiding distinctions. I propose to use technique of production to meet this need.

This usage requires, in turn, that we take more account of time in our definition of the firm's objective. We are considering a technique as something which will endure for a certain period of time. And it is only in exceptional cases that the particular set of products which the firm will produce will be invariant, in both composition and proportions, over that period. If, therefore, we are to continue to think of managements as choosing particular techniques to achieve defined objectives, we must define those objectives in such a way as to cover this liability to change. We need, in fact, to define the objective of the firm in terms of both the composition and the flexibility of output.

What I have termed the production objective of the firm can be described more familiarly as the market for which it is going. One firm, for example, may choose to concentrate on producing long runs of a limited variety of standardised products; this is mass production. Another may think it better to produce a wide variety of products and change from one to the other at frequent intervals; such is the jobbing firm. Whatever the market which a firm chooses to go for, there will be one method of production —one technique—which is most appropriate to it. This is so because of certain fundamental characteristics of technology which must briefly be examined.

Improvements in productive efficiency can come about in only two ways; the first is from the discovery that certain operations

previously performed are in fact unnecessary, that is, more generally, by preventing energy from being wasted; the second is by the substitution of non-human for human energy in the productive process.[1] The second, which rests on the simple fact that a worker can release far more stored-up energy in the form of, say, coal than he consumes in digging it, has been by far the more important source of progress. Such non-human energy can be brought to bear in a production process, however, only through the agency of a machine (whose construction, of course, absorbs energy). Hence it is equally and identically true to say that the growth of productive efficiency has resulted from increasing mechanisation.

Machinery has certain general characteristics. First, any productive operation can be performed more quickly by a man with a machine than by a man without one.[2] Second, the greater the number of different operations which a machine is constructed to perform the more expensive it will be, and/or the less effectively will it perform any one of them.[3] Third, machines, like men, grow old and die at a rate which is only mildly influenced by the amount of work which they do. Fourth, machines, unlike men, do not benefit significantly in efficiency or length of life by resting from their labours.

While I have stated these characteristics in a less precise way than would be desirable were I writing a treatise on technique, I think that both their meaning and their validity are clear in a rough and ready fashion. If this be so, then a little reflection establishes that, for a firm with any given number of workers, the greater the number of products which it produces during any

[1] My formulation does not exclude Smith's pin factory. The greater skill which his operatives acquired after the division of labour was precisely the ability to avoid unfruitful activities which concentration on one particular operation gave them. Since I am not trying to write a treatise on technology I have felt justified in regarding arrangements for concentrating energy at the operative point—such as the pulley and the lever—as being subsumed under my heading of preventing energy from being wasted.

[2] I do not intend this as a tautology; i.e. I do not *define* a productive operation as something in which a machine can be used. I am asserting that, as a matter of empirical fact, the production of any good *could* be decomposed into a series of operations capable of being mechanised; i.e. capable of having energy applied at a greater rate per second than a man unaided is capable of.

[3] Less effectively in the sense, either, that less energy per unit of time is effectively applied, or, that more energy is wasted.

period the smaller will be the cost advantage of more over less mechanical methods of production. This will be true indifferently, whether we consider the greater variety of output to be secured by producing more products at one particular time or by more frequent variations in the pattern of output. Or, alternatively, if the production objective is given, the bigger the output of a firm the greater the extent to which it can with advantage push mechanisation. In general, there will be one degree of mechanisation which results in production at lowest cost,[1] for any given production objective and size of firm. This degree of mechanisation and the organisation of production associated with it I will call the most appropriate technique.

Let us suppose that blue-prints and operating instructions for the most appropriate techniques for all conceivable production objectives were publicly available for consultation by managements. This is roughly the state of affairs implied by specialists on the production function when they speak of a given state of the arts or of technical knowledge. Then, if we abstract from differences in the effectiveness with which techniques are applied, there seem to be only three main reasons why the efficiencies of firms should differ:

(*a*) some firms had chosen the wrong production objectives;
(*b*) random influences were at work—I define these below;
(*c*) new and more efficient techniques had been discovered since the date at which some firms selected their objectives and techniques.

There is no doubt that each of these reasons helps to explain the differences in efficiency which are observed in reality. I shall suggest below, however, that none of them is adequate for our purpose.

If there were a single "best" production objective in any particular industry with a given state of technical knowledge—i.e. a unique optimum combination of objective and technique—it is unlikely that mistakes in choosing production objectives would in fact occur. But it is seldom or never the case that such an

[1] On the assumption, of course, of given prices for labour, machines, materials and fuel and a given life of plant.

optimum exists. For markets are not homogeneous. Most markets are such that there is room in the industries which serve them for both mass producers and jobbing firms and for intermediate species between them. And because there is scope for genuine choice there is scope for mistakes to be made. Indeed, it is a common complaint against British industry that too many firms continue to cater for special requirements, maintaining therefore a less mechanised technique, although customers have (or could be induced to have) no very rigid ideas on precise individual requirements but set (or could be induced to set) great store by low price and quick delivery. Moreover, not all firms have a free choice between the alternative objectives with their associated techniques. There are some objectives which, if operated with the most appropriate techniques, will involve a scale of output beyond the resources of most firms in the industry. The fact that the larger the firm the greater its freedom to choose between the alternatives is the most general justification for the doctrine that there are economies of scale.

But a mistaken choice of production objective cannot be the main explanation of the efficiency differences we are investigating. For a production objective and its associated technique are things which, once it has chosen, the firm has to live with for some considerable time. As I shall discuss more fully below, major reorganisations are expensive affairs which are not embarked on very often. Hence, a bad choice would condemn the firm to languish for a considerable period at the bottom of the class and the fairly frequent changes in relative efficiency which are needed to counter the transfer mechanism would not occur.

Consider now the influence of random factors, as I described them above. What I have in mind here is that, for example, a firm may rightly decide that it will get the best results by organising itself so as to produce three types and four sizes of pullover a year; it may, that is, have selected the appropriate objective. But it is always possible, and indeed likely, that from time to time it will choose an unpopular colour or style of garment. Its efficiency in such a year would be low by our measure although in more normal years it was high. Alternatively, if there is a general recession in trade in an industry, the firm which has

chosen to remain unspecialised with a relatively low degree of mechanisation will be better equipped to deal with the variety of small parcels into which demand has disintegrated than is the highly mechanised mass-production firm which is most efficient in the years of normal trade.

These, and similar cases which readily spring to mind, constitute a warning that if we examine the dispersion of efficiency at any particular time there will be some, so to speak transient, features about it which give an exaggerated appearance to the problem we are investigating. Some of the differences in efficiency which are measured by the index I have suggested will stem from causes which are not persistent enough either to activate the transfer mechanism or to necessitate anyone's worrying about them. But I start from the presumption that the differences which have provoked so much public comment are more fundamental. Having noted the warning, therefore, we must press on.

I suggested technical progress as the third reason why efficiencies might be expected to differ in the circumstances I have postulated; some firms would be less efficient than others because the technique they were employing, though once appropriate, had been superseded by later discoveries.

This, of course, is quite a likely state of affairs. As I have already indicated, techniques have a certain enduring quality; they cannot be changed overnight. Nor is this wholly, or perhaps even largely, a matter of the length of life of the specialised kinds of machinery which a particular technique may require. "Redeployment", of which a great deal used to be heard in the cotton industry some years ago, involved not so much changes in the types of machinery in use as its re-siting and the planning of new channels along which the work should flow. Work must often come to a stop, sometimes for quite long periods, while reorganisation of this kind is in progress.[1] Moreover, as I understand it, the labour skills required by modern techniques of production are, increasingly, the capacity to participate effectively in teamwork, rather than the ability to do any job within the trade which

[1] A more recent and dramatic example is the reorganisation of the Standard Motor Company's tractor plant which first brought the term "automation" into the vocabulary of the breakfast table.

characterised the old-style craftsman. The craftsman working on his own or with an assistant requires a comparatively short time to overcome transitional difficulties when he moves from one type of work to another. But anyone who has gone through the experience of retraining, say, a gun-crew in the drill for a new gun and has seen the length of time required before the same perfection of co-ordination is achieved on the new as on the old weapon will appreciate that, in industrial operations also, working up a new technique which involves team-work is not something to be lightly undertaken.

If, then, better techniques have been discovered—an addition has been made to the file of blue-prints which we are supposing to exist—a certain divergence of techniques of production from those most appropriate will be a normal feature of any industry. Moreover, in so far as firms have adopted their techniques at different dates, there will be a certain dispersion of efficiency as between firms for this if for no other reason. It is equally clear that in these circumstances the relative efficiency of any individual firm will not be constant. As the technique practised by a firm diverges further and further from that most appropriate there will at last come a point when it pays to make a change; technique will be reorganised and the firm will move from the ranks of the outmoded to those of the up-to-date.

If all differences in efficiency were of this kind the tendency of the transfer mechanism to produce an ever-increasing degree of concentration would be effectively counteracted. The relative efficiency of any individual firm would pursue a cyclical path, with its seven fat years balanced by seven lean years, and the average efficiency of all firms (averaging over a cycle) would be equal. The amplitude and duration of the cycles would be calculable, given the costs of changing techniques and the rate of technical progress, and, if the date of birth of all firms were known, the degree of dispersion of efficiency also.

But again we must suppose that those who have criticised the dispersion of efficiency in British industry have had in mind something more fundamental than the fact that change takes time. Were the criticism based on naïve interpretation of data on costs or profits it would be necessary to discount it heavily in

the light of the preceding discussion. But, in fact, the main evidence quoted has been the qualitative judgment of experienced industrial eyes.

So far we have not found what we want. None of the three reasons for efficiency differences which suggest themselves on the assumption that all technical possibilities are known by all is satisfactory. The first is too powerful; the other two are too superficial. The cause of the trouble is, fortunately, not far to seek. It is the assumption that the given "state of the arts" can be represented as a library of blue-prints, freely accessible to all. For reality is far removed from this.

There is, of course, something called the practice of the trade and, indeed, there are Professors of Technology, whose job it is to investigate and classify the state of the arts in different industries. But the literature of technology is a far cry from the universal systematisation of knowledge which we have so far supposed. I believe, in fact, that the state of the arts can properly be defined only as what the firms in an industry are doing. For in talking of techniques we are thinking, not of the countless ways in which production could in principle be organised, but only of those ways which, as we say, are practical possibilities. And practice is the only sure test of what is a practical possibility. The production engineer, or whoever's task it may be, can push forward the design of a new technique on his drawing board for quite some way. But there comes a point where only the test of actual application can establish its validity. And the practice of the firm which finally emerges from a long process of trial and error—the new technique as a practical possibility—will usually differ in many respects from the original conception of the production engineer. In this respect, as in many others, the development of language is a good guide to the nature of reality. The very fact that the barbarous neologism "know-how" has crept into the language points to the existence of something that we are unwilling to denote by a term, "technical knowledge", which can be taken to imply "capable of being written down and taught by professors".

Having redefined a given state of technique as being the aggregate of what firms are actually doing in any particular period,

we must note two abiding features of what firms do. In the first place, firms tend to be more or less secretive about what they are doing. In the second place, they tend to change their practices from time to time, as we have already seen.

The motives for secrecy are sufficiently obvious. If the firm believes itself to have found a better technique (or a more lucrative production objective) than those practised by its fellows in the industry, it has every incentive to preserve its singularity and profit thereby for as long as possible. Keeping secret the nature and the organisation of the productive operations which it performs is one way of guarding technique. But men have eyes and workers move from firm to firm, and, were this the only safeguard, knowledge of individual techniques would be diffused through an industry fairly rapidly. What can be guarded much more easily is knowledge of the sort of results which the firm obtains from operating its particular technique. It is of comparatively little value to firm A to know how firm B is organising its production if it does not know whether the latter's costs are higher or lower than its own. This is why a firm's costs of production are one of its most jealously guarded secrets, a fact which it has always been difficult to explain on the basis of the traditional theory of the firm.[1] Profits will, of course, provide a general guide to the relative merits of the practices of different firms, but financial data as published are notoriously difficult to interpret. In the absence of information on turnover—another figure which firms usually treat as confidential—the possibility that higher profits result simply from working more nearly to capacity cannot be excluded. The possibility that another firm gets more advantageous terms from its suppliers also blurs the picture. The extent to which the success of the technique used by one firm depends on adhering to a particular rather rigidly defined production objective is another unknown. The list can be extended indefinitely.

This ignorance and uncertainty about precisely what other firms are doing and the results they are obtaining is, I believe, a large part of the explanation of efficiency differences. It is, so

[1] The desire to weaken the bargaining strength of the buyer is, of course, another strong motive for secrecy about costs.

to speak, the grit which prevents a successful advance in technique by one firm from being imitated immediately by all firms which are in a position to take advantage of it. And it is this grit which is, at the same time, responsible for the fact that advances in technique do take place. For if any firm knew that it could copy immediately and painlessly all improvements made by other firms, then no firm would have any incentive, other than the instinct of workmanship, to originate an improvement. And the instinct of workmanship is a frail vehicle to rely on to carry the torch of progress.[1]

On this view the degree of dispersion of costs in an industry will depend primarily on the rate at which innovations in technique and production objectives are being made—their frequency and their magnitude—and on the speed with which they are diffused. And the rate at which innovations are being made will depend, in a fashion which I now want to analyse more fully, on the speed of diffusion. In the course of this analysis I hope to show that there is some reason to expect that the lead in innovations will not rest always with the same firm. For this is a necessary condition for the operation of the transfer mechanism not to push the industry, albeit now more slowly, towards a steadily increasing degree of concentration.

I have already suggested that a change in technique or production objective is not something which a firm undertakes lightly. It involves expense and the risk that the expense will be wasted if the change is one for the worse. It is perhaps even more important that it involves thought, for rethinking of established habits is a painful process which most men will do a great deal to avoid. It follows, therefore, that a change is likely to be made only if a firm has the expectation of great gain from making it or, perhaps more commonly, the fear of loss from not doing so.

The firm which at any particular time ranks amongst the most efficient in the industry is unlikely to feel either of these compulsions. Having set a wide gap between itself and a substantial part of the industry it will feel little immediate fear of being overtaken by others on a scale sufficient to threaten its position.

[1] On a strict interpretation of the frictionless situation I have envisaged *no* innovations would take place if there were the slightest chance that innovations might be unsuccessful.

Instead, it will be expanding in size at the expense of the industry's tail; all the capital which it can get hold of will be readily absorbed in expanding capacity and any money diverted into innovational expenditure will be at the immediate expense of increasing its size. Indeed, on a less abstract view of the process, one should probably envisage the firm as having planned for expanding capacity at the same time as it adopted the new technique (in the event successful) to which it owes its present preeminence. At the most general level, it is surely true that further progress can be conceived as a practical possibility rather than an abstract idea only when there is discontent with present practice, and discontent is most easily aroused by the example of others. In short, in the carefree and complacent atmosphere of easy expansion there will be little incentive for managements to force themselves to the painful and difficult task of re-examining the fundamentals of their practice.

All these arguments in reverse will be impelling the less efficient firms in the industry towards such a re-examination and towards experimentation with new methods. They will be bending all their efforts to divine the secret of the success of their more fortunate fellows. But their aim will not be simple imitation; it will be to do even better. Indeed, it is doubtful whether the distinction between innovation and imitation has much value in this context. All innovation contains an element of imitation, in the sense that new practice is rooted in, develops from, the practice of the present. And the successful innovator is he who can view present practice with an objective as well as an imaginative eye, which neither finds change an end in itself nor is wedded to the past by bonds of sentiment and habit. Therefore, in one sense, the outside firm may be in a better position to improve on the best present practice than is the firm which originated it. The former suffers from the disadvantage of less intimate knowledge. But it is free from the distorting influence of the pride of creation, and can judge with an unprejudiced eye the *pis-allers* which the originator has developed in the process of translating the idea of the production engineer into the practice of the firm.[1]

[1] It is often said that the relative decline of the Ford Company during the 1930's was the result, in part at least, of over-long attachment, through sentiment or habit, to a recipe which had ceased to be appropriate.

To set against the greater incentive and the less prejudiced eye, the firms which have been left behind in the race suffer the disadvantage of shortage of capital. Innovation is, to repeat, an expensive process and its results are never predictable with certainty. If the dispersion of costs in the industry is at all substantial, the highest cost firms in the industry—they may well be so because of unsuccessful innovation—are likely to be making losses instead of profits and may find it virtually impossible to find the money needed to finance a reorganisation. Even for those whose position is less critical, there will be the fear that a change may be only for the worse, that good money may be poured after bad, and the temptation to believe that what is really wrong is the "state of trade" and that patience and economy are all that is needed.

It is impossible to be at all precise about how these opposing forces will balance themselves out. But it seems highly probable that the next advance in technique will be made by some other firm than the one which, by means of the last advance, made itself into the most efficient in the industry. And this carries us some way towards finding the opposing force which is needed to prevent the operation of the transfer process from producing steadily increasing concentration. For in such circumstances the work of the process would, so to speak, be constantly undone. The most efficient, and therefore growing, firm of today would find itself outclassed tomorrow, with a consequent slackening in its rate of growth or even an absolute contraction. And contrariwise.

What more can be said about this innovation mechanism which I am opposing to the transfer mechanism? Something, I think, about the factors which will determine the intensity of the drive for successful innovation in the industry as a whole. It is a legitimate conclusion that the longer relative efficiencies within the industry remain undisturbed the more powerful and persistent will become the efforts to disrupt them. For we have seen in the preceding chapter that the longer these relatives persist—the longer the period of uninterrupted activity enjoyed by the transfer mechanism—the smaller will the dispersion of costs become and the slower the rate at which it will contract further.

And this is simply another way of saying that the industry will come to be more and more dominated by those firms whose costs are low, and that their ability to grow further, except at each others' expense, will become less and less. The low-cost firms having eaten most of the fat from the high-cost tail, each will find that the price of continued expansion is an improvement in his relative efficiency, and, each feeling this, the spur of expected gain will be reinforced by the fear of certain loss should another forestall him.

Moreover, the longer the transfer mechanism has been operating unchecked the more will the sizes of firms have changed. Instead of the smooth and continuous process of expansion or contraction in firm size assumed in the argument of the preceding chapter, we should think, more realistically, of changes in capacity taking place in bursts. And changes in size of this kind, whether upwards or downwards, will themselves usually involve a substantial process of reorganisation in the firm, which is unlikely to result simply in an exact re-duplication of the existing technique on a larger scale. This is all the more true if the industry is one in which "economies of scale" are important. For this reason also, therefore, it will be true to say that the longer the transfer mechanism operates the more likely is there to be a disruption of the constellation of relative costs on which it is working.

Let me now try to sum up the argument. I have suggested that the dispersion of efficiency at any particular moment of time results from the fact that innovations in technique (which may also be associated with innovations in production objectives) take place, that they are made by individual firms within the industry, and that knowledge of them is diffused only relatively slowly. The transfer mechanism, operating in the way described in the preceding chapter, progressively reduces the average degree of cost dispersion, by exalting the low cost at the expense of the high. But the smaller the dispersion becomes the slower the rate at which it contracts further. This is all the more true because the rate of experimentation[1] with new techniques will tend to intensify as the dispersion of costs is reduced. On the assumption

[1] I call it thus to cover unsuccessful innovations also.

of a self-financing industry made in the previous chapter, the cost of unsuccessful experimentation is something which has to be financed from the profits of the industry, so that the average rate of profit in the industry is higher.[1] This acts—just as would an increase in α—to slow down the rate at which the dispersion contracts. At some point the process of experimentation results in successful innovations. Since these will usually be made by firms other than those which were successful on the last occasion, the effect will be a disruption of the old cost-relatives, and much of the redistribution of business between firms which the transfer mechanism has brought about it will now have to set to work to undo.

In actual fact the process of relative growth, of innovation, and of changes in relative efficiency will take place in a much more complex and continuous fashion than has been implied above. The process of competition—for I think it is something like this which people have in mind when they use that term—cannot be reduced to such mechanical terms without losing a great deal of its reality. Nevertheless, the cyclical form in which I have found it most simple to expose the process may have more than merely pedagogic value. It is plausible to suppose that in reality also there will be periods of what may be called ingestion, during which the structure of efficiency-relatives is broadly undisturbed and the strong are engaged in consuming the weak, and that these will be followed by periods of revolution, when technique is in the melting pot, old kings are being dethroned and new ones are coming to the fore. We should at least bear this in mind when confronting empirical material.

* * *

Having analysed the two mechanisms under simplified assumptions we are now in a position to refine somewhat the formulation of the problem given in Chapters IV and V. The analysis so far has failed to suggest anything analogous to the optimum

[1] Experimentation will compete with expansion for the retained profits of the firm. Successful innovations will be amortised in the usual way, but the costs of unsuccessful experiments are likely to be written off specially. For the industry as a whole, the cost of unsuccessful experimentation is a drain on profits.

size of firm of the traditional theory, against which reality could, in principle at least, be measured. Progress, I have suggested, consists in the creation of cost dispersion, and it is the tendency of cost dispersion to be eliminated which generates progress. Hence it does not make sense to contemplate an "ideal" or "minimum" degree of cost dispersion. Indeed, I have suggested that a tendency to fluctuate, although perhaps without trend, will be one of the characteristics of cost dispersion. Nor does it make sense to think of an "ideal" rate of progress.

What then is the meaning—for meaning I still feel it to have—of a statement to the effect that such and such an industry is unprogressive and has too widely dispersed costs as a result of the rules of the game under which it is operating? It can be interpreted only as an allegation that the rules of the game are interfering with the operation of the mechanism which I have analysed; more specifically, with one of the three basic features on which its functioning depends—growth of the more at the expense of the less efficient, freedom to experiment to the limit of one's resources, and some delay in the diffusion of new techniques. These then are the empirical questions to which we must address ourselves. Do certain rules of the game diminish the penalty for failure to innovate or the reward of success, where penalty and reward are in terms of relative growth? Do they restrict freedom to experiment with new ways of doing things? Finally, and probably less important, do they influence the rate at which knowledge of new techniques is diffused?

To frame the questions thus is not to attribute any absolute value to the sort of mechanism which I have described. It is not to say that a social engineer with a completely free hand could not design a system which would work better. Nor is it to suggest that the analysis I have given is anything more than a very abstract and much too generalised picture of what people mean when they speak of a competitive economy. It is merely to imply that it is difficult to envisage any alternative set of forces which could be substituted for this mechanism *in a market economy*. If unfitness carries no penalty, then a disinterested desire to do things better will be the sole force making for progress. But if the unfit are to be eliminated it is only their fellows who can

remove them. And, if the slaughter is not to be senseless, freedom to mutate must be open to all.

It is time, now, to drop the assumption that the industry is isolated and to consider how our conclusions will be modified if firms can enter or leave the industry.

Chapter VIII

ON ENTRY AND EXIT

I HAVE so far discussed the functioning of the transfer and innovation mechanisms as if no firm were in more than one industry, no firm could move from one industry to another, and no new firms came into existence. None of these assumptions accords with reality.

We need first to consider the type of firm which may establish itself in an industry as an addition to the existing population. The distinction I have in mind relates to the firm's antecedents. All additions to the list of producers in an industry during any period can be divided into two types; firms which have a previous history in other industries, in the sense that they are offshoots of firms already established in those industries; and firms with no such previous history, which are new-born and have no link with other firms. I shall term these dependent and independent respectively. I hope to show that there will be important differences between the behaviour of the two types. Hence, we must enquire about the relative importance of these two constituents of the stream (or trickle) of new entrants.

So long as we assume that potentiality for subsequent growth is randomly distributed between dependent and independent entrants the question as to their relative importance can be given an unambiguous meaning. We are asking in what proportions that part of the total addition to an industry's capacity during a period which does not belong to firms already operating at the beginning of the period is divided between dependent and independent entrants. There are as yet, unfortunately, no empirical data which provide a direct answer to this question. We can, however, get some light by indirect means.

In order to establish a new firm, its creators must have at their disposal a substantial amount of money which will enable them to acquire a part of the economy's current output. Since the greater part of the country's wealth is owned by individuals and

the distribution of wealth is very unequal, there must be a considerable number of individuals who own wealth equal to the cost of establishing a viable new firm in most lines of business. But what they hold is titles to wealth, share certificates and the like, rather than money. Hence, if one of them wished to set up a business he would have to sell his titles to wealth. This requires that there should be someone else in the system who is willing to make over, in exchange for titles to existing physical (capital) goods, the money of which he finds himself in possession through spending less than he earns; i.e. by and large, someone must have saved from current income if someone else is to be able to turn his assets into money. By and large again, the amount of titles to wealth (old goods) which asset-owners as a class can turn annually into money (command over new goods) cannot exceed the annual savings of the community.[1] In fact, it will be far less.

As we have already noted, a large part of the nation's savings is made by existing firms (profits put to reserve) and used by them to finance their own expansion. So the ceiling on the amount of titles to wealth which individuals can convert into money contracts from total savings to personal savings. This last is no mean sum; it may have been roughly one-third of total savings in the United Kingdom in recent years.[2] But there are other claims upon it. A large part of personal savings, as defined in national-income accounts, is in fact saving by established businesses which have not adopted a corporate form, and will be ploughed back into these businesses just as are companies' undistributed profits. Another part, also large, is (from a conceptual standpoint) invested by those who make it, in new house property. And for what remains, existing companies will be making a strong bid by issues of new shares. All in all, therefore, the scope for holders of titles to existing wealth to turn them into money without coming into fierce competition with other seekers after funds must be pretty small. This is roughly

[1] This statement, like most of those which accompany it, requires a whole set of implicit assumptions to make it unambiguous and fully true. My only defence for cutting corners in this way is that to tell the whole story would require a book, and this is a book about competition and not, more than is necessary, about anything else.

[2] The share of personal savings is available on a gross basis in the annual national income estimates. But in the absence of any usable definition of net savings the more relevant proportions of these can only be surmised.

what is meant by saying that the Stock Exchange is a narrow market. If any substantial number of persons tried to convert their shares into money the result would be a catastrophic fall in share values.[1]

There is a further barrier between the would-be (independent) new firm and the money it needs. It is by no means necessary that those who possess large amounts of wealth are the same as those who wish to set up in business. Indeed, the existence of the term "idle rich" (the best English translation of *rentier*) implies that frequently they are not. And even when the rich are not idle they will often already be in positions of command in established businesses, with little incentive to go madcapping on their own. Hence the business-man-to-be typically has the task of persuading others to undertake the trouble and expense of converting their titles to proven wealth into money, which he, an untried fledgling, will then risk for them in a new venture. It is little wonder that investigation always brings to light as characteristic of any capital market the difficulty of new (independent) firms in securing capital.

I conclude, therefore, that we are justified in regarding entry by independent firms as quantitatively small in relation to that by dependent firms, unless indeed new entry in total is so small as compared with additions to capacity by existing firms as to make its investigation not worth while. I am strengthened in this view by the further conclusion that growth-potentialities will not in reality be randomly distributed as between dependent and independent firms, but will be biased in favour of the former.[2] It follows, therefore, that what we have to examine is the implications for our analysis of the possibility for existing firms to cross industrial frontiers. I shall postpone to the next chapter consideration of the special case when a firm steps across the last frontier and dies.

We have to ask under what circumstances firms are impelled to cross industrial boundaries and what determines the direction in which they then migrate.

[1] Most transactions on the Stock Exchange are, effectively, exchanges of one title to wealth for another.

[2] Empirical evidence on this point has been published by Steindl. See his *Small and Big Business.*

The answers to both these questions are conditioned by two basic facts. The first is that the direction in which the firm will grow will be that in which it believes it can grow quickest. Like a plant, it will follow the light—or what it believes to be the light. The second is that an established firm, unlike a new (independent) firm, is always a specialised firm, specialised in both its technological and its market experience. Indeed, it will be recalled that this was the basis of our definition of an industry, and is fundamental to the whole analysis. Such specialisation means that at any point of time there will be what we may call a technological horizon, within which the firm will follow the light but beyond which it will not normally leap. Within its own industry, growth, by definition, faces it with no new (technological) problems. Industries which are, in a technological sense, neighbouring will present it with new problems should it decide to enter them, and it will be less certain about the quality of its performance in them than in its own industry. As its vision takes in more and more distant industries the country will seem more and more strange, and some point is eventually reached when, as we say, the firm would be out of its field. This is the technological horizon.

The radius of this horizon will vary from firm to firm. For there is not only technological specialisation of firms, but also specialisation of managerial functions within firms. The larger the firm the greater the extent to which it is possible and profitable to push such a division of managerial labour. The greater the division of managerial labour the less technologically specialised will be the higher levels of management. The captain of industry needs, not technological knowledge, but the ability to bring together, organise and supervise the people who have it. It follows, therefore, that the larger the firm the greater will be, generally speaking, the radius of its technological horizon. Conversely, there will be some minimum size below which the technological horizon is coterminous with the boundary of the industry as I have defined it.

This minimum size, below which movement across an industry boundary is not seriously entertained, does not depend for its reality merely on the smaller firms' recognising their own limitations. Potential lenders also are likely to take the view that the

smaller firm had better see itself strongly rooted in the favourable soil of its home ground before it begins to send out suckers to neighbouring territory. Both these considerations are reinforced by the fact, referred to in Chapter VI, that, quite apart from technical considerations, the resilience of a firm increases with its size; as it grows larger it begins to acquire a certain monumental character. Since a venture into the (relatively) unknown will normally involve greater uncertainties than continued expansion in the traditional groove, there is a strong incentive to wax fat at home before looking farther afield. And if there are moderate (technical) economies of scale in the firm's own industry and those which lie within its technical horizon the case becomes overwhelming.

It goes almost without saying that it is only for the relatively efficient firm that migration presents itself as a real choice. There will, of course, always be instances where a firm, as a last resort, seeks escape from Queer Street by a radical transformation of the nature of its activity. But in the normal run the firm which is struggling in the lower levels of the efficiency ranking will be devoting its energies and such money as it can get its hands on to keeping alive and raising its rank in the line which it knows.

I conclude, therefore, that the class of potential migrants consists of largish efficient firms, and that their trespassing will be confined within a technological horizon whose radius is the greater the larger is the firm.

The potential migrant becomes an actual crosser of industrial frontiers when it believes that its combined rate of growth in two (or more) industries will be greater than that which it would achieve in only one.[1] This possibility of increasing the rate of growth by entry into a new industry can arise for two reasons. Operating the two lines of activity may result in productive economies; that is, the firm's unit costs in one or both industries may be lower than they would be if it were in only one. Alternatively, there may be no productive economies, but the rate of profit which it expects to be able to secure in the new industry may be higher than that which it is earning in the one in which it operates at present. Migration which is inspired by the first

[1] Growth in this context means rate of increase in total assets.

motive is usually termed vertical integration. I shall refer to that called forth by the second motive as conglomeration.[1]

Now, if in fact it yields substantial economies, integration will become a condition of survival for firms in one or both of the industries in question. Entry will then be on a massive scale. But if, as we are assuming, industries tend always to be in approximate equilibrium, the mass entry of new firms into either would result in excess capacity on a large scale; a fundamental disequilibrium would be created and, unless total demand were increasing at an improbably high rate, the entered industry would be in for a prolonged period of very low prices. The prospect of such a development will be attractive to neither the intruders nor those intruded upon. Hence, a good deal of the process of integration is likely to take place through the coming together of existing firms in the two industries. According to the relative sizes of the firms which come together, their union will be described as merger or capture.

There will sometimes be a unique relationship between the two industries between which integration occurs, such that one finds almost the whole of its market in the other. Crude oil production and oil-refining and, to a slightly lesser degree, blast furnaces and steel-melting are pairs which have this relationship. In such cases, the effective result of integration, whether it be by merger or capture, will be that, where formerly there were two industries, now, for practical purposes, there is only one. It would clearly be profitless to try to modify the general analysis to accommodate such cases, for the basic concept of an industry, on which the whole analysis has been built, itself melts away. Revolutions require special tools for their analysis, and no good can come of trying to fit them into general frameworks.

But the more normal case will be one in which there is no such unique relationship and in which, therefore, only part of an industry is captured (or captures). The result will then be that both industries will contain a number of firms for which the relation between efficiency and growth is potentially abnormal,

[1] The expression "conglomerate bigness" was, I think, coined by Corwin Edwards to describe the multi-industry firm for whose variety of activity no convincing rationalisation can be found on technical grounds. See his essay on "Conglomerate Bigness" in *Business Concentration and Price Policy*.

since success in one industry can be used to provide the sinews of growth for that part of the firm which lies in the other, or *vice versa*. In this respect, the situation is exactly analogous to one in which the firm has been induced to put a foot in both industries by the second of the two motives which I suggested. I shall examine it, therefore, in conjunction with conglomeration.

As I have said, conglomeration will take place when a firm believes that it can achieve a higher rate of growth (profit) than it could by remaining uniquely specialised. The important word in this proposition is "it". The firm is not *directly* interested in average or marginal rates of profit or in any other rate of profit actually being earned by others; it is interested in the experience of others only in so far as this provides a guide as to the rate of profit which *it* can expect to earn. This expected rate of profit, in either its own industry or another within its technological horizon, can formally be split into two elements; the average rate of profit which will be earned in the industry in question and the excess over the average which the firm itself expects to be able to earn.[1] The "excess profits" will depend in turn on the dispersion of efficiency within the industry and the firm's position in the hierarchy. Such is the translation into formal terms of the homely, and plausible, notion that in deciding where to put its money the firm will pay regard both to the general prospects of the industries with which it is confronted and to the quality of the rivals which it will meet in each.

The firm's views on both questions will be heavily influenced by the current and recent experience of the industries which lie within its horizon. The comparison of average rates of growth (and changes in them) for different industries will present few difficulties, although successful extrapolation calls for both ability and good luck. As regards its future efficiency ranking in the industries which it may enter, or in its own, I find no assumption more plausible than that the firm will assume its efficiency relative to the best firms in new industries to be roughly the same as currently in its own industry. Thus, if its costs are 10% higher

[1] Since we are assuming that only efficient firms are potential migrants this excess will normally be positive.

than those of the best firm in its present line, it will assume that a similar relationship would hold, once it was established and had settled down, in other industries, although its view may become somewhat more pessimistic as the horizon is approached. And, if this be in fact the sort of assumption which the firm makes, its view of the "excess profits" offered by expansion in alternative industries will depend on differences in the extent to which efficiencies are dispersed within them.

We may further suppose that differences in dispersion (the quality of rivals) will have more influence on the firm's decision as to the industry in which its best interest lies than will differences in average rates of profit (general prospects). For, given the size of capital/output ratios in most industries, a small difference in the proportion by which a firm's unit costs fall below the average will easily outweigh the sorts of differences in average profit rates which would rule in the absence of entry. Suppose, for example, that the rates of growth of demand in two industries are 5% and 15%, that taxation takes one-half of net profits, and that the ratio of capital per unit of output to average costs is one-third in both industries. Then if a firm can achieve unit costs equal to roughly 83% of the average in the industry with the lower average rate of growth (profit) it will secure the same rate of growth (profit) as if its costs were 90% of the average in the more rapidly growing industry. While such wide differences in average growth rates are uncommon, the difference in relative costs which is sufficient to extinguish them is modest in comparison with those suggested by the empirical data of later chapters.

We may imagine the firm, therefore, as (in formal terms) trying to form a view of the dispersions of efficiency in the different industries which lie within its horizon and as allowing this comparison to have a heavy weight in determining its ultimate choice. But it would be wrong, I think, to suppose that the choice between migration and continued expansion in its original industry is on all fours with that as to which industry to enter when once migration has been decided upon. The temptation to continue in an accustomed groove is strong in any line of activity. For, with whatever rationality, one attaches

greater uncertainty to one's view of the future of the less familiar. Moreover, the matter which is most certain in this question of inter-industry movement is that which least favours movement; the firm can be certain that, whatever its ultimate efficiency in the new line, there will be a painful, expensive, and sometimes prolonged interregnum while it is learning the job. It is more realistic therefore to think of the firm, not as eagerly scanning the horizon for a chance to break into new fields, but as forced reluctantly from its accustomed matrix by the increasing difficulty of further growth there.

We have seen from the analysis of the transfer mechanism that its effect is steadily to reduce "excess profits", even though the (weighted) average rate of profit in the industry may be constant.[1] As time goes on, therefore, the large efficient firm will find that, although it is still making profits, it is increasingly less satisfactory to invest them in further expansion in its own industry. Its rate of growth will be tending to slow up, irrespective of any damage which it may be suffering through the innovation mechanism, for the very success of it and others like it steadily withers the fruits of that success. In so far as other firms in the industry are also large, the difficulty of further growth for any one is further increased by the special resilience of the large firm to which I have several times referred. It is, in fact, the fall in dispersion in its own industry which stimulates the firm to look to other fields.

My suggestions are, therefore, that what we may call imperialist tendencies begin to develop in an industry when a considerable part of its output is in the hands of large firms which do not differ greatly in efficiency. This state is what is meant when we speak of an industry's being concentrated or oligopolistic, and, as I have tried to show, it is one which the transfer mechanism is constantly tending to produce. The counters to this tendency are the process of innovation and the entry of new firms. But, as I have also indicated, there is no necessary reason why, in any given period, the innovation mechanism should act so as exactly to offset the transfer mechanism. Indeed, it is highly improbable that it will. There may well be lengthy periods when no firm is successful

[1] Since output prices fall over time through the operation of the transfer mechanism.

in making a major innovation in technique, or the privilege may fall twice on the same firm or firms. Moreover, the innovation may frequently be a method of production which can be operated effectively only on a large scale. There is, therefore, a perennial chance that concentration will develop, and to say that there is a chance that something will happen means that sooner or later it will.

If I am right in suggesting that imperialism becomes significant only after concentration has developed in an industry, and that entry is primarily by dependent firms, it follows that the ultimate appearance of concentration in some industries is more or less inevitable, quite apart from any effect of economies of scale. Once such concentration develops, firms in the industries affected acquire a stimulus to look outward, to seek other industries where the dispersion of efficiency is wide, the average rate of growth is high, and a few large efficient firms have not yet secured a dominant position. In other words, the phenomenon of entry becomes significant for the system as a whole only when islands of concentration have appeared. The inhabitants of such islands develop imperialist tendencies and their colonisation is directed primarily to those industries in their neighbourhood which exhibit the greatest dispersion of efficiency. It follows, therefore, that migration is a one-way traffic from the concentrated to the less concentrated and not, normally, either in the reverse direction or between pairs of concentrated industries. And this means that there is a certain irreversibility about the process of concentration. Once an industry becomes concentrated it will tend to remain so.

The establishment by firms in concentrated industries of subsidiaries in other industries is equivalent to a partial suspension of the transfer mechanism in the former. Capital will now be flowing out of the industry and this raises the average rate of profit which is consistent with maintaining equilibrium. The rate of contraction of σ will diminish and this *a fortiori* since it is likely to be the firms earning the highest profits (the most efficient) which are the most active imperialists. There will tend to be therefore, a certain ossification in the structure of firm sizes. Moreover, once a firm has taken the step of establishing itself in

another industry (or more than one) its total powers of resistance to the transfer mechanism are strengthened. Funds can now flow without impediment between the several parts of its body corporate. There is, as we have seen, an element of luck in successful innovation, and the chance that a firm will be unsuccessful in the innovation competition in two or more industries at the same time is clearly less than the chance that it will be unsuccessful in one industry at some point of time. Hence, part of the reward for spreading into a number of industries is a greater stability in size of both the firm as a whole and its several industrial parts. We may say, therefore, that the imperialist process is equivalent to a partial suspension of the transfer mechanism and that the colonies which are the fruits of the process cause a lasting impairment of the mechanism.

In so far as the transfer mechanism in the imperialist industry is weakened the innovation mechanism is weakened also, since the fear of loss, which is a major stimulus to innovation, becomes less acute. There still remains the hope of gain. And there is always the hope (and fear) of some innovation which will be so great (and can be guarded so long) as to break down the resistance of other firms even in a highly concentrated industry. Live and let live is a policy to be abandoned at the prospect of a quick safe killing. Nevertheless, the net effect can scarcely be other than a weakening of incentives for innovation. And this means that the tendency to ossification in the structure of concentrated industries will be all the stronger.

The effect on the non-concentrated industries with relatively widely dispersed costs, which are the objects of colonisation, will be a more rapid rate of contraction in the dispersion than would otherwise occur. The creation of capacity financed from sources outside the industry is equivalent to a reduction in the equilibrium average rate of profit in the industry. And we have seen that the transfer mechanism will work the quicker the lower is the average rate of profit. Moreover, if the large efficient firms which provide the entrants are justified in their expectation that they will be relatively efficient in their new industry, the rate at which average unit cost falls will be all the more accelerated and so, therefore, the rate at which prices fall and the least

efficient are eliminated. If, however, the argument of Chapter VII is correct, the effect of this activation of the transfer mechanism will be a similar stimulus to the innovation mechanism; innovation will proceed at a faster rate also. It is impossible to form any general conclusion as to the net effect on dispersion of these opposing stimuli. But it is reasonable to suppose that progress will be faster than had the industry been free from entry.

These are, so to speak, the impact effects of new entry. Once the colonies are established, the effect is, as we have seen, that the industry now contains a number of enterprises which are, potentially and within limits, immune from the transfer mechanism. If their parents are willing and able to finance them, they can continue to operate at a loss without suffering any contraction in size, and, conversely, high rates of profit may not lead to high rates of growth.

To sum up. The general effect of relaxing the assumption that firms do not move across industry boundaries is to reinforce the conclusion of the previous chapter, that to look for "ideal" values for dispersion or progress against which to compare reality is to seek the philosopher's stone. We cannot *predict* what dispersion would be under given circumstances if the circumstances assumed are to have any relation with reality. What has been added in a positive sense to the conclusions of the preceding two chapters is that, in the absence of rules of the game which prevent it, concentration is likely to arise in some industries; that it will then tend to persist; and that the existence of such concentrated industries constitutes a force tending to prevent wide and persistent divergences between the dispersions in non-concentrated industries. The practical significance of this force depends on the quantitative importance of new entry and multi-industry firms, information on which is considered in Chapter XI.

Chapter IX

HOW EQUILIBRIUM IS ACHIEVED

IN CHAPTER VI, I defined equilibrium as a situation in which demand and capacity were equal and an equilibrium price as one which would maintain this equality. I have assumed throughout the ensuing discussion that equilibrium and equilibrium prices were continuously achieved. It is now time to examine this assumption. The prices which are charged in an industry and the rate at which capacity is added to result from the actions of individual firms, each of which has only very imperfect knowledge of the rate at which total demand for the industry's products is growing and of the actions and financial situation of its competitors. The mechanism—reaction system—through which equilibrium is approximated is by no means self-evident.

I gave a very provisional account in Chapter VI. If a firm finds that it is not producing to capacity but is still earning the profits which would permit further growth, the appropriate corrective action will be to find more customers, by offering more attractive terms than it believes some other producers are willing or able to give. Conversely, if demand persistently exceeds its capacity it will take action to accelerate the rate of growth of capacity, by earning more money, by raising prices. If most firms find themselves in the same position, if there is a *persistent* buyer's (seller's) market, a majority of firms will be tending to reduce (raise) their prices. Other things being equal, the lower (higher) is the general level of prices and profits the slower (faster) will be the potential rate of growth of the total capacity of the industry. But the rate of increase of demand for the products of the industry will, if affected at all, vary inversely with such changes in prices. Prices will, therefore, tend to move in the direction required to restore equilibrium in the industry.

This is all very well so far as it goes. But it does not explain why the industry and its prices should not fluctuate wildly about the equilibrium position. What is it that prevents the price cuts which a buyer's market induces from greatly overshooting the

mark, so that the industry moves, not to equilibrium, but to an acute seller's market?

The immediate answer lies in the fact that firms have an idea of what constitutes a "normal" price or, more usually, a "normal" margin of profit over costs. That the notion of normal prices is a real one is evidenced by the fact that business-men speak of "fancy" and "give-away" prices or of prices being at a level which can't last. Moreover, they refer to normal margins or mark-ups when questioned about how in fact they set their prices; to such an extent that there is now a school of "full-cost" price theorists among economists. It is this idea of the normal margin which acts as a breakwater against the waves of instability which would otherwise rock the industry.[1]

I suggest that the system works in the following way. Firms will always plan to be in equilibrium, because it is wasteful not to do so. They will plan also to charge "normal" prices. But their plans may go wrong, because demand can change much faster than capacity, and they do not expect always in fact to be in equilibrium. If a firm finds itself in disequilibrium the natural tendency is to assume that this state will be only temporary, since it has planned to be in equilibrium. And, since such changes are difficult, painful, and sometimes irreversible, it will make no fundamental move to correct the disequilibrium. This is why prices are proof against passing upsets. If the disequilibrium persists, however, it will come to be recognised as more fundamental and action will be taken to correct it. Price changes are the appropriate action. But such changes will be departures from the deep-rooted notion of what is normal, and this sets limits to the size of change which will be entertained. The notion of normal price will, in fact, provide an anchor, which tends to confine price changes within not too wide a range.

It seems then that all is well so long as the normal price or profit margin which provides business-men with their point of reference approximates reasonably closely to the equilibrium price as I have defined it. And this will usually be so. The most convincing interpretation of the normal price is that it is the

[1] Like most ideas (in this book and others) this appears in Marshall; business-men are restrained from extreme price-cutting or excessive profiteering by fear of "spoiling the market".

price which business-men believe will rule when the industry is in equilibrium and will tend to keep it in equilibrium in the future. And what is more natural than that they should base their view of the price or margin which will do just this by reference to the prices which have in fact (or they think would have) maintained equilibrium in the past. Since it contains a reference to the future, the normal level of prices is not an incorrigible value. It will be modified in the light of experience. But, since it is rooted in the past, the modifications will be made with caution; the anchor will drag only slowly.

There is, however, one further problem, which will compel us to plunge more deeply into the practice of pricing. The stabilising force of the notion of normal prices will become progressively weaker the longer a disequilibrium persists; the longer the anchor drags, the faster will it move. The plausibility of my account rests, therefore, on the assumption that fundamental disequilibrium will be corrected fairly quickly.

Such an assumption seems fully justified if the disequilibrium consists of a deficiency in capacity; new capacity can be created quite quickly in most industries. And I have so far acted as if capacity could be eliminated with equal speed. I have assumed, in fact, that just as the earning of a 5% rate of net profit in one year provides the means for a 5% expansion of capacity in the following year, so does a negative profit of 5% entail a shrinkage of capacity by a similar proportion. This assumption has made it possible virtually to ignore the fact that an expanding firm needs to pursue an active selling policy in order to win new customers. The argument has been conducted as though customers were driven to it by the shrinkage, or failure to expand, of the capacity of their previous suppliers.

But the assumption that capacity is so eliminated is, in fact, invalid. It represents a false assimilation of the financial concept of depreciation and the physical concept of capacity. The value of a machine diminishes continuously throughout its life because its life is not eternal; the shorter is its remaining life the smaller will be its value. But there is no physical counterpart of depreciation in the shape of a continuous decline in the productive capacity of the machine. By and large it will retain the same output

capacity throughout its life, its capacity falling sharply to zero when it is scrapped.[1] Thus, failure to cover depreciation, which is the meaning of negative net profit, does not entail any immediate fall in the capacity of a firm, but merely that when its assets come to be scrapped it will not have the money with which to buy new ones. We have, therefore, to examine two problems. What perturbations are introduced into an industry when excess capacity is created and is not quickly cured by the growth of demand? And how can the creation of excess capacity be avoided in an industry in which the transfer mechanism is operating? I shall deal with these in reverse order.

The first thing to note is that what we are concerned with in analysing the transfer mechanism is not assets but firms. The tie between a firm and its assets can be broken in other ways than by the scrapping of the latter. Thus it is possible, in principle, to conceive of a firm which was earning negative profits as selling part of its assets in the second-hand market and using the proceeds to replenish its inadequate replacement fund. In practice such a procedure will not usually be a real possibility.[2] But it directs attention to the fact that assets can disappear while firms remain and *vice versa.*

A collection of assets will be scrapped when their owners and all other sane men believe that in no circumstances which it is reasonable to contemplate will it be possible to earn by working them a revenue greater than the direct costs involved in so doing; that is, when the gross margin of profit expected is zero.[3] But a firm will die when it believes that it can do better for itself by disposing of its assets and going out of business than by continuing to operate. This decision may be free, in the sense that the firm could stay in business but receives an offer attractive enough to persuade it not to; or, alternatively, the firm may be

[1] A young (not too young) hen fetches more than an old, not because it is a better layer—indeed, having had less practice it may be a worse—but because its laying life is longer.

[2] A factory is planned to have a certain capacity and its different items of equipment are complementary. Hence, it will usually be impossible to take away items piecemeal without disrupting the whole organisation of production. Separate establishments, or even specialist departments of the same establishment, may be susceptible to this cannibalisation procedure, but such cases will be exceptional.

[3] This is only strictly true on the assumption that the assets have no scrap value.

forced to close down because, having failed to earn a positive gross profit margin for some time, it has exhausted its cash and credit and can no longer pay its bills. But in either event there is no necessary reason why the firm should live as long as its assets. A firm cannot exist after its assets are scrapped, but the reverse is not true.

It is clear, therefore, that there is an important discretionary element in the working of the transfer mechanism. At the one extreme, we can envisage the firm which is earning negative profits as staying in business to the bitter end, "working its capacity to death", as we say. At the other extreme is the case where the firm sells out to a more efficient rival as soon as it begins to go into the red. Indeed, this last may be regarded as only a special example of the general possibility of fusion between any two firms in the industry—horizontal integration, as it is usually called—even though both of them may be earning positive net profits. I do not propose to pursue the general case further, except to take note that it provides yet another reason for not expecting to find any simple, easily detectable, relationship between relative efficiencies and relative growth rates in empirical material.[1] But with the question of the speed of extinction of the inefficient firm we begin to come nearer to the heart of the matters with which rules of the game are concerned.

The immediate choice as to the manner and speed of its death lies with the negative profit firm itself. But it is the efficient expanding firms who make the ultimate decision. For it is they who provide the market for going, though unprofitable, concerns, and so determine the attractiveness of a quick death. It is they also who, by the policy they adopt, largely determine the length of time for which a firm can go on working its capacity to death. What is needed to kill off the firm is that *gross* profits should become negative for a period. And, typically, the relapse from negative net into negative gross profits will result, not from the increased cost of working machines tied up with string, but from

[1] One public company may acquire another either by buying up its shares for cash or by offering a new issue of its own shares in exchange. In the first case, the cost of the acquired capacity will depend on the vagaries of Stock Exchange values; in the second, no draft on the firm's liquid resources will be required. In either case, therefore, the cost of the expansion may be very different from that of building a new plant.

a combination of falling prices and lost customers. The onset of this combination depends on the strength of the efforts which the expanding firms in the industry are making to win new customers.

The two alternative modes of death which we have been considering correspond, in fact, to two alternative policies for the efficient. On the one hand, they may pursue a rather passive policy, allowing the inefficient slowly to decay, and earning in the meantime a higher rate of profit than they would otherwise receive, at the cost of a delay in their own expansion. On the other hand, they may combine an aggressive policy for winning more custom with a readiness to take over firms in difficulty. Reality will be somewhere between these extremes. The extreme to which it inclines will depend to a marked degree on the nature of the aggressive weapons at the firms' disposal. These we must now consider.

The immediate arbiter of whose capacity shall be employed is the customer. It is he who places orders, and we are imagining him to have a choice as to where he places them. He will place his order with the firm which he believes to be offering the best combination of price and quality—the best value for money. The concordance between his belief and reality will depend on two things; first, the information available to him about alternative possibilities; second, his ability to assess that information.

We must distinguish two situations with respect to the information available about price; the posted price and the negotiated price. I mean by posted prices the case where the producer publishes to all interested parties the prices at which he will sell his goods, and where the individual purchaser can only accept or reject the offer and has no possibility of bargaining. Negotiated prices are those which result from bargaining between the individual customer and the supplier. By definition, there can be no universally known price in such cases, since the same commodity may be sold at different prices to different customers. The customer will seek out offers and make counter-offers; the producer can never be certain as to the prices which his rivals are really offering—as opposed to what they or his customers say they are offering; and the customer can never be certain as to how far

the price asked by the producer is above what he would really be prepared to accept.

The distinction arises from the characteristics of the sale or the customer rather than from any inherent characteristics of the goods. For it is a necessary condition for negotiated sales to be possible that the number of sales be relatively few; otherwise there would simply not be time for negotiation. And this means in turn that each individual sale must be large relative to the total turnover of the producer—and, for that matter, to the total expenditure of the customer. Hence negotiated prices are largely confined to inter-firm sales—though not to all of these—and posted prices are characteristic of sales to personal consumers.[1]

The information available on comparative qualities is inherently less clear-cut than that on price. Some qualitative characteristics can be known by inspection or analysis before purchase; others can be known only through experience. The proportions of these two will vary from case to case. There is again a distinction between the industrial and the private customer. The former is, comparatively, an expert buyer, better equipped to exploit every possibility of pre-purchase judgment by inspection. But even for him there is a residuum of qualities which can be known only through experience. Even in the most favourable case, the reliability of the supplier in fulfilling his contract, as regards both what he supplies and when[2] he supplies it, can be discovered only by trial.

It is this issue of reliability, together with the fact that customers vary in their tastes, which accounts for the existence of goodwill; i.e. that firms have habitual customers, some at least of whom will not be detached from them at the first whisper of a more attractive offer elsewhere. This attachment of customers, which neo-classical theory has, for historical but unfortunate

[1] The manufacturer seldom sells directly to personal consumers, but to distributors. It might seem, therefore, that the condition of large and few sales is more nearly satisfied. But, in most cases, the manufacturer *feels* himself to be selling to consumers and regards the distributors simply as providing a necessary service (like transport) for which he pays through the distributive margin. And in this he is right, in so far as the distributor's willingness to buy will depend more on the speed of turnover than on the size of the margin. This attitude goes a long way towards explaining why manufacturers favour resale price maintenance and are profoundly irritated by any attempt on the part of the distributor, whose services they "buy", to take a hand in determining the final price of what they regard as still their product.

[2] In full employment conditions delivery dates assume great importance.

reasons, called an imperfection of competition, should rather be considered a normal feature of the market for manufactured goods. Indeed, it is difficult to see how without it manufacturing industry could avoid some of the extreme instability which chronically afflicts the market for primary commodities, in spite of the existence of professional speculators.

When we say that a firm is engaging in active selling we mean that it is trying to change the structure of values for money (price/quality combinations) known to customers in such a way that its offers will appear attractive to customers who previously went elsewhere. It is *changes*, therefore, which are important. Now, as I have said, price, being a quantity, is clear-cut and definite, whereas quality is—a quality, and so much less so. Hence we may expect that, if both are unmatched and equally known, a price-cut will win new customers more speedily and certainly than will an improvement in quality (real or imaginary) costing a similar amount. Conversely, however, an improvement in quality (or, by advertising, in what customers think they want) can be matched less quickly than can a change in price.

A firm which wishes to add more customers to its habitual circle will wish to *detach* them from others at the smallest cost and *attach* them to it as firmly as possible. Its success in this aim will be the greater the more it can confine and direct its selling effort. By confining its effort, I mean confining the change in its price/quality bundles to the new customers which it seeks, without giving any benefit to those it already has. By directing its effort, I mean concentrating on winning over the customers of a particular firm (or a few firms) which is in a weak position and so can be driven into bankruptcy (negative gross profits) or bought up at the end of the campaign.

Conditions for confining and directing competition are clearly most favourable in industries where negotiated prices are the rule. For discrimination between customers, and the ability to choose which to favour, is the essence of negotiated prices. Hence, we should expect to find such industries characterised by a rather continuous process of aggression by the efficient against the inefficient. The transfer mechanism will tend to work both smoothly and relatively swiftly.

When prices are posted, however, the scope for confined and directed competition will be much smaller. Discrimination between customers is not possible, so that any cut in price or improvement in quality has to be extended to the old as well as the new. Moreover, instead of the aggressive firm being able to direct its competition against the moribund, the effects of its every move will be felt by all firms producing the goods for which the aggressive firm is now offering better value for money. These limitations will be the greater the more standardised are the products in question and the more their qualities can be known by inspection—in traditional language, the more perfect is the market. For, if customers are relatively firmly attached to a particular supplier by the belief that it offers them something which other firms do not, there will be greater scope for any one of the other firms to focus its attack. We may expect, therefore, that in industries characterised by posted prices and little "imperfection" the efficient firm will be more reluctant to embark on aggressive selling policies than it will in negotiated-price industries. This does not mean that the transfer mechanism will not operate, but rather that it will tend to work more jerkily and that the rate at which the least efficient fringe is killed off will tend to be slower. There will be periods of relative tranquillity, with little change in the structure of values for money on offer, when the least efficient are quietly working their capacity to death. These will be broken when the accumulating profits of the efficient begin to burn holes in their pockets and they feel able and impelled to enter on a struggle for *lebensraum*. Periodic price-wars as they are called will, in fact, be characteristic of the posted-price industry, even if fundamental disequilibrium is avoided.

This brings us to the second of the problems I posed earlier. More particularly, we have to examine the behaviour of an industry when capacity has been created in excess of demand. For, as I said at the beginning of this chapter, the birth of capacity being a much quicker affair than its death, no special problem arises when the disequilibrium consists of a deficiency in capacity.

What is needed to eliminate excess capacity is not the extinction of firms but the sterilisation or scrapping of capacity, and

competition in selling (price-cutting) is much less effective in killing capacity than firms. As I have said, what is needed to kill a firm is a period of negative gross profits, or a good take-over bid from another. But what is needed if capacity is to be scrapped is that reasonable men should believe that under no future conditions which it is reasonable to envisage will it be possible to earn any positive gross profit by working the capacity. Such a view will usually be taken only if the capacity is either very decrepit or, technical innovation in the industry having been rapid, very old-fashioned. In other words, firms can be killed by prices, but capacity only by time.

We may see the implications of this by envisaging a rather extreme case in which, the industry being previously in equilibrium, there is a sudden fall, of say 20%, in demand, which thereafter remains perpetually at this level. It would then be necessary for prices ("values for money") to fall to a level at which the owners of 20% of the capacity of the industry were willing to hold it voluntarily inactive, and that these prices should persist for a period long enough to see the condemnation, by the criterion of the previous paragraph, of 20% of the industry's capacity. After that period prices could return to a level which made average net profits in the industry zero.

It is possible, in principle, to conceive of this as happening by prices falling to a level at which (the least efficient) firms accounting for 20% of the industry's capacity were earning negative gross profits and, in consequence, either closed down to wait for better times or went bankrupt. Those firms which stayed in business could then be working to capacity. Given that direct costs can be reduced considerably below their normal levels for substantial periods by the deferment of all postponable expenditures, the price necessary to bring this about might be very low indeed, and might entail an average net profit, year in year out, for the active part of the industry which would leave capacity seriously deficient when the period of scrapping arrived.

In practice, such extreme solutions are not encountered. Some of the least efficient firms do go out of business during a slump. But enough remain in operation for most firms to be working below capacity. Factories being planned to work at a particular

capacity (more strictly within a certain range), the effect of working below it will be to raise total costs per unit of output and, probably, direct costs also. We have to ask, therefore, what it is which prevents each firm from indulging in the aggressive selling to try fully to occupy its capacity which would, in the event, force prices down to the level suggested in the previous paragraph. The answer is to be found in the continued force of the notion of normal price, reinforced by the articulate fear of just such a self-defeating eventuality as has been described. But the notion of normality grows weaker as time goes on, and when firms are working from hand to mouth perspectives are likely to be short, and the fear that the last state will be worse than the first is always liable to be overborne by the chance (real or imagined) of securing a short-term advantage from price-cutting.[1] In short, in conditions of fundamental disequilibrium prices increasingly cease to have any firm base comparable to the notion of "normal" prices in "normal" times. They are unlikely to fall to the firm floor at which the excess capacity is closed down, but their position above that is inherently unstable.

I suggested above that the length of time needed to destroy any given excess of capacity will be largely independent of the course of prices. The slow process of time can only be hurried by technical innovation, which makes capacity old-fashioned before it becomes decrepit. But the rate of technical innovation is likely to be slowed rather than speeded by fundamental disequilibrium. Innovation involves expenditure and slumps are traditionally regarded as bad times in which to spend money. While this attitude is no doubt compounded of both reason and unreason, the desire to remain liquid at a time when all around you are going bankrupt has an obvious common sense. It seems, therefore, that the innovation mechanism as well as the transfer mechanism will tend to be suspended for the duration of the disequilibrium, and the duration of the disequilibrium will be the longer in consequence.

Capacity is built by firms to meet a demand which they hope to create and not for a queue of customers already waiting. The

[1] Price-cutting is likely to be favoured over quality improvement in these circumstances, because the latter probably involves increased expenditure and there is an understandable reluctance to throw good money after bad.

past experience which guides them can never be a certain pointer to the future. Hence there is a constant possibility of fundamental disequilibrium being created. Since the rate of decay of capacity is so slow it may be asked, therefore, whether industries will not have excess capacity almost continuously, and, since the transfer and innovation mechanisms are then suspended, whether the attention devoted to them is not beside the point. The answer is that excess capacity may disappear through the growth of demand as well as the extinction of capacity. So long as there is the general full employment which I am assuming throughout this book, it is reasonable to assume that the saving grace of growing demand will normally remedy fairly quickly the excesses of over-optimistic industries, and to treat the disequilibrium situation discussed in the second part of this chapter as an abnormal one. The less one's confidence in continuing general full employment the less legitimate my conclusion. A very qualified confidence is one of the main stimulants to rules of the game, which I now consider.

Chapter X

THE NATURE AND IMPLICATIONS OF RULES OF THE GAME

At the end of Chapter VII, I suggested that our enquiry should be directed towards three questions.

(1) Do certain rules of the game diminish the penalty for failure to innovate or the reward of success, where penalty and reward are in terms of relative growth?

(2) Do they restrict freedom to experiment with new ways of doing things?

(3) Do they influence the rate at which knowledge of new techniques is diffused?

In the course of Chapter VIII, I concluded that, even in the absence of rules of the game, concentrated (oligopolistic) industries would develop; that once concentration was established it would tend to persist; that both the transfer and innovation mechanisms would be partially inhibited in such industries; and that this inhibition would be compensated by colonisation of other (non-concentrated) industries. This analysis suggests three further questions.

(4) Do certain rules of the game tend to develop or preserve concentration?

(5) Do they interfere with the process of entry?

(6) Is there any positive means by which the process of concentration can be arrested or reversed?

We are now in a position to consider what answers should be given to these questions. While so doing, we must consider, in the light of the arguments of those who make rules of the game, whether any further questions need to be asked and answered before judgment on the rules can be made.

As I emphasised in Chapter I, it would be vain to expect incorrigible answers to these or other empirical questions, even if the factual information at our disposal were far greater than

it is. Economic events are always unique and the degree of similarity between any two is very debatable. We can, nevertheless, hope to advance some way from the Cleopatra's nose end of the spectrum of knowledge. To do so involves two activities. The first is to consider, from examination of the rules, whether there is a *prima facie* expectation that the answers to the questions just posed will be affirmative. The second is to examine empirical data to see whether the *prima facie* expectation is validated and, if so, whether the effects revealed are quantitatively significant. To repeat, we can never find *direct* empirical evidence relevant to the proposition, "Things would have been different in such and such a way had the rules of the game been different." What is needed is to show *either* that, following a change in its rules, the behaviour of an industry altered significantly, *or* that industries with different rules of the game behave differently, allowance being made in either case for differences in other factors which our analysis has suggested to be influential. This chapter is concerned with the *prima facie* indications. In the chapters which follow I examine the very incomplete factual evidence I have been able to find, and suggest some directions in which further empirical research should be profitable.

As regards the rules themselves, the analyst suffers from an *embarras de richesse*. The rules of Rugby Football are simple and compact compared with the plethora of regulations created by the typical Trade Association, and the Monopolies and Restrictive Practices Commission has made a generous sample available to the public in the bulky appendices to its reports. Given the differing circumstances of each trade, a precise classification of the rules revealed would require a book in itself. But the great majority can, without too Procrustean violence, be grouped under the following four broad heads:

(1) Rules limiting the freedom of the individual producer to change the price/quality combination (value for money) which he offers to his customers. These are the most common of all rules. They range in severity from a mere obligation to report in advance the intention of varying price, through agreement not to cut prices below some minimum level, to the imposition of a common selling price for all producers in the industry.

Quality specifications and conditions of sale are laid down, in the degree of detail necessary to make a reality of the obligations accepted by the firms party to the agreement.

(2) Rules laying down the proportions of the total custom of an industry which shall be taken by each particular firm. In the extreme cases such agreements are satisfied currently by firms, where necessary, passing orders from one to another. In other cases the reality of the rules lies in the provision for payment after the event by firms which exceed their quotas to those who fall short. Quotas may be fixed rigidly for a period of years or may be adjusted annually, by some established formula, in the light of actual sales.

(3) Rules providing for the exchange of technical information between the firms in an industry.

(4) Rules providing for concerted action to enforce adherence to the preceding rules. The most common instrument of enforcement is the denial of supplies, either absolutely or on customary terms, to customers (usually distributors) who trade with a firm which is not a member in good standing of the rule-making association. In some cases, however, more active methods are employed; special companies[1] may be formed to indulge in intensive directed competition[2] against an offender or outsider; or, where these are controlled by compliant firms, essential supplies may be denied the rule-breaker.

The particular combination of these various heads, and the forms of rules adopted under each head, will tend to vary with what I shall term the strength of the association. I draw this distinction to mark the fact that agreements vary considerably with regard both to their durability and to the circumstances in which they are conceived. There are three main types:

(1) The excess-capacity agreement. By this I mean a set of rules which are agreed upon during a slump, and do not long survive the disappearance of the excess capacity which constitutes the slump.

(2) The unstable agreement. The history of certain industries

[1] These are usually termed "fighting companies"; their purpose is to undersell the outlaw (making a loss if necessary) in the particular lines which he is offering.

[2] See Chapter IX.

is characterised by the constant making, breaking, and remaking of agreements on rules of the game. In a typical case-history an agreement is made to observe minimum or common prices, which may sometimes be bolstered by agreement on market shares. The rules are observed for a time, but the machinery of enforcement is rather poorly developed and there are discreet departures from them as firms see a chance to prosper thereby. Such departures become more and more blatant until the agreement collapses in mutual recrimination. There is then an interregnum during which no holds are barred, which is brought to an end by the formation of a new agreement. The cycle then repeats itself.

(3) The entrenched agreement. Unstable or excess-capacity agreements sometimes acquire a quality of persistence; the customary collapse of the rules does not take place. In such cases the extent and severity of the rules, and the effectiveness of the rules regarding enforcement, tend to increase with their duration. It is usually only in such entrenched agreements that type (3) rules—providing for technical collaboration—are found.

It follows from the discussion in the previous chapter that there is little reason to condemn excess-capacity agreements. I there suggested that when there is a fundamental excess of capacity the transfer and innovation mechanisms will, in any event, be largely suspended. I suggested, further, that the only firm floor to prices in such a situation—the level necessary to keep firms owning $x\%$ of the industry's capacity out of the market—may be well below the level which, *ex post*, would be seen to have been necessary to restore eventual equilibrium. Hence an agreement which gives more stability to prices, as they float like Mahomet's coffin above this floor, is at least as likely to be a good as to be a bad thing.

This neutral attitude does not, however, cover attempts to anticipate excess-capacity situations. It is only natural that industries which have experienced fundamental disequilibrium in the past should wish to have machinery in being which, should this recur, will prevent the unpleasantness which preceded the establishment of agreement on the last occasion. But, given my conclusion below, that entrenched agreements are likely to be against the public interest, such anticipation also stands condemned.

Either the rules are inoperative so long as the industry is in equilibrium, in which case there is no assurance that they will be effective if and when need should arise, or they are operative, in which case they must be characterised as entrenched agreements. Many of the entrenched agreements which exist today came into being as excess-capacity agreements. Their persistence, and the honest conviction with which the participants defend their necessity, reflect a belief, which I have already suggested is unwarranted, that the conditions of the inter-war period are a better guide to "normal" economic conditions in the future than is our experience so far in the post-war period.

Consider now the unstable agreement. Any form of price agreement requires that the industry establishing it shall be, by and large, a posted-price industry. It is further necessary that there shall be a fair degree of standardisation in the products of the industry, since the task of drawing up lists of agreed prices would otherwise be impossibly complicated.[1] In other words, the conditions which make price-agreements possible are also those which I earlier suggested are likely to lead to periodic price-wars, in which the general level of prices may fall temporarily considerably below the equilibrium level. Now the difficulty with a price-war is to bring it to an end, even when the weaker brethren, whose liquidation was its essential object, have in fact been removed. No one wishes to be the first to raise prices when the speed with which others will follow is a matter for conjecture, so that prices may continue to languish below their equilibrium level. It is plausible, therefore, to regard some agreements as directed to, and necessary for, restoring "normal" (equilibrium) prices, and we may regard these, not as an impediment to the transfer mechanism, but as a necessary part of it in industries of this kind.

But if the agreement is not ultimately to become an impediment to the transfer mechanism, and is to be what I have called an unstable agreement, the peace which it establishes must be limited in duration. If the transfer mechanism is to operate there must be periods of war as well as of peace. Agreements break down when there is dissension as to what prices should be, or

[1] This is one reason for the existence of agreements as to types and qualities of goods to be produced in industries with well-entrenched agreements.

more or less covert breaches of the rules. And such dissension or rule-breaking develops when some firms develop an interest in prices coming down. Some firms (the more efficient) will begin to develop such an interest if they are accumulating profits which can be transmuted into new capacity only in their own industry, and that capacity can be supplied with customers only at the expense of other firms. It is, therefore, a condition for an agreement to be unstable that there should be no significant opportunity for firms in the industry to establish new branches in other industries. And the condition for this is, in turn, that the industry should not have achieved any significant degree of concentration.[1]

I think that this last condition will normally be sufficient to secure the instability of price and output-sharing agreements.[2] In formal terms, the requirement is that a stable price shall be incompatible with a stable average rate of profit in the industry and this requirement will be met if the average cost of production in the industry is being lowered through the relative expansion of the more efficient firms. In more homely language, efficient firms will break loose from an agreement if it begins to be a serious bar to their expansion, and less efficient firms will revolt if the agreement is operated so as simply to be a smoother mechanism for their extinction. So long as the expansion of the efficient involves the contraction or extinction of the inefficient, there is a conflict of interest which no verbal agreement can long extinguish, and which can be resolved only if there is a safety valve for the excess profits[3] of the efficient to escape to other industries. In short, it is internal pressure which breaks most agreements; lack of opportunity for intrusion into other industries rather than intrusion by them.

If the preceding analysis is correct it follows that it is in concentrated industries that we shall find entrenched agreements. Such industries have the necessary safety valve and there is a basis, therefore, for community of interest between the less and the more efficient firms in the industry. The more efficient can maintain excess profits, which would otherwise be eaten away

[1] See Chapter VIII.

[2] It is not a necessary condition since unstable agreements are also found in concentrated industries.

[3] See Chapter VIII.

by their very success in further expansion in their own industry, and these provide the sinews for war in the more rewarding fields of other, less concentrated, industries. Moreover, it is comforting to feel that, should hard times fall on their industry, there exists a tail of less efficient firms on whom, in the last resort, the main impact can be made to fall. The less efficient will be glad to exchange an arduous, and perhaps losing, battle for survival for a tolerated stability, which carries with it as *douceur* the chance to fight another day. Indeed, as I have argued in Chapter VIII, the check to the transfer mechanism which imperialism makes possible, and which is necessary if agreements are to become entrenched, develops quasi-automatically, even without an agreement, as concentration grows. The purpose of an agreement is, therefore, to strengthen tendencies which are already present, and it is from these that the agreement gains its durability.[1]

An agreement, together with its attendant enforcement machinery, may strengthen the tendencies which gave them birth in two senses. In the first place, if durability is intended, the parties to the agreement must intend that it shall contribute to a policy of live and let live within the industry, and there is no reason to doubt their belief that it will. Formal agreement is clearly a firmer basis than tacit consent—even though business morality seems to be peculiarly lax about breaking such undertakings—and once machinery for enforcement is established the chance of a successful break-through by any individual firm is greatly reduced. But the agreement contributes in a second way, which is not necessarily always foreseen when it is made. Any organisation, once established, tends to take on a life of its own; the instrument which serves passively the purpose of its makers is found only in the engineering world; the performance of social instruments almost always differs from that envisaged by their creators, and, in particular, their scope is usually wider. The mere practice of co-operation is habit-forming, to such a degree that its absence comes soon to be regarded as abnormal.[2] Moreover,

[1] It is also true, of course, that the sheer mechanics of organising and policing an agreement are simpler the fewer are the firms involved.

[2] The most obvious reason for the sincere indignation with which the makers of restrictive practices greet any attack upon them is that they have lived with the practices so long that they cannot envisage life without them.

it is evident in any walk of life that rules are, so to speak, self-proliferating. Once a rule is made—say against price-cutting—the ingenious and ambitious firm will seek a way of violating its spirit with advantage which leaves its letter sufficiently intact to nullify any prosecution. The natural reaction of other parties to the agreement is to make a further rule—perhaps limiting advertising, specifying qualities more closely, or laying down market shares—which will prevent a repetition of the contravention. And so, through its own momentum, the apparatus of rules extends in scope and severity.

If my argument in Chapter VIII is correct, this reinforcement of the tendency for relative market shares to be roughly stabilised, which is the object and result of entrenched restrictive agreements, will equally accentuate the weakening of the innovation mechanism which characterises the concentrated industries in which such agreements are found. The net incentive to innovation will be diminished. And this for a less obvious reason also. What helps to keep innovation alive in concentrated industries is the hope (and fear) of some innovation which will be so great (and can be guarded so long) as to break down the resistance of other firms. Live and let live is a policy to be abandoned at the prospect of a quick and safe killing. But one of the commonest types of innovation consists of a firm's adopting a less diversified production objective which enables it to take advantage of a more mechanised (lower-cost) technique. And the condition for its success is that, by underselling other less specialised producers, it should be able to persuade away from them their customers for the more limited range of goods which it now produces. In other words, a change in *relative* prices may be necessary if the innovation is to succeed, and it is likely to be as difficult to win agreement for this as for a change in the general price level.

One further way in which entrenched agreements nourish the roots from which they spring rests on the way in which the machinery of enforcement is operated. The original purpose of such machinery is to deny advantages or impose positive pains on any firm which breaks the rules. But it may equally well be used, if a majority of the parties are so agreed, against firms which might be quite prepared to abide by the rules were they allowed.

And it is reasonable to suppose that it will be so used against firms from other industries (or new independent firms) which seek to enter the industry. In the eyes of established firms the new entrant is always the bearer of excess capacity, as he is in truth, unless total demand be expanding very fast. Hence it is easy for them to convince themselves that the public interest is at one with private interest in requiring action to block his entry. And this means that the irreversibility which I said was characteristic of the process of concentration will be strengthened all the more.

My conclusion so far, therefore, is that, while excess-capacity or unstable agreements may well be justified, the entrenched agreement stands condemned as yielding affirmative answers to most of the questions which I posed at the beginning of this chapter. Although itself the child of concentration, it may help to create the conditions characteristic of concentration—partial atrophy of the transfer and innovation mechanisms—earlier than they would otherwise have arisen, and will certainly tend to strengthen them. By erecting obstacles to the entry of new firms into the industry it makes concentration more irreversible. And if the agreement develops to the point where there is interchange of technical information it removes the barriers to the diffusion of innovation which I have taken to be one of the conditions for innovation to proceed at a fast rate.

It is now time to examine the justifications advanced by the rule-makers. They fall into two classes; those which accept the assumption, which I have made throughout this book, that a market economy must have characteristics broadly similar to those which I set out in Chapter V; and those which, in effect, deny the efficacy of the market mechanism.

Arguments of the first class are all essentially variants of the same theme; that it is only by agreement amongst producers that prices can be prevented from resting, quasi-perpetually, at the level to which I have suggested they may fall during periods of excess capacity or in the course of a price-war in a posted-price industry. In the majority of cases, such arguments depend for their validity on the assumptions that periods of excess capacity will be far more frequent than I believe is likely when there is

general full employment, or that price-wars will be unduly destructive unless arrangements to avoid them are made in advance, an argument which I have already dismissed. There is, however, one more complex thread in the contention.

On this line of argument the producers stand forward as defenders of the customer against his own incapacity. We are told that the customer is incapable of detecting debasement of quality and that, in the absence of rules against price-cutting, the few unscrupulous producers to be found in every trade will, therefore, seek to prosper at the expense of their fellows by cutting price and quality simultaneously. Their fellows being forced to follow suit in self-defence, the end result will be a general debasement of quality.

It is at first sight curious to find this negative justification combined with the positive contention that, if prices cannot be cut, sellers' competition will operate by each producer trying to offer better value for the same money, so that quality steadily improves. But we need not dismiss this allegation, that the customer is incapable of detecting debasement of quality but is keenly appreciative of an improvement, as merely a confused or unscrupulous attempt to have it both ways. I earlier drew a distinction between aspects of quality which could be known by inspection and those which could be known only through experience, and suggested that the latter explained the existence of goodwill. If a firm is willing to risk its goodwill it can secure a temporary advantage by debasing quality in a way which is not apparent to inspection and will, therefore, be discovered by the customer only after some time. Conversely, improvements in quality, if not immediately obvious, may at first have little effect in detaching customers from other suppliers but, if genuine, will operate more powerfully as time goes on. In these senses, therefore, the customer is both insensitive and sensitive to quality changes.

Such a reconciliation of the two contentions, however, robs this line of argument of any new force. Goodwill, once lost, is difficult to recover, and a firm is unlikely to risk its goodwill by concealed quality debasement unless this be the only alternative to extinction. And no significant number of firms will be in a

position where dishonour is the only escape from death unless the industry is in a state of fundamental disequilibrium. In other words, the threat of a cumulative debasement of quality does not arise except when there is substantial excess capacity, a situation with which I have dealt already.

The second class of arguments—those which, in effect, deny the efficacy of the market mechanism—stands on two legs. The first is a transformation of the contention just considered, that the customer can sometimes be tricked in the short run, into a thoroughgoing denial of his ability to know quality when he sees or uses it.[1] The second is the proposition that innovation proceeds faster, less wastefully, and generally more satisfactorily if there is co-operation between firms, in which discoveries are shared and overlapping is avoided. The corollary of the first proposition is that quality specifications should be laid down and enforced by the industry, and if this stage of co-operation is reached it is only natural that there should be agreement on prices also. When the technical inter-change involved by the second proposition is in force it becomes difficult to see in what sense, other than a legal or book-keeping sense, the firms constituting the industry retain any individuality. The hard concept of an independent self-seeking entity has melted away, leaving us with an industry touched but lightly with nebulous lines of internal demarcation.

The attitudes displayed in the arguments just deployed find their reflection in two charges which are frequently made against British industries; that the British firm insists on giving the customer what it thinks he ought to have rather than what he wants; and that British industry is eminent in technological discovery but slow in application. The gossip of the market-place should not be given more attention than it deserves. But it is valuable in this instance in shedding light on the misconceptions on which these two defences of restrictive practices are based.

The instinct of workmanship is an admirable thing, and so is collective pride by firms in the good standing of their industry. But it is not desirable that private firms should set themselves up as defenders of the public interest, when private interest and

[1] This was the argument of the electric cable makers in the United Kingdom.

public good are so easy to confuse. A private enterprise system rests on the assumption that truth will out, sooner rather than later. If this assumption cannot be validated, with some occasional help from the government in laying down standards or assisting in testing, the logical implication is not that producers should have collective power but that private enterprise is the wrong form of organisation.

The defence based on the superior power of technological co-operation rests on a fundamental misunderstanding of the nature of the innovative process; on a confusion of innovation with research. There is ample evidence that there are economies of scale in scientific research, and it is obvious that, in one sense, it is wasteful for different research staffs to be conducting identical experiments. But, as I have emphasised earlier, the ultimate test of successful innovation is industrial practice. Since the translation of an idea from the research laboratory into industrial practice is an expensive and risky business, there is a strong and natural inclination to delay it until every endeavour has been made to foresee and eliminate possible snags. Unless, therefore, this inclination is counterbalanced by some strong compulsion, the practical development of ideas will be slow. This compulsion is normally supplied by the fear that others will get in first. Since technical co-operation removes the fear, the net result will usually be a slower rate of innovation. In short, we must regard the waste of unsuccessful ("premature") innovation as a necessary part of the machinery of technical progress. Attempts to avoid the waste will be at the expense of the functioning of the machinery.

I see no reason, therefore, to retract my earlier conclusion that the effect of entrenched agreements will usually be harmful. We must now address ourselves to empirical material to see whether this conclusion finds confirmation and whether the harm is quantitatively significant or, alternatively, the whole issue we are considering must be dismissed as a storm in a teacup.

PART II: PRACTICE

Chapter XI

SUMMARY OF THE EMPIRICAL INVESTIGATION

In the theoretical chapters I have constructed a crude model of the competitive process. I have developed a set of concepts and arrived at conclusions about their interrelations on the basis of only very general assumptions about the nature of economic motivations and limitations. This analysis has suggested that concentration and entrenched restrictive agreements will usually interfere with the functioning of the system in an undesirable way. I gave it as one of my objects in Chapter I, however, that the model constructed should be capable of empirical verification, of development through the interplay of theory and practice. The chapters which follow are an attempt to discharge this commitment.

The statistical tools employed are very simple, being no more than the standard deviation and the co-efficients of correlation and regression. Furthermore, at the risk of irritating the professional, I have been at some pains to explain what I understand to be the meaning of the operations performed. Nevertheless, I am conscious that some non-statistical readers may regard these chapters as laborious and not very rewarding. There is a further point. The empirical material I have used, although the best which I could bring together in the time available, is by no means ideal for my purpose, and could be improved on considerably, given time and ingenuity. Similarly, more powerful statistical tools may be found to be applicable by students of greater statistical competence. This part of the book is to be regarded, therefore, more as an illustration of the kind of empirical approach which the theory suggests than as anything in the nature of a conclusive demonstration of the validity of its conclusions.

For both these reasons, I have thought it appropriate to summarise in this chapter the results which are elaborated at greater

length in those which follow. Those who are not interested in the technique of testing how well empirical facts can be cut to clothe a theoretical model can confine themselves to this chapter and Chapter XVII. But they are warned that in so doing they will ignore factual material some of which has, I think, intrinsic interest.

Chapter XII examines the basic bricks from which the model is built—the concepts of an industry and of efficiency. It is shown that when the Census classification is imposed on the universe of manufacturing industry the industries so defined are sufficiently clear-cut to make the industry a usable concept. There is fuzziness at the boundaries because some firms are in more than one industry, but not so much as to destroy the concept and force us to regard manufacturing industry as one continuous spectrum which cannot usefully be differentiated. As regards efficiency, there is convincing evidence that the differences in efficiency which appear from examination of data on profit rates for any one year are not the product of random and transient factors, but have sufficient persistence to provide the transfer mechanism with something to bite on and us with a problem worth investigating.

Chapter XIII describes the empirical material. This consists of Census of production data, since the Census is the most accessible source which covers a considerable span of years. The theory deals with growth, efficiency and changes in efficiency. The Census yields information on changes in size measured by net output (net output=wages and salaries plus gross profit) and on productivity (net output per person employed) and changes in it. The conditions under which productivity can be taken as a reasonable index of efficiency as I have defined it are described.

Chapter XIV establishes that the dispersion of efficiency is very substantial in most industries and that there is no systematic tendency for the dispersion to narrow over time. There are substantial differences in the degree of dispersion from industry to industry. And the theoretical expectation that industries with long tails of unassimilated high cost firms are likely to be ones in which concentration or restrictive agreement have impaired the transfer mechanism is shown to have some reflection in reality.

But by and large, there is no general tendency for the degree of dispersion to vary with the degree of competition.

Chapter XV investigates the transfer mechanism. The minimum evidence needed to impart empirical validity to the existence of a transfer mechanism is that there should be substantial changes over time in the relative sizes of the different firms in an industry. This evidence is amply supplied by the material examined. The second need is to show that growth is not randomly distributed, but that it is the most efficient firms which grow the fastest. Since productivity is likely frequently to be a poor index of efficiency in individual cases, we should not expect to find that rapid growth is correlated with high productivity at the beginning or end of the period considered. But it is shown that those firms which have *increased* their productivity most have also grown the fastest. Since increases in productivity are probably reasonable evidence of increases in efficiency, this goes some way to confirming the theoretical explanations, and reasons are given for believing that it is the increase in efficiency which causes the increase in size and not *vice versa*. There is, however, only very weak confirmation for the theory's conclusion that the transfer mechanism is likely to be impaired in industries which are highly concentrated or have entrenched restrictive agreements.

Finally, in Chapter XVI, the innovation mechanism is examined. There is ample evidence that relative productivity is not constant over time. The firm with the highest productivity in 1935 is not, in the typical case, found in the first rank in 1948. There is, in fact, a shift from the bottom towards the top and *vice versa*. I have already remarked on the obstacles to taking productivity as an index of efficiency for the individual firm. None the less the empirical evidence can be regarded as providing some support for the existence of an innovation mechanism of the kind postulated by the theory. There is, moreover, some evidence that the mechanism works more powerfully in those industries which are least concentrated and freest from restrictions.

All this may be summed up by saying that the very imperfect empirical evidence is not inconsistent with the theory and lends qualified support to certain parts of it.

Chapter XII

THE VALIDITY OF THE CONCEPTS

THE purpose of this chapter is to demonstrate that the concepts of an industry and of the efficiency of a firm, in terms of which the preceding analysis has been conducted, are usable tools.

I have defined an industry as, roughly speaking, the group of firms engaged in a common process of production which the Census of Production has until recently called a trade.[1] When I say that the concept is valid or usable I mean that it is possible to classify the firms operating in the United Kingdom[2] into trades, without overmuch duplication—in the sense that one firm appears in a considerable number of trades—or too much heterogeneity—in the sense that the activities in terms of which an industry is defined are only very partially representative of the activities of the firms classified to it. In other words, most industry boundaries are genuine boundaries, which only few firms straddle. Since I have treated the case of the firm established in more than one trade as exceptional,[3] the analysis would need considerable recasting if this were not true.

The basic unit for Census purposes is the establishment (plant) and not the firm, and only a limited amount of work has been done on linking establishments with firms.[4] Hence a definitive proof of my contention cannot yet be given. But if we can show, first, that the Census classification of establishments is free from duplication and significant heterogeneity, and, second, that it is the exception rather than the rule for a firm to own establishments in more than one trade, we may regard the point as sufficiently established for present purposes.

[1] In the reports on the Census for 1954 the terminology has been changed and what was formerly called a trade is there called an industry.

[2] Excluding, of course, those engaged in distribution and other activities which fall outside the scope of the Census of Production and this study.

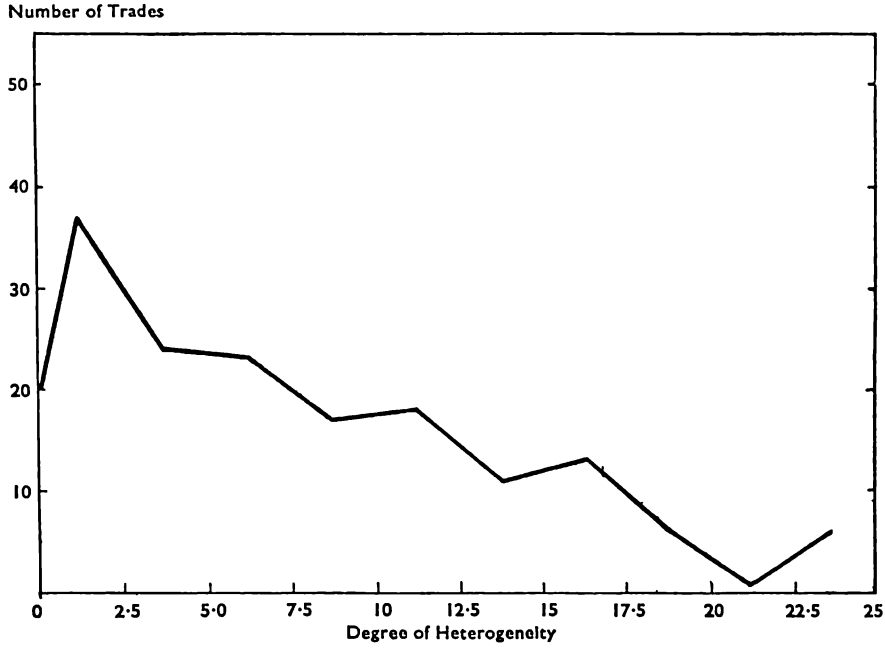
[3] See Chapter VIII.

[4] Notably, Leak and Maizels, *The Structure of British Industry*, Journal of the Royal Statistical Society, Vol. CVIII—1945 (New Series).

Freedom from duplication in the classification of establishments[1] is guaranteed by definition. An establishment is classified to one trade and one trade only. Heterogeneity cannot be banished so easily, but every effort is made to do so. The maximisation of homogeneity is one of the main criteria used in determining the definitions of industries and provides a supplementary criterion to that quoted in the footnote for deciding what is a single establishment.[2] The degree to which this aim is frustrated by the practical difficulty of unscrambling intertwined or undifferentiated accounts can be seen from Chart III. I have taken as an index of heterogeneity the percentage of the sales of an

Chart III

FREQUENCY DISTRIBUTION OF CENSUS TRADES BY DEGREE OF HETEROGENEITY



[1] "In the majority of cases an establishment comprises the whole of the premises under the same ownership or management at a particular address (e.g. a mine or factory)." *Final Report on the Census of Production for 1948, Introductory Notes*, London, H.M.S.O., 1951.

[2] "Where two or more distinct trades were carried on in separate departments of a single works, the firm was generally required to treat these as separate establishments and make a separate return for each department on the appropriate form." *Ibid.*

industry which consists of products which fall outside its definition, calculated this for each Census industry, and presented the results in terms of a frequency distribution.

The degree of heterogeneity is typically very small. In nearly two-fifths of the 156 industries with which the Census deals the degree of heterogeneity is less than 5%, and over four-fifths of the industries fall within a range of 15%. The weighted average degree of heterogeneity for all industries is less than 7%.[1]

So much for the classification of establishments. When we try to estimate the proportion of firms which own establishments in more than one trade we step on to more shaky ground. The main source of information is the work of Leak and Maizels on the results of the 1935 Census of Production, to which I have already referred, and the greater part of their analysis did not deal with firms employing less than 500 workers.[2] For all firms larger than this they presented a frequency distribution showing, for a number of employment size-classes, the relative importance of single-trade and multi-trade firms. This distribution is summarised in Chart IV.

It will be seen that the smaller is the firm the less likely it is to be a multi-trade firm, and it seems reasonable to infer from the line of (visually) closest fit which I have drawn through the plotted points that the single-trade firm must be virtually the rule in the large size-class below 500 employees on which we have no information. Hence, we may take it that Chart IV includes almost all multi-trade firms. If we add in the holding companies, which I excluded from the Chart for lack of comparable information by size-classes, we find that 68% of all workers in firms having more than 500 employees are employed by firms which operate in more than one trade. But firms employing more than 500 workers accounted for only slightly less than one-half of total employment in Census Trades. We may conclude, there-

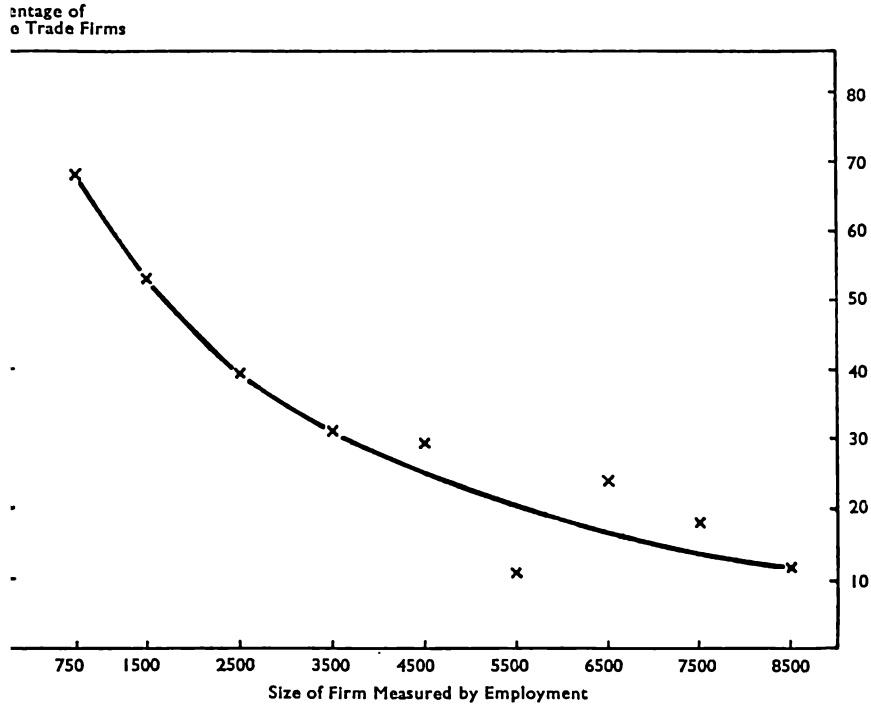
[1] This is an average weighted by sales. The use of sales for defining heterogeneity is open to criticism, but the nature of the Census material precludes any other measure. In any case the degree of duplication in the estimates of sales is insignificant for most trades. Estimates are given in Table 5 of Part II of the summary tables for the Census of Production for 1951.

[2] I mean by a firm what Leak and Maizels called a unit; namely the aggregate of establishments owned or controlled by a single company, where control means ownership or more than one-half of the voting power in a concern.

re, that multi-trade firms account for only about one-third of tal employment. In terms of net output the proportion would : somewhat higher; in terms of numbers of firms it would, of ›urse, be very much lower.[1]

Chart IV

RELATIVE IMPORTANCE OF SINGLE TRADE FIRMS IN DIFFERENT FIRM SIZE CLASSES



A proportion of one-third is not insignificant. What this means, owever, is not that the industry concept is invalid but that the ensus classification is not precisely right for our purpose. A conderable proportion of the multi-trade firms are undoubtedly

[1] A rough check on this last point was made from the British Company Index ublished by Moody's Services Ltd. This covers a wider range of activity than the ensus and distinguishes some 240 industries. The index gives the number of the age or pages on which any Company is listed. Since some pages cover more than ne industry, a single page reference may not entail that the Company is in only one idustry, but in most cases it will. Three pages were chosen at random from the idex and it was found that of the 361 Companies listed, 295 had only one page eference, 56 had two, 8 had three, and only 2 had more than three. Even after the ualification just mentioned, it seems legitimate to infer that most companies are

cases of vertical integration for technological reasons. If the pairs of trades where such integration is common were combined, as would be natural if the Census grid were designed to classify firms rather than establishments, the proportion of multi-trade firms would be much smaller. I conclude, therefore, that there is justification for assuming, as I have throughout this book, that the typical firm is engaged in one industry only and that its history depends on its behaviour in that line of business and on that of other firms which, in the main, will also be engaged in that line of business only.

It is convenient to pursue the matter a little further at this point and to consider what evidence exists for assuming that the (exceptional) case where firms migrate across industry boundaries accords with the pattern which I suggested in Chapter VIII. My suggestions were that, except where there were technological advantages in integration, only large firms would migrate across industry boundaries ("conglomerate"); that even they would not usually do so, unless their original industry were becoming concentrated; and that their migration would be confined within a technological horizon, whose radius would be the greater the larger the firm.

Lack of published data again precludes any definitive validation of these hypotheses. Migration is a process in time and its character can be determined adequately only by data which cover a period of years. Such data are available to the Census authorities, since we have now had a succession of Census enquiries on largely identical lines. But the provisions against disclosure of information relating to individual firms make it impossible for such analysis to be carried out by private research workers. It is, therefore, much to be hoped that the Census office will find itself able to do the job.[1] In the meantime, however, we may regard the points as partially validated if we can show that it is, typically, the large firm which spreads across

[1] As a minimum, one would like to know how the growth of net output in different industries between two Census years is divided between firms which were already established in the industry at the first date, firms which have entered the industry since then, and (as a negative component) firms which have disappeared from the industry since the first date. This could be supplemented by a classification of the new firms into "independent" and "dependent" and an analysis of the provenance of the latter.

industry boundaries; that the multi-trade firm has a centre of gravity, in the sense that one or two trades account for a large part of its total output (for it is then reasonable to infer that the flow has been from the centre to the marginal trades); that it is concentrated industries which provide such centres of gravity; and that the trades into which a firm migrates are, technologically, neighbours of its industry of birth.

The data already presented in Chart IV provide convincing evidence that large firms are more prone to migration than are small. Their source, however, does not provide any information on either the number or the nature of the industries in which multi-trade firms are engaged. As a first resort, therefore, I sought supplementary evidence in the Stock Exchange Official Year Book and Company Reports. A sample[1] was drawn, from the Iron, Coal and Steel and Industrial and Commercial Sections of the Year Book for 1955, of firms in different size classes, and an estimate was made of the number of trades in which each was engaged, on the basis of information in the Year Book supplemented in some cases by Chairmen's speeches. The measure of size was the nominal capital of the company. The results are presented in Table 2.

Table 2

INDUSTRIAL DIVERSITY OF MEDIUM AND LARGE FIRMS

	1-3	4-6	9-11	Over 20
1. Nominal capital, £m. .				
2. No. of firms in sample .	23	21	18	18
3. Average No. of trades in which firms involved .	2·73	3·24	3·5	5·94
4. Average No. of orders in which firms involved .	1·87	1·95	2·28	3·11

So far as they go, these data support the hypotheses. Diversity of activity is shown to be a function of size (line 3). Moreover, if, as seems reasonable, we assume that the trades within an

[1] Samples of 1 in 20 and 1 in 3 were drawn for the smallest and next smallest size classes respectively; for the upper two classes the sample included all firms. After elimination of subsidiaries and firms engaged in activities falling outside the Census of Production the numbers in the samples were as shown in Table 2.

Order of the Standard Industrial Classification have a greater technological affinity with each other than with trades in other Orders, line 4 of the table suggests that there are technological horizons and that their breadth increases with the size of the firm.

But this evidence is clearly very defective. Nominal capital is a poor measure of size of firm. A good deal of judgment was involved in deciding, from the published evidence, how many trades a firm was engaged in. And perhaps most important, no systematic idea could be formed of the relative importance of the different trades in which a firm was engaged. It is fairly certain that essentially ancillary activities—such, for example, as the building work carried out in its own plant by a firm in the steel industry—were excluded by the nature of the basic material. But there can be little doubt that Table 2 exaggerates the diversity of firms' activities, by including trades in which a firm's interest is so small as to be insignificant.

I decided, therefore, to make a direct approach to the above firms, requesting information on their size (in terms of value of assets), the number of Census trades in respect of which they completed returns, the relative importance of each trade (in terms of sales or net output), and the number of new trades which they had entered since 1948. A letter embodying these four questions—each of which could be answered by reference to completed Census returns—was sent to the eighty firms. Twenty-five complete replies were received; six firms sent reports and accounts in lieu; two firms refused their co-operation; and the rest neglected the courtesy of a reply or acknowledgment.[1] The resultant sample was, therefore, too small to justify any elaborate analysis or to provide firm assurance that conclusions derived from it are truly representative of industry as a whole. The main results are summarised in Tables 3 and 4.

It will be noticed that I have dealt in terms of the number of trades or orders required to account for a given proportion of a firm's output rather than the number of trades in which a

[1] The suspicion of or lack of interest in economic research which this response denotes can be paralleled, though perhaps not equalled, by the experience of many other research workers. To the frequent complaint of British business-men that their intentions or practice are misunderstood, one is tempted to reply that they have only themselves to blame.

Table 3

INDUSTRIAL DIVERSITY BY SIZE OF FIRM

1. Size of class (£m., total assets) . . .	2·5-5·0	5·0-10·0	10·0-25·0	Over 25·0
2. Number of firms . .	5	2	8	10
3. Percentage of total assets of all public companies in class . . .	2	2	11	53
4. No. of trades required to account for 90% of output . . .	1·8	1·0	1·71	2·22

Table 4

FREQUENCY DISTRIBUTION OF FIRMS BY NUMBER OF TRADES/ORDERS ACCOUNTING FOR 80%/90% OF TOTAL OUTPUT

Number of trades/orders	*Number of firms*	
	Trades	*Orders*
1	12	13
2	6	9
3	3	3
4	3	–
5	–	–
6	1	–

firm is engaged. This procedure was adopted because the replies showed that, in the typical case, a firm engages in a number of activities which represent so small a proportion of its total output or turnover that it is clear that it is producing primarily to meet its own needs. Such lines of business are obviously not potential avenues for growth, alternative to the main lines, of the kind considered in connection with conglomeration,[1] but are minor forms of integration in which convenience is frequently the motive. Thus, Imperial Chemical Industries Ltd. are engaged

[1] See Chapter VIII.

in twenty-six trades and eleven orders. But five of the orders taken together account for less than 2 per cent. of total turnover, and in only eight trades is output greater than 2% of the total.

Even after this correction has been made, there is reason to suppose that Tables 3 and 4 exaggerate the extent and frequency of *conglomeration*. A number of the firms in the sample are engaged in steel-making, and the diversity of their activity reflects the technological advantage of integration between blast furnaces and steel melting and rolling plants. Moreover, it is possible that one reason for the poor response to the questionnaire in the lower size groups was the belief by single-industry firms that lack of diversity would rob their replies of any interest.[1]

The data—which, be it remembered, are most representative of the largest firms—can be regarded as consistent with the view that the firm which sprawls over a multiplicity of different industries is quite exceptional, and that even the largest companies have firm centres of gravity in a very few trades. By and large, firms stick fairly closely to their accustomed lasts, and when they pry themselves loose their wanderings do not usually extend beyond the order. The sample is unfortunately too scanty to provide any evidence on my suggestion that the motive force in the process of conglomeration comes from industries with a relatively high degree of concentration. Nor have I been able to find any other published material directly bearing on this point. In so far as concentrated industries are characterised by large firms, the evidence already examined suggests that concentrated industries will produce a higher proportion of migrants than will non-concentrated. But the supplementary proposition—that large firms in non-concentrated industries will be less prone to migration than their fellows in concentrated industries—must remain at present not-proven.

I turn now to efficiency. I defined this in Chapter III as $\epsilon = 1 - \beta\,(r - \bar{r})$. If $r = \bar{r}$ continuously for all firms, then $\epsilon = 1$ for every firm. There would then be no dispersion of efficiency, and the problem with which this book has largely been concerned would have no existence in fact. Cursory inspection of the most

[1] The questionnaire was headed, perhaps unwisely, "Diversity of Activity of Large Firms".

elementary information on company accounts is sufficient to establish that individual r's are not identical in an industry during any particular (short) period of time, and it is unlikely, to say the least, that variations in β will be such as exactly to offset variations in r. But this is not enough to establish the reality of the problem with which we have been concerned or the usefulness of ϵ. It is worth troubling ourselves about the distribution of business between firms of greater or lesser efficiency only if there is something enduring about efficiency. Suppose, for example, that firms with small ϵ's this year had large ϵ's next year and, more generally, that over any given (fairly short) period of years the average was identical for every firm. We should then feel that it did not matter what was the relative size of any particular firm. For growth takes time—at the very least the time taken to build new capacity or organise the transfer of its control—and, in the conditions envisaged, the justification for such transfer would have disappeared before it had been effected. If reality were like this, our enquiry would be directed to discovering the factors responsible for the constant changes of rank rather than to the nature of the connection between rank and growth.

I felt it necessary, therefore, to satisfy myself that ϵ is a reasonably stable magnitude over short periods of years. I was greatly assisted in this endeavour by the generosity of the National Institute of Economic and Social Research in making available to me the results of their work in reducing the accounts of public companies to a standard form.[1] At the time of my enquiry, standardised accounts were available for all public companies engaged in the seven industry groups[2] listed in Table 5, and a random sample[3] was drawn for each group for the five years 1949-50 to 1953-4. Given capital employed and gross profits, rates of gross profit on

[1] Some preliminary results of its enquiry into the sources and uses of funds of public companies have been printed by the Institute for private circulation under the title of *Company Income and Finance 1949-53*, London, 1956.

[2] I use this term since the classification used by NIESR is considerably coarser than that of the Census of Production.

[3] The samples consisted of 40 companies from each group, except for the tobacco industry, where all companies were taken. Firms which were born into or disappeared from an industry during the period or which changed their accounting dates radically were excluded before sampling. The proportion of the sample to all public companies engaged in the industry throughout the period varied from two-fifths to two-thirds.

capital employed could be calculated[1] but β could not be determined.

ϵ cannot be calculated without β, but our interest at this point is not in absolute values of ϵ but in its stability over time. If we assume that β is itself relatively stable for any individual firm in the period covered—and this seems reasonable—then stability of a firm's relative rate of profit can be taken as indicating that ϵ also was stable.[2] The question to be addressed to the data therefore was how far the relative rates of profit of the individual firms in an industry were constant.

Within each industry, therefore, firms were ranked by rate of profit in each year and the data examined to see whether most firms tended to maintain roughly the same rank in all years. Chart V illustrates the situation which was typically found to obtain.[3]

It is visually evident in the case illustrated that there was in fact a high degree of stability in relative rates of profit (and therefore ϵ) during this period. But the chart is a complicated way of expressing this truth, even though there are only eleven companies in this industry. In the other industries, where the samples consist of forty companies, proof by inspection is virtually impossible. It is, therefore, desirable to have some summary numerical measure of the degree of stability.

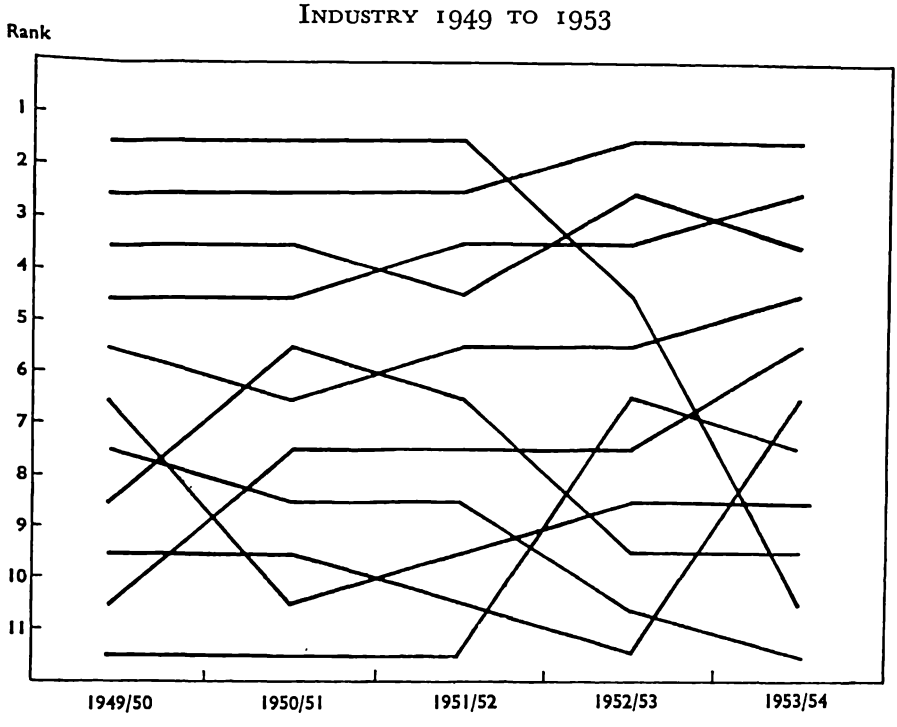
[1] The numerator of the rate was operating profit before deduction of depreciation and amortisation, this procedure being adopted so that profit movements would not be distorted by any year-to-year changes in depreciation practice in line with changes in tax laws. For similar reasons, fixed assets were taken gross of depreciation in the denominator of the expression, which I shall term operating capital. This last was, in any year, the arithmetic mean of the values of fixed assets and stocks and work in progress at the beginning and end of the year. Liquid assets, debtors and trade investments were excluded from capital (and dividends and interest received from profits), on the grounds that the yield on such assets is likely to differ significantly from that on capital more "closely" invested in the business and that their size is liable to considerable short-term fluctuations. The accounts of the Ford Motor Company Ltd. for 1955 provide a particularly striking example. Liquid assets and cash at the Bank amounted to £47 million and the value of fixed assets and stocks and work in progress £54 million. The yield on liquid assets was roughly $2\frac{1}{2}\%$ whereas the ratio of gross operating profit to operating capital was in the neighbourhood of 40%. Since it seems certain that the large investment in liquid assets was a purely temporary holding, destined to finance the firm's current expansion programme, it would clearly be extremely misleading to include it when calculating the firm's rate of profit.

[2] An addition to capacity during any year is unlikely to be fully in operation during that year. Hence β and r will be subject to fluctuation for this reason. The analysis, therefore, probably understates the "true" stability of ϵ.

[3] I am indebted to Mr. F. R. Oliver of Nuffield College for analysing these data.

Chart V

RANKING OF COMPANIES BY RATE OF PROFIT IN THE TOBACCO INDUSTRY 1949 TO 1953



The degree to which companies retain in one year their rank in another year can be expressed by a familiar measure—the coefficient of rank correlation.[1] And we might measure the stability of relative profit rates over the period as a whole by means of an average of the rank correlation coefficients between all possible pairs of years (irrespective of their contiguity). It is, however, more economical in computation to employ an alternative measure with which such an average has a precise relationship, namely the coefficient of concordance.[2] Were each firm to retain the same rank in every year, the total of its ranks over

[1] Spearman's coefficient of rank correlation ρ is defined by $1-\frac{6\Sigma(d^2)}{n^3-n}$, where d is the difference in rank in two years and n is the number of firms.

[2] If we write $\rho_{\text{av.}}$ for the average value of the Spearman coefficient, W for the coefficient of concordance, and m for the number of years, then $\rho_{\text{av.}}=\frac{mW-1}{m-1}$. See M. G. Kendall, *Rank Correlation Methods*, Chapter VI.

m years would be mx, where x is the (persistent) rank and the totals for all firms would form a series m, $2m$ . . . nm, where n is the number of firms.[1] The dispersion[2] of this series about its mean is clearly the maximum possible. If the ranking of firms changes from year to year the totals of their ranks will be less dispersed than the series just given, and the greater the variability of ranks the smaller will be the dispersion. When variability is at a maximum, the totals of ranks will be equal for all firms and the dispersion zero. The coefficient of concordance is the ratio of the actual dispersion to the maximum possible,[3] namely:

$$W=\frac{\Sigma(d^2)}{\frac{1}{12}m^2(n^3-n)}$$

where d is the deviation from the mean.[4] Since the numerator cannot exceed the denominator and has a minimum value of zero, the possible values of W are from 0 to 1.

Values of W for each of the seven industries are given in Table 4.

Table 5

VALUES OF THE COEFFICIENT OF CONCORDANCE OF PROFIT-RATE RANKINGS, 1949-54

1. Mining and treatment of non-metalliferous mining products	·808
2. Chemicals and allied trades	·721
3 Vehicles	·709
4. Cotton	·594
5. Wool	·706
6. Other textiles	·647
7. Clothing	·791
8. Tobacco	·836

These values indicate that there is, in fact, a strong tendency

[1] This ignores the possibility that some firms may have equal rates of profit.

[2] Measuring this, as in Chapter IV, by the sum of the squares of the deviations from the mean.

[3] The denominator is the formula for calculating the sum of the squares of the deviations of the series m, $2m$ . . . nm about its mean.

[4] See Kendall, *loc. cit.*

for relative rates of profit (and therefore ϵ) to remain constant over the period considered.[1]

This conclusion seems all the more justified in view of the broad scope of most of the industry groups listed, which represent aggregations of the sorts of industry (Census trades) which we have considered elsewhere. It follows from our earlier analysis that the average rate of profit ($\bar{r}$) in an industry is likely to be affected by the state of trade. Hence, if the different trades here aggregated experience unsynchronised changes in their state of trade the values of ρ and W will be lowered, even though the ranking of firms within each constituent trade remains completely stable. It is significant that the lowest coefficients are those for Cotton and "Other Textiles" which were hit by the cyclical fluctuation of 1950-53, whose timing was significantly different in different branches.

I conclude, therefore, that the problem of the relation between efficiency and growth is a genuine one which warrants the attention it has been given; that efficiency is a valid concept; and that in the further empirical work which follows we may take measures of efficiency drawn from one year as reasonably representative of a longer period within which that year is set.

[1] Some idea of the strength of the tendency may be gained by observing that the Spearman coefficient of rank correlation between the first forty natural numbers and a series identical in every respect save that 1 and 40 are transposed is 0·72.

Chapter XIII

THE EMPIRICAL MATERIAL[1]

THERE are two difficulties in empirical research. In the first place, it is usually impossible in practice to get the precise data which one needs. The alternatives then are to modify the concepts to suit the data or to torture the data until they yield approximate answers to the concepts. These alternatives are usually not exclusive in practice. In the second place, since the collection and processing of data is a lengthy process, it is frequently necessary to set this in train before the concepts and hypotheses to which the data are to be applied are fully thought out. The theory which prompted their collection may well have been modified by the time the data are available. In any event, it will be changed by the first confrontation, which will suggest new questions, requiring different data for their answering. These difficulties may serve in part to explain why the tests applied in the remaining chapters may not always be identical with those which would suggest themselves to an analyst who approached the problems *de novo*.

My basic need was measures of efficiency and size for firms in different industries over a period long enough for it to be reasonable to expect significant changes in both these variables to occur. It seemed to me that this requirement could not be satisfied by data from the published accounts of firms of the kind used in the preceding chapter. The watershed for such material is the Companies Act of 1948, which laid certain minimum obligations on firms regarding the publication of financial results. Prior to that date there were wide and uncertain differences in the practice of firms, with respect to both the amount of information published and the definitions employed. It is, therefore, uncertain whether any satisfactory attempt can be made to extend to

[1] The argument in the second part of this chapter is somewhat technical. It is intended only to establish that the ratio of the standard deviation of net output per head and average gross output per head can be taken as a reasonable approximation to the standard deviation of efficiency. Those who are prepared to take this result on trust may omit this part of the chapter without loss.

the pre-war period the sort of compilation carried out by the NIESR for the period since 1948. Certainly it has not yet been done. Moreover, as I have already noted, it is usually impossible to get from published material any precise idea of the industrial composition of the multi-trade firm, so that the industrial classification has to be rather coarse and the degree of heterogeneity within industries is a matter for speculation. I took, therefore, as my basic source the Census of Production, which would yield data for 1935 and 1948—an adequate span of years—addressed a common set of questions and definitions to its firms, and defined its industries in a satisfactory fashion.

I selected a sample of thirty trades from the one hundred and fifty-six into which the Census authorities divide the universe with which they deal. There were three principles of selection; first, that there should be industries of differing degrees of concentration; second, that there should be industries with and without entrenched agreements; third, that the sample should be reasonably comprehensive with respect to the type of market, in the sense that industries selling consumer goods, intermediate products and capital goods were included in roughly equal proportions. Degree of concentration was measured in the only way which is practicable without a special enquiry, by the proportion of total employment in an industry accounted for by the three largest firms and, the article by Leak and Maizels being the source, the concentration measured was that of 1935. The classification of industries into those with significant entrenched agreements and those without was a highly subjective affair, based on a variety of sources of information, which ranged from the reports of the Monopolies and Restrictive Practices Commission to credible gossip. An industry was classed as reasonably free from agreements in the absence of evidence to the contrary. I may, therefore, have understated the prevalence of agreements. The industries are listed in Table 6.

I was not given access to any information which might enable particular results to be identified with particular firms. Hence, when I wanted to use information of this character, my needs could be met only by partial processing of the primary statistics in the Census Office. I am very grateful to the Census Office

Table 6

CENSUS TRADES COVERED BY THE ENQUIRY

Degree of concentration	*Industries having restrictive agreements*	*Industries without serious restrictive agreements*
A. *High* (over 70%)	*Cement.*	*Dyes and dyestuffs.*
	Wrought-iron and steel tubes.	Soap, candles and glycerine.
	Batteries and accumulators.	*Margarine.*
	Rayon, nylon, etc., and silk.	Sugar and glucose.
	Wallpaper.	
	Tobacco.	
B. *Medium* (35%–50%)	*Rubber.*	Blast furnaces.
	Glass containers.	*Motor vehicles and cycles (manufacturing).*
	Electric wires and cables.	Printing and bookbinding machinery.
	Textile machinery.	Biscuits.
	Jute.	
C. *Low* (under 15%)	*Canvas goods and sacks.*	Mechanical handling equipment.
		Brick and fireclay.
		Hardware, hollow-ware, metal furniture and sheet metal.
		Leather (tanning and dressing).
		Wooden containers and baskets.
		Furniture and upholstery.
		Cotton weaving.
		Woollen and worsted.
		Hosiery and other knitted goods.
		Boots and shoes.

Note. The trades italicised are those for which the most complete Census information was secured (see below). The names of some trades are abbreviated, and further abbreviation is practised in subsequent tabulations.

for the co-operation which they showed, both in making available tabulations of primary statistics and in carrying out such partial processing. But the Office is fully (or over-fully) employed on official work and there was a limit to the amount of additional analysis for which I could reasonably ask. Moreover, there are some questions which can be answered satisfactorily only by visual inspection of primary data, for which processed data cannot be a substitute. For these reasons some questions which could be answered from Census data do not, in fact, receive fully satisfactory answers in what follows.

In the outcome, my basic data were as follows:[1]

(1) Frequency distributions of net output per head in 1948 by net output and employment. That is, for each of the thirty industries listed in Table 6, I had a table showing, for different values of net output per head, the net output (and labour employed) of the establishments in which net output per head had that value. The tabulation of net output per head was by £100 intervals.

(2) For the 16 industries italicised in Table 6 similar frequency distributions, based on a sample of establishments, for 1935.

(3) The samples consisted (except when the total number of firms in the industry was smaller than this) of 40 firms drawn at random from all those firms existing in 1948 which had also been in existence in 1935.[2] A firm was defined for this purpose as the aggregate of establishments in an industry trading under the same name.

(4) For each firm in the sample, I had net output per head and the ratio of net output to gross output in 1935 and 1948; the average annual earnings of employees (hereafter called the wage rate) in 1948; and the percentage change in net output between 1935 and 1948.

(5) Because of the provisions against disclosure, I could not have the values of net output for individual firms. In order partially to fill this gap, the Census Office were good enough to

[1] Such of the terms used in this summary as require it are defined in what follows.

[2] The procedure was, in fact, to draw a sample of more than 40 from the 1948 register of firms in a trade and then to exclude those which had come into being since 1935.

calculate for me the correlation between net output and both net output per head and the change in net output per head.

I now define the terms used in the preceding paragraph.

Gross output is virtually identical with turnover, being the total value of goods made and other work done during the year.

Net output is the value of gross output less the cost of material and fuels used and the amount paid for work given out. It will usually approximate closely, therefore, to the sum of payments to employees (hereafter called wages) and gross profits, but also includes certain payments for services.[1]

Net output per head (hereafter called productivity) is the ratio of net output and the number of employees.

The wage rate is the ratio of payments to employees and the number of employees.

Certain empirical simplifications are necessary before the data on productivity can be made to yield estimates of efficiency.

I defined $\epsilon = 1 - \beta(r - \bar{r})$

Average efficiency $= \bar{\epsilon} = \frac{\Sigma \epsilon s}{\Sigma s} = 1 - \frac{\Sigma k(r - \bar{r})}{\Sigma s} = 1$, since $\Sigma k(r - \bar{r}) = 0$

So $(\epsilon - \bar{\epsilon}) = -\beta(r - \bar{r})$

$(\epsilon - \bar{\epsilon})^2 s = \beta k(r - \bar{r})^2$

Let us suppose that β is identical for all firms in an industry.

Then $\sigma_\epsilon^2 = \frac{\beta \Sigma k(r - \bar{r})^2}{\Sigma s} = \beta \cdot \frac{\Sigma k}{\Sigma s} \sigma_r^2 = \beta^2 \sigma_r^2$

and $\sigma_\epsilon = \beta \sigma_r$

Let the rate of *gross* profit on capital employed be r', and the rate of depreciation in any year be

$$d = \frac{\text{Depreciation}}{\text{Capital employed}}$$

So $r' = r + d$

and $\bar{r}' = \bar{r} + \bar{d}$

Let us suppose that d is identical for all firms.

Then $(r' - \bar{r}') = (r - \bar{r})$,

and $\sigma_{r'}^2 = \sigma_r^2$

[1] For example, the hire of plant, machinery and vehicles; payments in respect of advertising and market research; repair and maintenance of buildings, plant and vehicles. See Table 18 in the Census reports.

Let us define the following further symbols.

n = net output

l = number of employees

W = wages

$p = \frac{n}{l}$ = productivity

$w = \frac{W}{l}$ = the wage rate

$i = \frac{k}{l}$ = capital employed per employee, or capital intensity

By definition, $r' = \frac{p - w}{i}$

Therefore
$$\bar{r}' = \frac{\Sigma k \frac{(p-w)}{i}}{\Sigma k} = \frac{\Sigma l(p-w)}{\Sigma l} \cdot \frac{\Sigma l}{\Sigma k} = \frac{\bar{p} - \bar{w}}{\bar{\imath}}$$
$$\sigma_{r'}^2 = \frac{\Sigma k(r' - \bar{r}')^2}{\Sigma k} = \frac{\Sigma k r'^2}{\Sigma k} - \bar{r}'^2$$

If now we write
$$w = \bar{w}(1 + x)$$

then
$$r' = \frac{1}{i}[p - \bar{w}(1+x)] = \frac{1+x}{i}\left(\frac{p}{1+x} - \bar{w}\right)$$
$$\sigma_{r'}^2 = \frac{\Sigma k}{\Sigma k}\left[\frac{(1+x)^2}{i^2}\left(\frac{p}{1+x} - \bar{w}\right)^2\right] - \bar{r}'^2$$
$$= \frac{\Sigma}{\Sigma k}\left[\frac{l(1+x)^2}{i}\left(\frac{p}{1+x} - \bar{w}\right)^2\right] - \bar{r}'^2$$

Now if $\frac{(1+x)^2}{i}$ is uncorrelated with l and x is uncorrelated with p, this becomes approximately
$$\sigma_{r'}^2 = \left[\frac{\sigma_p^2}{\bar{\imath}} + \frac{(\bar{p} - \bar{w})^2}{\bar{\imath}}\right] \cdot \frac{\Sigma l}{\Sigma k} - \bar{r}'^2 = \frac{\sigma_p^2}{\bar{\imath}^2}$$

Since $\sigma_{r'}$ is approximately equal to σ_r, and $\sigma_\epsilon = \beta\sigma_r$ we may write
$$\sigma_\epsilon = \frac{\sigma_p}{\bar{\imath}} \cdot \beta = \sigma_p \cdot \frac{\Sigma l}{\Sigma s}$$

$\frac{\Sigma s}{\Sigma l}$ is average gross output per head, for which we will write $\bar{t}$.

So $$\sigma_\epsilon = \frac{\sigma_p}{\bar{t}}$$

The assumptions made in the preceding analysis are the minimum necessary for deriving a measure of the dispersion of efficiency from the Census information available. What is their plausibility? The least plausible is that the capital/gross output coefficient (β) should be identical for all firms.

Let us write k_1 and k_2 for fixed capital and working capital respectively.

So $$\frac{k}{s} = \frac{k_1 + k_2}{s} = \frac{k_1}{n} \cdot \frac{n}{s} + \frac{k_2}{s}$$

It is by no means unreasonable to suppose that the ratio of fixed capital to net output varies little from firm to firm within the same industry. Increases in fixed capital per employee are usually regarded as resulting in greater productivity, and, given constant prices, all that is required to produce constancy of the capital/net output coefficient is that the increase in productivity should be proportionate to the increase in capital.[1] The ratio of stocks and work in progress to turnover will be a function largely of the average period of production in a firm. And, since the ratio of net output to gross output measures the amount of work done by the firm on the materials which it buys, we may take differences in this ratio as indicating, in a rough way, differences in the length of the period of production. Taking $\frac{k_1}{n} = b$ as constant and assuming that the relation between $\frac{k_2}{s}$ and the period of production is of the form $\frac{k_2}{s} = \lambda \frac{n}{s}$, we may therefore write

$$\frac{k}{s} = \frac{n}{s}(b + \lambda).$$

On this line of argument, therefore, the size of the error resulting from the assumption that β is identical for all firms in an

[1] The assumption that the marginal and average productivity of capital are equal does not contradict the classic "law of variable proportions" (diminishing returns to a variable factor), since I am not assuming the single-technique world from which the law is drawn.

industry will depend on the degree to which the ratio of net to gross output differs from firm to firm. Table 7 gives estimates of the variability of this ratio for the sixteen industries for which I have data.

Table 7

VARIABILITY OF THE NET OUTPUT/GROSS OUTPUT RATIO (g) WITHIN INDUSTRIES

	1935		*1948*	
Trade	$\sigma_g \times 100$	$\frac{\sigma_g}{\bar{g}} \times 100$	$\sigma_g \times 100$	$\frac{\sigma_g}{\bar{g}} \times 100$
Wallpaper	16·0	29	10·9	18
Dyes	13·4	35	15·3	41
Margarine	17·9	69	12·7	27
Wool	19·7	57	21·9	61
Blast furnaces	9·1	41	6·1	27
Batteries	12·5	26	7·5	20
Boots and shoes	11·8	25	10·8	23
Rubber	14·5	26	12·3	23
Leather	11·5	38	11·9	32
Textile machinery	13·4	22	11·4	18
Vehicles	14·9	31	14·8	31
Cotton	12·0	37	12·1	38
Glass	10·0	16	7·7	13
Canvas goods	12·9	41	12·4	43
Cement	15·7	32	11·8	28
Wires and cables	9·1	23	12·2	30

It is clear that, in certain trades at least, a significant error results from assuming that β does not vary from firm to firm. Until such time as estimates of β for individual firms are available we cannot, however, either quantify or remove the error in any precise way.

The other two assumptions give less cause for concern. I show in the next chapter that, although there is some tendency for high wage rates to be associated with high productivity, the correlation is usually small; that is, the tendency is only weak. To the extent that the association exists σ_p will be an over-estimate

of σ_ϵ. I cannot check from the data at my disposal whether large firms (size being measured by employment) tend to pay relatively high wage rates and to have relatively high capital intensities. There is evidence elsewhere in favour of a rather feeble positive relationship between size and wage-rates,[1] and there is a long-standing tradition amongst economists that bigger firms are more capital intensive in the sense in which I have defined the term. There may be, therefore, some association between l and $\frac{(1+x)^2}{i}$. But it is probably not strong and, in any event, so long as l is uncorrelated with p, which I later try to show is the case, the former association will introduce little error into our results.

[1] See, for example, "The Variability of Engineering Earnings", by T. P. Hill and K. G. J. C. Knowles in *Bulletin of the Oxford University Institute of Statistics*, Vol. 18, No. 2.

Chapter XIV

THE EXTENT TO WHICH EFFICIENCY VARIES

My object in this chapter is to show what support is given by the material just described to two propositions made earlier in the book. First—the starting point of the whole enquiry—that efficiency varies sufficiently from firm to firm or establishment to establishment within the same industry to warrant serious political attention. Second, that the degree of dispersion of efficiency observable at any moment of time is unlikely to reflect the "conditions of competition" within the industry in any simple and direct fashion, since it is itself the product of two opposing mechanisms, each of which is likely to be affected in the same direction (strength or weakness) by those conditions.

I give estimates in Table 8 of the standard deviation of efficiency in the 30 trades, calculated in accordance with the formula, established in the preceding chapter, that

$$\sigma_\epsilon = \frac{\sigma_p}{\bar{t}}$$

The reasons for selecting the standard deviation as the measure of the dispersion of efficiency were discussed in Chapter IV, where I concluded that it provided the best numerical measure of the inherently eclectic and slightly vague set of ideas which we have in mind when discussing this issue. It must be confessed, however, that it is difficult to give a verbal translation which is at once accurate and meaningful of what is meant by a particular value of σ. The best I can do—and this in the certainty of attracting the contempt of the statistical purist—is to suggest that σ_ϵ measures the extent to which, on the average, the efficiency of any individual firm will differ from the average efficiency of the industry. Courting still further disapproval, I take σ_ϵ to indicate, at any time, the proportion by which output in the industry could be increased, without bringing in additional resources, if production at below-average efficiency were eliminated.

It will be seen that in some industries the dispersion of efficiency exceeds one-fifth. The unweighted arithmetic mean for

Table 8

DISPERSIONS OF EFFICIENCY IN THIRTY INDUSTRIES IN 1948 AND 1935

Industry	£ *1948* σ_p	$\bar{l}$	σ_ϵ%	£ *1935* σ_p	$\bar{l}$	σ_ϵ%
Wallpaper	313	1,355	23	89	536	17
Dyes	290	2,195	13	105	995	11
Margarine	325	1,840	18	375	1,780	21
Wool	473	1,900	25	91	553	16
Blast furnaces	229	3,370	7	105	1,335	8
Batteries	230	1,320	17	97	541	18
Boots and shoes	116	1,160	10	49	342	14
Rubber	256	1,445	18	82	510	16
Leather	371	2,640	14	62	860	7
Textile machinery	166	881	19	63	305	21
Vehicles	167	1,480	11	45	625	7
Cotton	176	1,660	11	74	421	18
Glass	129	866	15	65	370	18
Canvas goods	149	2,240	7	59	620	10
Cement	194	2,710	7	165	960	17
Wires and cables	230	1,560	15	88	641	14
Tobacco	361	14,950	2	—	—	—
Tubes	190	1,540	12	—	—	—
Printing machinery	146	843	17	—	—	—
Biscuits	220	1,320	17	—	—	—
Handling equipment	155	1,080	14	—	—	—
Brick and fireclay	165	766	22	—	—	—
Hosiery	161	1,150	14	—	—	—
Wooden boxes	249	1,135	22	—	—	—
Sugar and glucose	216	7,500	3	—	—	—
Hardware	194	1,030	19	—	—	—
Furniture	191	1,040	19	—	—	—
Jute	110	1,285	9	—	—	—
Soap	319	2,670	12	—	—	—
Rayon	213	1,290	16	—	—	—
Unweighted average	—	—	14	—	—	14

Note. There is a formal inconsistency in these estimates. l is taken from the published reports of the Census. σ_p is calculated from the frequency distributions already described. The data in these were grouped and it was necessary to make an arbitrary assumption as to where the average output per head lay within the bounds of each class. The assumption made was that the average in each class was at the centre of the class.

all trades is in the neighbourhood of 14% in both 1948 and 1935. In other words, the "cost" of efficiency dispersion, by the arbitrary standard of measurement which I have adopted, is roughly equal to the annual value of gross fixed investment. Evidently, therefore, it is worth making sure that this dispersion is genuinely the cost of progress; i.e. that it is the "natural" product of the interaction of the transfer and innovation mechanisms, working with full efficiency, and not of their stultification.

I examined the main errors involved in the approximation to a true standard deviation of efficiency which I have been forced to adopt in the preceding chapter. Since all of them tend in the same direction, it is worth seeing whether any estimate of their magnitude is possible.

The first approximation lies in assuming constancy of the capital/output ratio from firm to firm. There is no systematic information available on how this ratio varies in the industries under examination. It is evident from inspection, however, that in certain industries a considerable part of the difference in productivity which we are measuring must be due to differences in capital per head. The very high productivities in certain establishments in the wool trade, for example, are explicable only on the assumption that they are wool brokers with high capital intensity. Similar factors account for part of the dispersion in Batteries. The point is further explored in the note at the end of the chapter.

The second assumption was that any association between wage rates and productivity is too weak to import significant error. The evidence on this is summarised in Table 9.

Since continuing use will be made of coefficients of the kind presented in Table 9, it is convenient to say something of their characteristics at this point. When we say that two quantities are strongly associated we can have two things in mind. First, that higher than average values of one will, almost invariably, be associated with higher (or lower) than average values of the other, in such a way that the higher the one the higher (or lower) the other. Second, this association ruling, that a given proportionate change in one quantity will be associated with a significant proportionate change in the other. The first of these meanings is

Table 9

RELATION BETWEEN WAGE-RATES AND PRODUCTIVITY IN 1948

Trade	£ *Standard deviation of wage rates* σ_w	*Coefficient of correlation between wage rates and productivity* r_{wp}	*Coefficient of regression of wage rates on productivity* b_{wp}
Wallpaper	28	0·25	·017
Dyes	69	0·35	·027
Margarine	67	0·62	·132
Wool	74	0·93	·179
Blast furnaces	49	0·43	·358
Batteries	64	0·71	·214
Boots and shoes	36	0·56	·138
Rubber	60	0·58	·135
Leather	80	0·43	·115
Textile machinery	55	0·63	·172
Vehicles	61	0·62	·216
Cotton	28	0·22	·051
Glass	45	0·40	·107
Canvas goods	54	0·29	·105
Cement	56	0·35	·059
Wires and cables	56	0·65	·163

Note. The coefficients are calculated from the data on average remuneration per head and average net output per head for samples of 40 firms described in Chapter XIII.

expressed by the coefficient of correlation and the second by the coefficient of regression.

The coefficient of correlation is defined as

$$r_{wp}=\frac{\Sigma(w-\bar{w})(p-\bar{p})}{N}\,\frac{1}{\sigma_w\sigma_p}$$

where N is the number of pairs of values included in the summation. It indicates the degree of error involved in assuming a linear relation between two variables. Its possible values vary

from 0 to ± 1, the higher the value the closer being the approximation.[1] The coefficient of regression of wage-rates on productivity is defined as

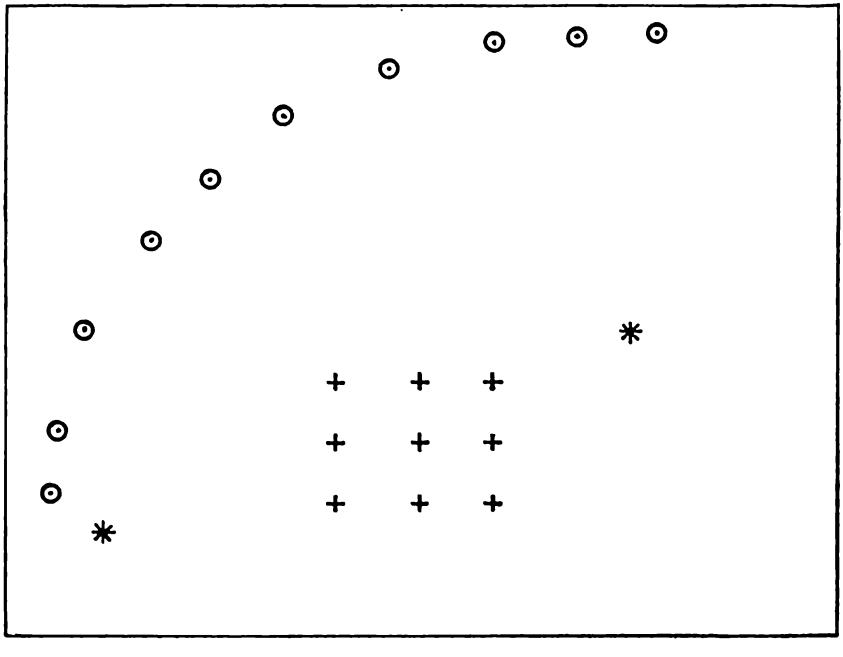
$$b_{wp} = \frac{\Sigma(w-\bar{w})(p-\bar{p})}{N} \cdot \frac{1}{\sigma_p^2}$$

b is the coefficient in the linear relationship assumed. A value of ·17 means that, in so far as the relationship holds, if there is a difference of £100 in productivity between two firms there will be a difference of £17 in the average wage-rates which they pay.

The coefficient of correlation is, unfortunately, by no means unambiguous in its meaning. Visual inspection of the data is needed before any firm conclusion can be drawn from a particular value of the coefficient. This can be seen from Chart VI.

Chart VI

ILLUSTRATIONS OF THE CORRELATION COEFFICIENT



[1] A positive value indicates a direct relationship between two variables and a negative value an inverse relationship.

If the two extreme pairs of values, distinguished by asterisks, are excluded the value of the correlation coefficient for the lower part of the chart is negligible; when these two are included the value becomes appreciable. In the absence of visual inspection, the temptation would be to conclude that there was a significant association in the one case but none in the other. It is obviously very much open to question whether this conclusion is warranted. Alternatively, the correlation coefficient for the values in the upper part of the chart is very low. But a very definite *non-linear* association can be seen to exist.

Let us return now to the particular issue of the strength of the association between wage rates and productivity with which we are immediately concerned. In order to justify the method of estimating σ_ϵ it is necessary that r_{wp} be small and/or that b_{wp} be small.[1]

It will be seen that the coefficients of correlation are in some cases substantial, although the largest of all—that for the wool industry—has to be rejected as being of the kind illustrated in Chart VI. But in no case does a high coefficient of correlation co-exist with a high coefficient of regression. Indeed, most of the coefficients of regression are rather small. We may conclude, therefore, that the error imported into our estimates of σ_ϵ by the assumption that wage rates do not vary with productivity is not large.

There is, finally, the assumption that high productivity and large size, measured by employment, are also not strongly associated. Approximate estimates of the coefficients of correlation between these variables are presented in Table 10. Since most of the coefficients are so small there is no point in estimating the coefficients of regression.

[1] Let productivity be £400 and £300 in two firms, the average wage rate in the industry be £200, and capital per worker be i. Assuming that both firms pay the same wage, we shall calculate their rates of profit as $\frac{200}{i}$ and $\frac{100}{i}$. If wage rates are linearly related to productivity, the coefficient of regression is 0·10, and the firm with lower productivity pays wage rates of £200, the rates of profit will be $\frac{190}{i}$ and $\frac{100}{i}$, so that the change is small. If, however, the coefficient of regression were 0·80, the rates would then be $\frac{120}{i}$ and $\frac{100}{i}$, a large change.

Table 10[1]

COEFFICIENTS OF CORRELATION BETWEEN PRODUCTIVITY AND SIZE (MEASURED BY EMPLOYMENT) IN THIRTY INDUSTRIES IN 1948

Industry	*r*	*Industry*	*r*
Wallpaper	−·47	Wires and cables	−·08
Dyes	−·10	Tobacco	+·07
Margarine	+·57	Tubes	−·03
Wool	−·01	Printing machinery	−·06
Blast furnaces	−·07	Biscuits	−·12
Batteries	−·04	Handling equipment	−·01
Boots and shoes	+·01	Brick and fireclay	+·13
Rubber	+·05	Hosiery	+·02
Leather	+·24	Wooden boxes	−·12
Textile machinery	−·08	Sugar and glucose	−·14
Vehicles	+·03	Hardware	·00
Cotton	−·07	Furniture	+·02
Glass	−·09	Jute	−·16
Canvas goods	−·01	Soap	−·05
Cement	+·57	Rayon	+·12

In view of the rough way in which the coefficients had to be calculated and of the impossibility of any visual assessment of their meaning, I am not disposed to build overmuch upon them. It seems legitimate to conclude, however, as with the other

[1] The coefficients were calculated by the following method:

$$(p-\bar{p})(l-\bar{l})=pl-p\bar{l}-\bar{p}(l-\bar{l})$$

$$\text{So } \frac{\Sigma(p-\bar{p})(l-\bar{l})}{N}=\bar{n}-\bar{p}\bar{l}, \text{ and } r_{pl}=\frac{\bar{n}-\bar{p}\bar{l}}{\sigma_p\sigma_l}$$

The averages refer to establishments. Estimates of $\bar{n}$ and $\bar{l}$ are directly available from the published reports on the Census. σ_l was calculated from Table 4 of the Census. It is a minimum estimate, in as much as it assumes that all establishments within a given employment size class are of identical size. $\bar{p}$ and σ_p were calculated from my frequency distributions of productivity, weighting by the number of establishments in each class of output per head. The estimate of $\bar{p}$ is subject to error, since it assumes that average productivity within a given class is equal to the central value of the class. σ_p is a minimum estimate since it assumes no dispersion within productivity classes. It is difficult to be at all certain about the degree to which these errors cancel out. Information on establishment sizes being not available for the 40 firm samples, no visual check on the meaning of these coefficients could be made.

assumptions examined, that there is no reason to regard the estimates of the standard deviation of efficiency in Table 8 as gross exaggerations of the true dispersion which we are interested in measuring.[1]

Table 11

AVERAGE DISPERSION OF EFFICIENCY IN DIFFERENT GROUPS OF TRADES IN 1948

Degree of concentration	*Industries having restrictive agreements*	*Industries without serious restrictive agreements*	*Total*
High	12·0	11·5	11·8
Medium . . .	15·2	13·0	14·2
Low	7*	16·4	15·6
Total. . . .	13·0	15·9	14·7

Note. This table is derived from Tables 6 and 8, except that two trades—wool and batteries—have been excluded for the reasons given. Unweighted arithmetic averages of the standard deviations for individual trades have been taken. The asterisked compartment contains only one entry.

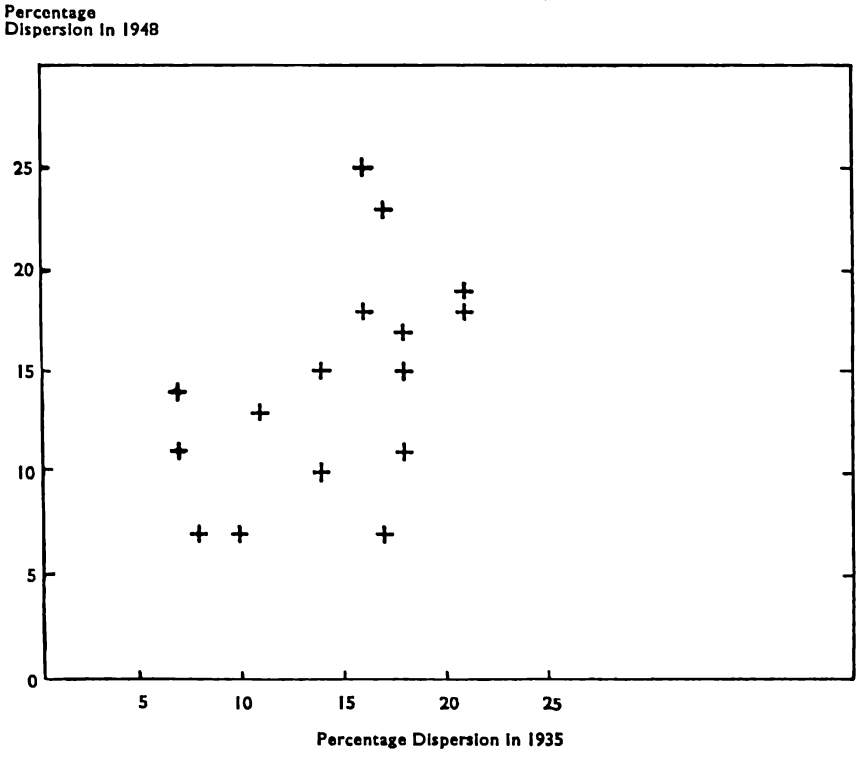
In Table 11 I have grouped the results from Table 8, in order to see whether there is any evidence of an association between the dispersion of efficiency and the "conditions of competition". The table suggests that there is, in fact, a tendency for the degree of dispersion to diminish as the degree of concentration increases, a result which conforms with the analysis of Chapters VI and VIII. There is also come suggestion, greatly weakened by the fact that the lower left-hand cell in the table contains only one entry, that industries with restrictive agreements have higher dispersions than those without. Unfortunately the variation of the values of σ_ϵ for the individual trades in each cell about the average for the cell is so considerable that little significant meaning

[1] The low values of the correlation coefficient in Table 10 may cause some surprise to those whose ideas have been formed by Table 4 of the 1948 Census. The reasons why grouping of data, as in that table, should suggest a strong correlation even when there is no association between the underlying variables have been set out by a number of writers. (See, for example, "Labour Productivity and Size of Establishment", by J. Johnston, in *Bulletin of the Oxford University Institute of Statistics*, Vol. 16, Nos. 11 and 12.) More generally, there seems no theoretical reason to expect productivity and employment size to be correlated. Indeed, if productivity and output size were highly correlated, which some might think more likely, productivity and employment size would not be.

can be attached to these differences between the averages. Moreover, as is evident from Chart VII, changes over time in the dispersion of efficiency of an industry can be very substantial.

Chart VII

DISPERSION OF EFFICIENCY IN 1935 AND 1948



There is only a minimal tendency for industries with high (low) dispersion in 1935 to have high (low) dispersion in 1948. This chart implies, and the evidence of Table 11 cannot be taken to negate the implication, that the dispersion of efficiency which is found in an industry at any particular point of time is itself of little significance as a pointer to either the behaviour of that industry over time or the rules of the game which help to determine the behaviour.

Before this conclusion is accepted, however, it will be useful to examine the *shapes* of the distributions with which we have

been dealing. There are two reasons for so doing. First, as was pointed out in Chapter IV, there are some shapes of curve where the summary measure (σ) which I have been employing in this chapter will not in fact express properly what we have in mind when we talk of the dispersion of efficiency. Second, we need to satisfy ourselves that there is no challenge to the theoretical analysis, of the kind which would be presented by a systematic association between particular shapes of curve and particular conditions of competition. For lack of comparable frequency distributions of gross output per head or wages, the examination must be confined to the shapes of the productivity curves.

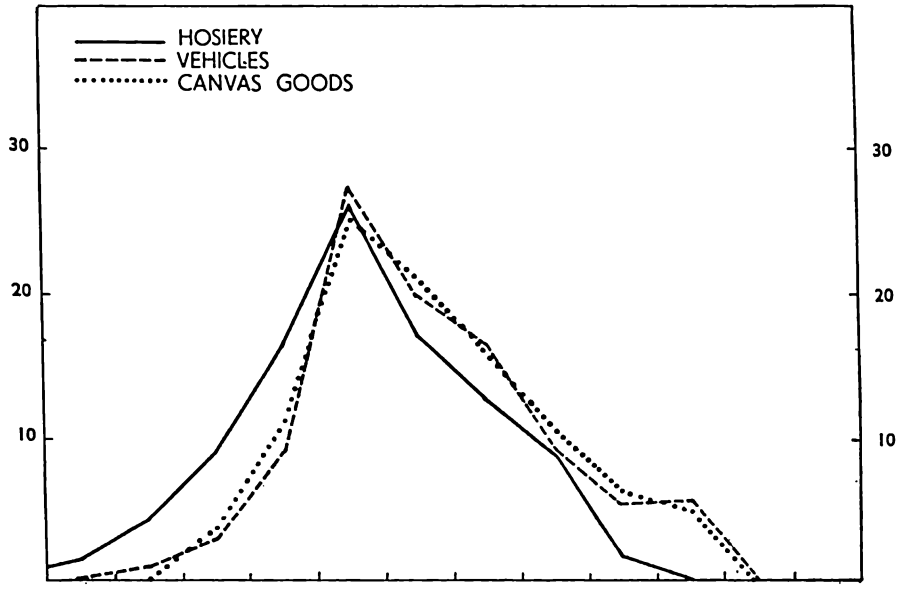
In order to compare the shapes of different distributions we must draw them all on the same scale. Differences in the vertical scale—representing differences in the sizes of industries—are easily eliminated by expressing the number of workers employed in establishments where productivity falls within a certain range as a proportion of the total number of workers in the industry. Harmonisation of the horizontal scales is necessary because of the wide differences from industry to industry in both average productivity and its dispersion. Comparability can be secured by taking as the dimension of the class interval for each industry the standard deviation of productivity or some fraction of it. In drawing Chart VIII, I have taken $\frac{1}{2}\sigma_p$ as the standard class interval.[1]

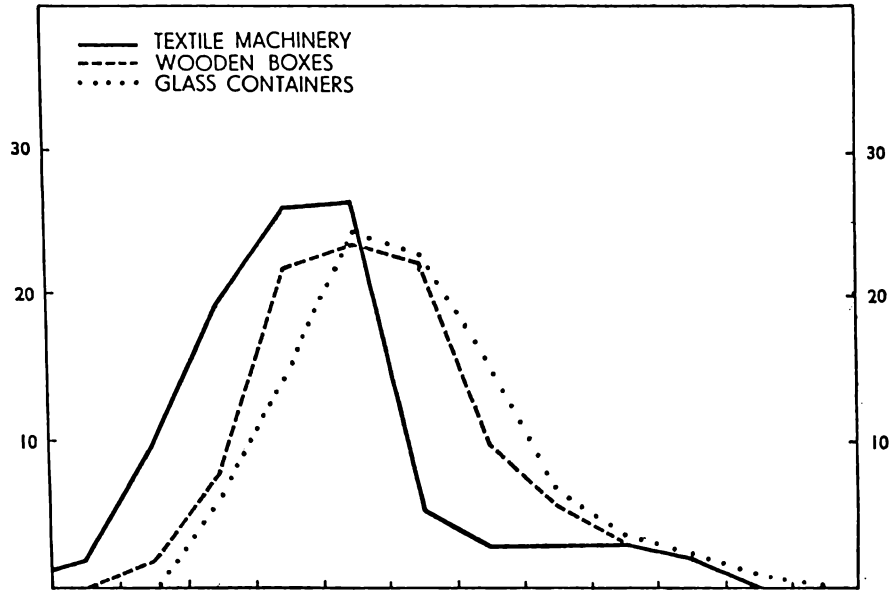
With two exceptions, I have shown the curves in triplets, to avoid confusion, grouping together the three curves exhibiting the greatest similarity in shape. The triplets themselves fall into three groups. The first is typified by the triplet Hosiery, Vehicles and Canvas Goods; a curve of moderate height, exhibiting a high degree of symmetry, with such tail as there is lying towards the right. In some cases, for example Dyes, Wool and Printing Machinery, the curve becomes much more sharply peaked, without, however, losing its essential symmetry. The other two groups are characterised by significant asymmetries, which amount in most cases to definite bi-modality; in the one group the lesser "most typical" level of productivity forms part of a long tail of

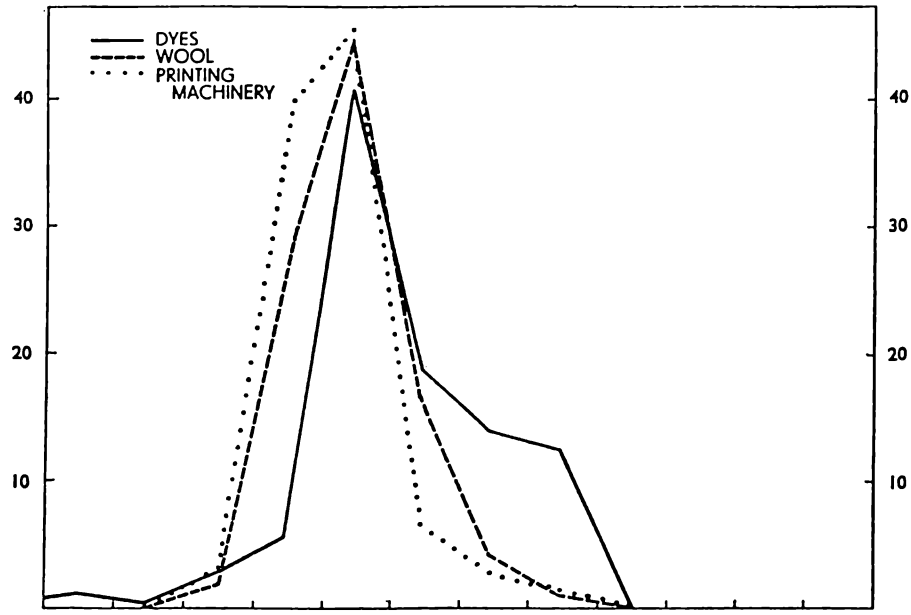
[1] As noted in Chapter XIII, the original data were grouped into classes of £100. Cumulative frequency curves were constructed from these, and the frequencies appearing in Chart VIII were then read off for the new intervals of $\frac{1}{2}\sigma_p$.

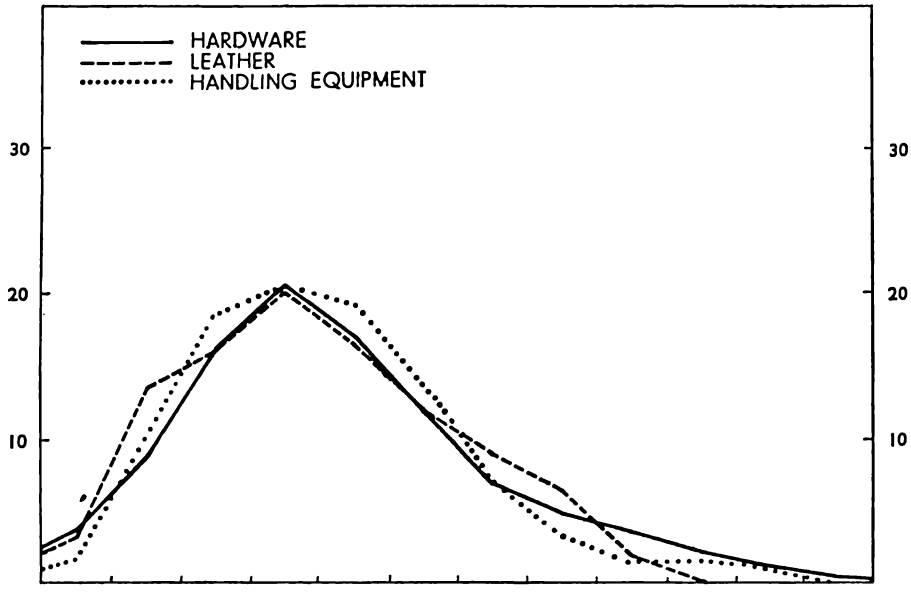
Charts VIII/1 to VIII/10

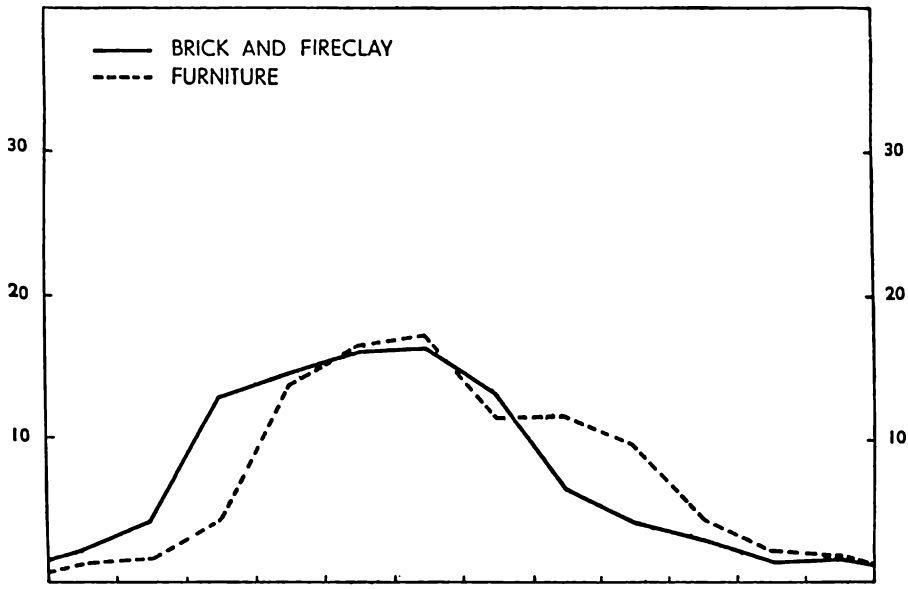
FREQUENCY DISTRIBUTIONS OF PRODUCTIVITY IN THIRTY INDUSTRIES
(1 CLASS INTERVAL $=\frac{1}{2}\sigma_p$)

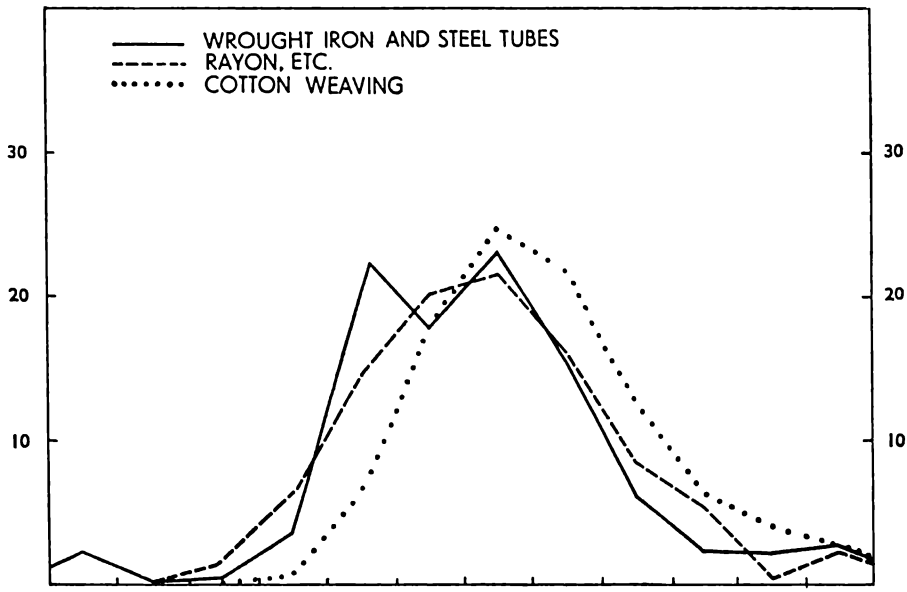


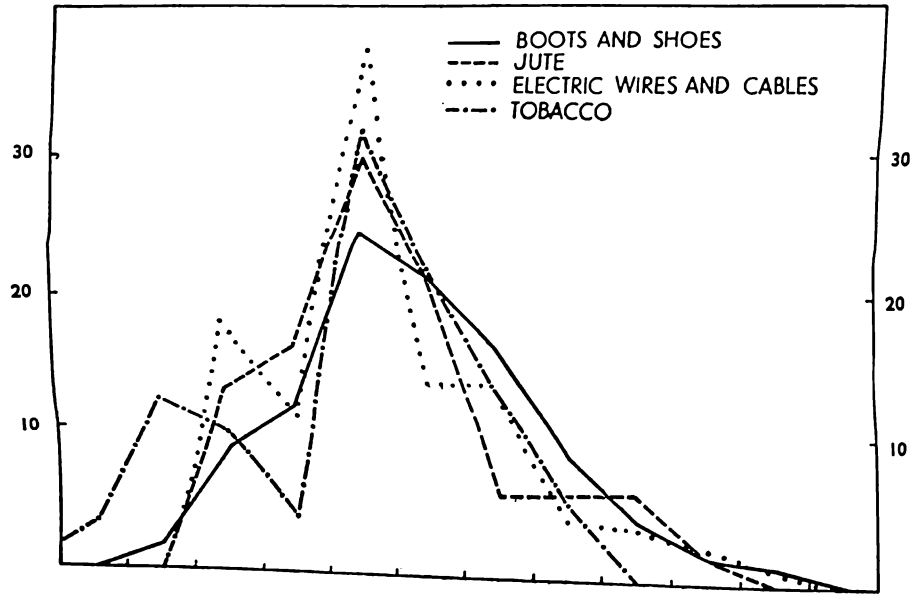


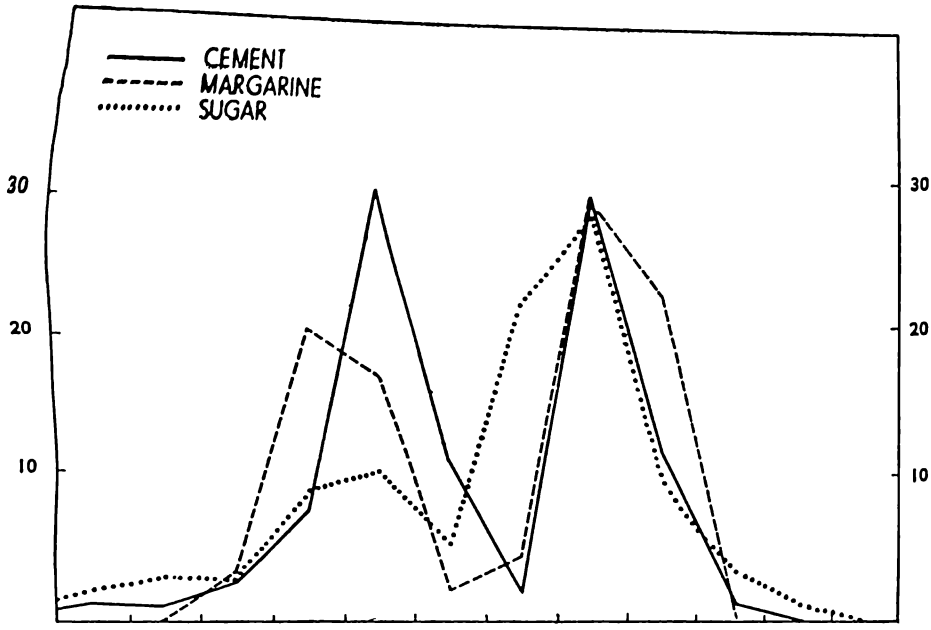

DYES
WOOL
PRINTING MACHINERY
40
30
20
10

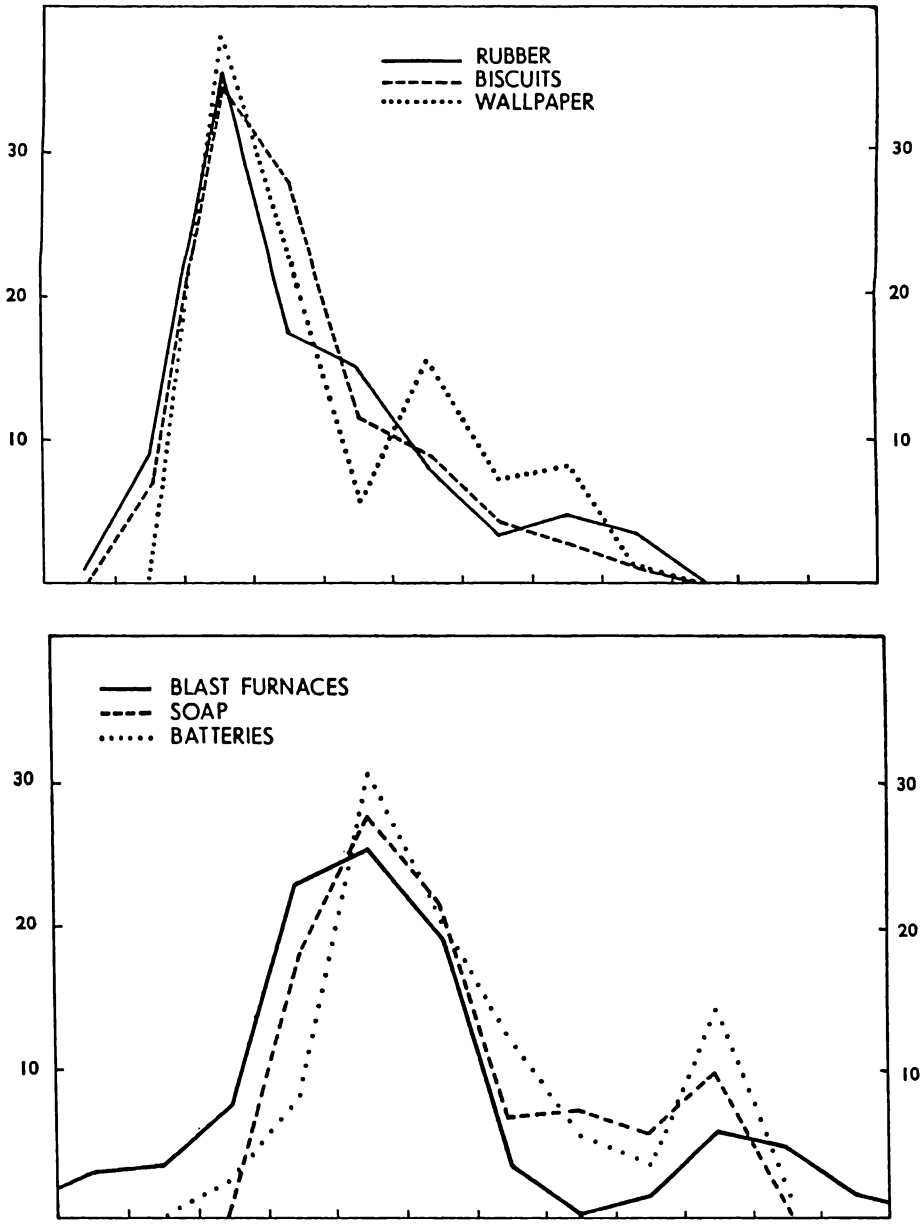

HARDWARE
LEATHER
HANDLING EQUIPMENT
30
20
10


BRICK AND FIRECLAY
FURNITURE
30
20
10
30
20
10


WROUGHT IRON AND STEEL TUBES
RAYON, ETC.
COTTON WEAVING
30
20
10
30
20
10


BOOTS AND SHOES
JUTE
ELECTRIC WIRES AND CABLES
TOBACCO
30
20
10
30
20
10


CEMENT
MARGARINE
SUGAR
30
20
10
30
20
10



Note. Net output per head is measured along the horizontal axis and one unit is equal to one half the standard deviation of productivity in the industry in question. Along the vertical axis are measured proportions of the total labour force of the industry employed in establishments whose average productivity is as indicated. For ease of comparison of the shapes of the distributions the modal classes are made to coincide within each triplet.

relatively low values; in the other, the tail and the minor mode lie to the right.

It is clear that the standard deviation does not always truly measure the magnitude in which we are interested, since extreme values, which represent very little employment or output, are nevertheless[1] able to exercise a powerful influence on the value of σ. Moreover, the double peak of a curve such as that for cement is an important feature which the summary measure conceals. In most cases, however, we can evidently rest content with the adequacy of our measure.

It will be seen that most of the industries with the single-mode moderately symmetrical curves are those which I have classed as non-concentrated and non-restricted. The correspondence is not complete; the dye industry, for example, finds its closest doubles in the wool and printing machinery industries. Moreover, there are purely statistical reasons why the more highly concentrated industries should display the greatest irregularity in shape. The smaller the numbers in a distribution the less smooth it is likely to be. The more concentrated industries tend, as a matter of course, to have fewer establishments than the less. The probability that this will lead to an irregular distribution is all the greater in so far as firms in such industries avail themselves of the option which the Census gives for a multi-establishment firm to render a consolidated return.[2] Nevertheless, it is at least suggestive of the value of further investigation that the industries bearing long unassimilated tails of low productivity (high cost) firms are also concentrated and known to have entrenched restrictive agreements. I suggested in Chapter VI that this was an essentially unstable situation, likely to be found only where the transfer mechanism had been suspended. *Prima facie* grounds for this suspicion is, I think, the most we can derive from data confined to one period of time only. In the next chapter I examine whether further light can be shed from information on changes over time.

[1] Because of the use of squares of deviations.

[2] In such cases the practice of the Census Office is to apportion net output, etc., among the different establishments in proportion to the numbers employed.

Note to Chapter XIV

I pointed out, in discussing Table 8, that certain of the Census trades cover a number of technologically heterogeneous activities, and that the structure of production costs is likely to vary substantially from one such activity to another. In such cases the standard deviation of productivity is likely to exaggerate the underlying dispersion of efficiency.

Some idea of the degree of this exaggeration can be derived. In all cases where this is important, the Census reports contain an additional table—Table 6, "Analysis according to specialisation within the trade"—which provides, *inter alia*, data on average productivity in the various sub-trades which can be distinguished within the main trade. This makes it possible to estimate how much of the standard deviation of productivity in the trade as a whole results from differences between sub-trades—which we may take to be irrelevant to our enquiry—and how much from variations within sub-trades. The ratio of the former to the total may be taken to provide a minimum index of the degree of exaggeration imported into our estimate of σ_ϵ. Values of the index of exaggeration are as follows:

PROPORTION OF DISPERSION DUE TO SPECIALISATION IN CERTAIN TRADES IN 1948

Trade	*Index of exaggeration* (%)
Batteries	36·0
Wool	27·0
Leather	12·5
Rayon	12·0
Sugar and glucose	7·1
Wooden boxes	7·1
Hardware	6·3
Furniture	5·3
Bricks and fireclay	5·1
Rubber	3·5
Canvas goods	3·5
Vehicles	2·5
Boots and shoes	1·5
Cotton	0·5

The index is calculated as follows:

Let sub-trades within an industry be denoted by subscripts a, b, etc.

$$\sigma_p{}^2 = \frac{\Sigma(p_a - \bar{p})^2 l_a + \Sigma(p_b - \bar{p})^2 l_b \ldots}{\Sigma l}$$

And $$\sigma_{pa}{}^2 = \frac{\Sigma(p_a - \bar{p}_a)^2 l_a}{\Sigma l_a} = \frac{\Sigma(p_a - \bar{p})^2 l_a}{\Sigma l_a} - (\bar{p} - \bar{p}_a)^2$$

Or

$$\Sigma(p_a - \bar{p})^2 l_a = \Sigma l_a \sigma_{pa}{}^2 + \Sigma l_a (\bar{p} - \bar{p}_a)^2$$

So $$\sigma_p{}^2 = \frac{\Sigma l_a \sigma_{pa}{}^2 + \Sigma l_b \sigma_{pb}{}^2 \ldots + \Sigma l_a (\bar{p} - \bar{p}_a)^2 + \Sigma l_b (\bar{p} - \bar{p}_b)^2 \ldots}{\Sigma l}$$

The second series of terms on the right in this expression measure the dispersion due to specialisation within the trade (S^2), the first series is the weighted sum of variations within sub-trades.

p_a, p_b . . . and l_a, l_b . . . etc., are provided in Table 6 of the Census reports, so that S can be calculated.

The index of exaggeration is $\frac{S}{\sigma_p}$.

Chapter XV

ON EFFICIENCY AND GROWTH

I WANT now to examine the evidence for three more of the propositions developed earlier:

(*a*) that there is a detectable process within industries of the kind which I have christened the transfer mechanism;

(*b*) that the changes in relative firm sizes which constitute the process are not random but associated with efficiency;

(*c*) and that the power of the transfer mechanism varies with the degree of concentration and of restrictive agreement.

The evidence available is the data for samples of forty firms from each of sixteen industries described in Chapter XIII.

The minimum evidence necessary to establish the first proposition is a demonstration that the relative sizes of firms within an industry change considerably over time. The most nearly appropriate measure of the change in the size of a firm which I have is the percentage change in its net output between 1935 and 1948. Such an index reflects, of course, changes in prices as well as changes in the volume of output, and the former have been very large over the period under consideration. Moreover, to the extent that different firms have had different price histories comparison of the changes in their net outputs may mislead us as to changes in their relative market shares. Nevertheless, while these qualifications need to be kept in mind, I think the data are capable of serving the purpose.

So long as the frequency distribution of changes in firm size within an industry is not too bizarre in shape, we may again make use of the standard deviation of the distribution to form a summary measure of the degree of flux.

The standard deviation itself will provide a measure of the extent to which individual changes in size differ from the average for the industry. A value of zero for the standard deviation would mean that all firms had grown at the same rate, so that relative

firm sizes were unchanged. Other things being equal, the greater the value of the standard deviation the greater the change in the structure of market shares within the industry.

But other things are seldom equal. The same value for the standard deviation may mean very different things in terms of size-rankings within an industry according to the value of the average change in size about which it is measured. A standard deviation of 200% means much less in terms of changes in the structure of an industry if the mean ratio of net output in 1948 to that in 1935 is 300% than it does if the average is only 100%. The implication is, therefore, that we should look to the coefficient of variation[1] rather than to the standard deviation alone.

Even this index, however, does not measure all that we have in mind. The coefficient of variation of (unweighted) changes in net output takes no account of any possible association between changes in output and the size of output in the base period. Hence, two industries might show the same value for the coefficient, although in one industry the large changes were all made by small firms and, in the other, they were randomly distributed with respect to size. In one very important sense, however, the structure of the second industry would have changed more than that of the first, and I think that this sense must be covered by our measure. Before, therefore, we can use the coefficient of variation with equanimity we must satisfy ourselves as to the relation between size and changes in size.

I have set out the relevant figures in Table 12. The third column demolishes the doubts which I have just raised about the propriety of treating the coefficient of variation as an index of changes in the structure of an industry. Such relation between initial size and subsequent growth as exists is in accordance with *a priori* expectations. Small firms are more likely to experience large percentage growths than are large.[2] But in all cases the size of the correlation coefficients is so small that it is legitimate to proceed as if the rate of growth of a firm were independent of its size.

The coefficients of variation themselves are surprisingly large.

[1] It will be recalled that this is the ratio of the standard deviation and the mean.
[2] This is the meaning of a negative coefficient of correlation.

Table 12

CHANGES IN NET OUTPUT FROM 1935 TO 1948

Trade	$\bar{s}$	$\frac{\sigma_s}{\bar{s}} \times 100$	$r_{s \cdot n_{35}}$
Wallpaper . . .	...	...	...
Dyes	457	84	—·01
Margarine . . .	208	92	—·09
Wool	351	76	—·09
Blast furnaces . .	541	88	—·18
Batteries . . .	349	55	—·12
Boots and shoes . .	422	80	—·17
Rubber . . .	502	94	—·14
Leather . . .	501	79	—·18
Textile machinery .	541	67	—·11
Vehicles . . .	541	125	—·17
Cotton . . .	297	75	—·38
Glass	357	65	—·17
Canvas goods . .	393	65	—·24
Cement . . .	314	51	—·17
Wires and cables . .	490	45	...

Note. $s = \frac{n_{48}}{n_{35}} \times 100$, where n is net output and subscripts refer to years.

$r_{s \cdot n_{35}}$ is the coefficient of correlation between net output in 1935 and change in net output between 1935 and 1948.

The shape of the frequency distribution for wallpaper is too bizarre to make the coefficients meaningful.

The average value for all industries in the table is 76%, and in only one industry is it less than 50%. Adopting the crude intuitive interpretation of coefficients of variation suggested in the previous chapter, we may take a value of, say, 75% to mean that if we selected a number of firms at random from an industry we should find that, on the average, their individual rates of growth tended to differ from the average rate of growth of all firms in the industry by as much as 75%. This means that there were pervasive and substantial changes in relative firm sizes in all the industries between 1935 and 1948. In other words, the first of the propositions with which I began this chapter is established.

Consider now the second proposition, that changes in relative firm size are not random but are associated with relative efficiency, the process being for relatively efficient firms to grow fastest. To test this we have the data on changes in net output for samples of firms which have just been examined and information on productivity in the same firms in 1935 and 1948, the initial and terminal dates of the period under review.

The number of equalities necessary to render relative productivities precise measures of efficiencies has been exposed in Chapter XIII. When we are constructing aggregative measures, like the standard deviation, the fact that, say, capital/output ratios will in fact differ from firm to firm does not necessarily mean that to calculate as if they did not will yield a wrong answer. All that is needed to yield a reasonable answer is that they should not be strongly correlated with productivity. But to test the proposition now under examination it is necessary to look at efficiency and growth for individual firms. We can no longer assume, therefore, that differences in the other variables entering into the definition of efficiency will cancel out in the wash so as to leave productivity a close index of efficiency. For this reason, no very high correlation between productivity and growth would be expected, even in industries where the transfer mechanism was operating in perfect conformity with the theoretical model developed in Chapter VI.

There is a further difficulty. The relative growth of a firm in an industry in which the transfer mechanism is operating in accordance with the theory will depend on its efficiency at each moment of time in the period over which growth is measured. And if the innovation mechanism is also at work the firm's relative efficiency will not be the same at all such moments. There is no particular reason to suppose that a firm's relative productivity in either 1935 or 1948, or even an average of the two, will give a good measure of its average relative productivity over a period so long as thirteen years.[1] But our information is confined to these two years.

Given these two qualifications, there is little likelihood of any

[1] The improbability is all the greater in view of the large changes in relative productivities which took place over the period. See Chapter XVI.

significant correlation between changes in net output between 1935 and 1948 and productivity in either of those years.[1] Indeed, if correlation were found it would be difficult to explain.

I therefore take a more limited proposition for statistical testing; namely, that those firms whose relative efficiency *increases* most over any period will also tend to grow most over the period. A demonstration that the statistical evidence is not inconsistent with this proposition does not, of course, prove that the transfer mechanism is operating in the way I have suggested. But the essential idea within the notion of the transfer mechanism is that virtue should be rewarded, and such a test *is* relevant to determining whether this law rules. Moreover, I believe that we shall be on much firmer ground in taking a *change* in relative productivity (which is all we have to go on) as indicating a *change* in efficiency than in taking relative productivity at any particular time as an index of efficiency. I have, therefore, calculated the coefficients of correlation between changes in net output and changes in net output per head between 1935 and 1948 for the sixteen industries for which I have the necessary information (see Table 13).

The coefficients suggest a very definite association in most industries between changes in productivity and growth. They do not in themselves, however, establish the proposition that virtue is rewarded. Correlation analysis can show that two variables are functionally related; it cannot establish that there is a causal relationship, still less which variable is cause and which effect. To assert causality on the basis of a systematic association we must, first, produce independent theoretical reasons why such a causal connection should exist, and, second, satisfy ourselves that no alternative explanation of the connection is plausible.

The first requirement has already been met. The second is more difficult to satisfy. On the evidence so far presented, it is perfectly possible to argue that the causal connection is the reverse of that which I have been trying to establish; i.e. that firms which grow fastest increase their productivity most, because large size leads to (causes) high productivity. Such an assertion has, indeed, a very respectable ancestry. The doctrine that there are

[1] A summary of the insignificant coefficients actually found is given in the Appendix.

Table 13

CORRELATION BETWEEN CHANGES IN NET OUTPUT AND CHANGES IN PRODUCTIVITY IN SIXTEEN INDUSTRIES BETWEEN 1935 AND 1948

	r
Wallpaper	...
Dyes	·24
Margarine	·46
Wool	·35
Blast furnaces	·18
Batteries	·57
Boots and shoes	·71
Rubber	·85
Leather	·94
Textile machinery	·71
Vehicles	·56
Cotton	·55
Glass	·10
Canvas goods	·65
Cement	·81
Wires and cables	·57

Note. The correlation is between $\frac{n_{48}}{n_{35}}$ and $\frac{p_{48}}{p_{35}}=\frac{n_{48}}{n_{35}}.\frac{l_{35}}{l_{48}}$. The calculation for wallpaper is so dominated by the enormous changes in output and productivity in a few firms that the coefficient has been omitted.

economies of scale is one of the most deeply entrenched beliefs in economics.

The only additional statistical evidence relevant to a choice between the two alternative hypotheses about the nature of the causal connection is that on the association between size of firm and productivity at any particular point of time. I have already shown that no such association can be distinguished when size is measured by employment. The result is in most cases not much stronger when size is measured by net output (Table 14).

The additional evidence is itself not easy to interpret. Suppose that, other things being equal, productivity did in fact rise with size. But let it also be the case that size at any time is a function

Table 14

CORRELATION BETWEEN NET OUTPUT AND PRODUCTIVITY IN 1935

	r_{np}
Wallpaper	·54
Dyes	·00
Margarine	·84
Wool	·14
Blast furnaces	·03
Batteries	·12
Boots and shoes	·16
Rubber	·08
Leather	·40
Textile machinery	·29
Vehicles	·05
Cotton	·62
Glass	·13
Canvas goods	·28
Cement	·46
Wires and cables	—

Note. The correlation coefficients were calculated for the samples of firms. The Census Office supplied totals for $\Sigma(n-\bar{n})(p-\bar{p})$ and the standard deviations of firm size in the different industries, so that the indirect methods of Table 10 could be avoided.

of productivity in the past, and that other factors beside size determine productivity. Then we should not expect to find any very close relation at any given moment of time between size and productivity. The large firm of today would be so because it had achieved high efficiency (productivity) in the past. Its growth would have raised its *potential* productivity still higher. But if the luck or skill which had favoured it in the past had by then deserted it, as the analysis of Chapter VII suggests is likely, its *actual* productivity would by no means reflect the additional advantages gained through growth. More generally, since the present size of a firm is the product of the whole of its past history, it may be argued that there is no incompatibility between the fact that productivity and size are only weakly correlated and

the theory that it is changes in size which cause changes in productivity.

It is not possible to disprove this argument definitively. But I think that the balance of the evidence is sufficiently against it to shift the onus of proof on to those who wish to maintain it. The relationship between changes in productivity and changes in size is a strong one, in terms of the coefficients of both correlation and regression. To explain it as reflecting a causal connection between size and productivity requires, therefore, that this connection also should be strong. If it were so strong it seems unlikely that it would be so swamped by past history as is implied by the low correlation coefficients of Table 14.

I conclude, therefore, that there is some positive evidence for the second of the hypotheses with which this chapter is concerned; namely, that the changes in relative firm size which have been shown to be characteristic of all the industries examined are not random but associated with (caused by) relative efficiency.

The evidence on this second proposition was less conclusive than that on the first. That on the third is weaker once more.

Table 15

AVERAGE CORRELATION BETWEEN CHANGES IN SIZE AND PRODUCTIVITY IN DIFFERENT GROUPS OF TRADES

Degree of concentration	*Industries having restrictive agreements*	*Industries without serious restrictive agreements*	*Total*
High . . .	+·69	+·35	+·52
Medium . . .	+·56	+·37	+·49
Low . . .	+·65*	+·64	+·64
Total . . .	+·61	+·50	+·55

Note. The wallpaper industry was excluded in striking the averages. The compartment marked with an asterisk contains only one entry.

The dispersions about the averages in the different cells of Table 15 are again too great, and the number of trades too small, for any firm conclusions to be drawn from the small differences

in the degree of correlation. It is perhaps significant that the lowest correlations are in blast furnaces, dyes, glass and wool. All of these are industries in which, either through integration or conglomeration, a substantial number of the firms are likely to be operating in other industries also, so that no strong relation between efficiency and growth is to be expected. To this extent the results are consistent with theoretical expectations. By and large, however, the conclusion must be that the data available do not themselves establish the proposition that the efficacy of the transfer mechanism varies with the degree of concentration and restriction.

Chapter XVI

ON CHANGES IN EFFICIENCY

I SUGGESTED in Chapter II that the second major aspect of the public interest, against which the virtues or vices of various rules of the game should be tested, was the rate of increase of industrial efficiency. This issue is the most difficult of all on which to bring empirical evidence to bear.

Some of the difficulties of principle in devising a measure of changes in efficiency were discussed in Chapter IV. The difficulties of practice are even greater. The problem of new products or changes in quality, for example, may in principle be ignored over short periods or dealt with by chain-index methods over longer ones. But in practice my period is long, and even the most rudimentary data are available only for its initial and terminal points. Moreover, even if a numerical measure of changes in efficiency can be constructed, there are immense difficulties of interpretation. There are fluctuations in the technical history of any industry—alternations between rapid advance and stagnation—which can only be characterised as random. To some extent "random" has, in this context, the meaning "not at present capable of being explained in terms of other factors". But it has also its proper meaning of "incapable in principle of causal explanation". Thus, even if we found that technical progress in one industry had been greater than that in another during a particular period it would be bold, and indeed foolhardy, to identify different rules of the game in the two industries as the cause. It might simply be that one industry was in the active and the other in the passive phase of its technical life-history. Given sufficient industries, some conclusion might be legitimate on the assumption that truly random factors would cancel out in the aggregate. But information on this scale is certainly not available at present.

It is idle to pretend, therefore, that any complete demonstration is possible. But this does not mean that we must remain totally stranded in the *a priori*. I concluded in Chapter VII that

the process of innovation in a market economy consists in the creation of cost (efficiency) dispersion, and that it is the tendency of cost dispersion to be eliminated which generates progress. I went on to suggest in Chapter IX that the uninhibited interaction of the transfer and innovation mechanisms will tend to generate a higher rate of progress than will more "managed" systems. This last claim must be left unproven by statistical evidence. But we can make some rudimentary tests of the proposition that concentration or restrictive agreements will inhibit the working of the innovation mechanism, and of the prior proposition that there is an innovation mechanism. If the statistically unproven is accepted, a demonstration of the two subsidiary statements will itself justify the conclusion that such agreements or industrial structures are detrimental to the public interest.

I again take changes in productivity as an index of changes in efficiency. The coefficient of variation of changes in productivity $\left(\frac{\sigma_\lambda}{\bar{\lambda}}, \text{ where } \lambda=\frac{p_{48}}{p_{35}}\right)$ provides a suitable measure of the extent to which relative efficiencies change over time. Small values for the coefficient of variation indicate that the change in productivity in most firms is similar, and large values that there are large variations from firm to firm in the extent to which productivity changes. Table 16 contains the coefficients for the sixteen industries.

There is evidence here of considerable flux, but not of any systematic tendency for the more concentrated or managed industries to be more rigid than the less. Indeed, the most striking feature of the table is the limited variation of the coefficient from industry to industry.[1]

In any event, however, the coefficients cannot themselves establish that an innovation mechanism such as I have postulated is operating. Changes in relative productivities form a necessary but by no means sufficient condition for this. For it is an essential characteristic of the innovation mechanism that, over time, the lowly be exalted and the mighty put down. We need, in other words, to see how far it is true that the biggest advances in

[1] The average coefficient of variation for all the industries is 51 and the coefficient of *their* variation 26.

Table 16

COEFFICIENTS OF VARIATION OF PRODUCTIVITY CHANGES (1935-1948) IN DIFFERENT TRADES

Trade	*Coefficient of variation* × 100
Wallpaper	—
Dyes	55
Margarine	64
Wool	81
Blast furnaces	51
Batteries	49
Boots and shoes	39
Rubber	76
Leather	63
Textile machinery	39
Vehicles	46
Cotton	42
Glass	43
Canvas goods	41
Cement	45
Wires and cables	36

Note. The coefficients of variation of $\frac{p_{35}}{p_{48}} = \lambda$ is $\frac{\sigma_\lambda}{\bar{\lambda}}$. The distribution of λ in the wallpaper industry is such as to make σ_λ and $\bar{\lambda}$ of very dubious value.

productivity are made by those whose productivity initially is lowest. Table 17 presents the evidence on this.

The fact that the sign of all coefficients in Table 17 is negative means that the innovation process is not, so to speak, explosive in its effects; relative productivities do not, in general, become more and more widely dispersed as time goes on. On the contrary, the implication is that the mechanism works as I have suggested, as an offset to the transfer mechanism, which prevents the latter from inexorably pushing every industry towards the ultimate goal of high concentration.

It may be seen from Table 18 that there is no systematic

Table 17

RELATIVE PRODUCTIVITY IN 1935 AND PRODUCTIVITY CHANGES, 1935 TO 1948

Trade	*Coefficient of correlation between* $p_{35}/\bar{p}_{35}$ *and* p_{48}/p_{35}	*Coefficient of regression of* p_{48}/p_{35} *on* $p_{35}/\bar{p}_{35}$
Wallpaper	—	—
Dyes	−·15	−0·3
Margarine	−·62	−1·6
Wool	−·21	−0·7
Blast furnaces	−·14	−0·4
Batteries	−·68	−1·7
Boots and shoes	−·72	−3·3
Rubber	−·43	−2·4
Leather	−·42	−4·1
Textile machinery	−·39	−1·6
Vehicles	−·73	−3·3
Cotton	−·64	−2·0
Glass	−·49	−1·4
Canvas goods	−·61	−2·2
Cement	−·49	−1·7
Wires and cables	−·75	−1·8

Table 18

RELATION BETWEEN RELATIVE PRODUCTIVITY AND CHANGES IN PRODUCTIVITY IN DIFFERENT GROUPS OF TRADES

Degree of concentration	*Coefficient of—*	*Industries having restrictive agreements*	*Industries without serious restrictive agreements*	*Total*
High	(*a*) Correlation	− ·58	− ·38	− ·48
	(*b*) Regression	−1·7	−1·6	−1·7
Medium	(*a*) Correlation	− ·50	− ·43	− ·48
	(*b*) Regression	−1·8	−3·3	−2·1
Low	(*a*) Correlation	− ·61	− ·49	− ·52
	(*b*) Regression	−2·2	−3·1	−2·9
Total	(*a*) Correlation	− ·54	− ·45	− ·49
	(*b*) Regression	−1·8	−2·9	−2·3

Note. Since regression coefficients have no significance when correlation coefficients are small, those for dyes, wool and blast furnaces have been excluded in forming the averages for regression coefficients.

tendency for the degree of correlation to vary with the degree of concentration, and there is considerable variation within each cell. But the correlation coefficients alone cannot measure the degree to which the structure of relative productivities in different industries has been transformed. For this we need to look also at the regression coefficients. An industry in which changes in productivity were perfectly correlated with initial relative productivities, but in which differences in productivity changes were only small (the regressions weak), would clearly exhibit smaller changes in its structure of relative productivities than would one in which, although the correlation were weaker, differences in productivity changes were larger (the regression was strong).

Although there are again substantial differences within the cells of the table, there is a fairly systematic tendency for the strength of the regression to vary with the degree of concentration and restriction. In other words, the data are at least not inconsistent with our theoretical conclusion that the power of the innovation mechanism will be weaker the greater the degree of concentration and restriction.

The same conclusion may be demonstrated more dramatically by correlation of productivity in 1935 with that in 1948, as in Table 19. Given the imperfections of productivity as a measure of efficiency we need not be unduly disturbed by the fact that

Table 19

COEFFICIENTS OF CORRELATION BETWEEN PRODUCTIVITY IN 1935 AND 1948

Industry	$r_{p_{35}p_{48}}$	*Industry*	$r_{p_{35}p_{48}}$
Wallpaper	—	Leather	·32
Dyes	·61	Textile machinery	·43
Margarine	·39	Vehicles	·05
Wool	·53	Cotton	·13
Blast furnaces	·52	Glass	·56
Batteries	·51	Canvas goods	·16
Boots and shoes	—·06	*Cement*	·56
Rubber	·53	Wires and cables	·46

virtually all the coefficients are positive. The significant fact is that the tendency for firms with relatively high productivity in 1935 to have high productivity in 1948 also is weakest in those industries where concentration and restriction have made the smallest headway.

Chapter XVII

CONCLUSIONS

I ENDED Chapter I with a quotation fr[illegible]eynes on the nature of economics. It is because economic[illegible]s he described that few books on economic subjects have a very lengthy chapter on conclusions. This book is no exception.

I cannot pretend to have moved my subject very far towards the "received ideas" end of the spectrum of opinion. Moreover, I have in one sense been overtaken by events. The Restrictive Trade Practices Act became law while this book was being written. By introducing the Bill the Government acted as if the undesirability of private rules of the game were a received idea, and Parliament, by legislating, confirmed that the idea was in fact widely received. It might seem then that the time for argument and analysis is past.

I think that this is not so. The law is what the Courts say it is, and the process of judicial interpretation and application of the will of Parliament is only about to start. Parliament has defined the restrictive agreements which it has prohibited, but the translation of words into working rules will be made by the practice of the Restrictive Practices Court. Moreover, Parliament has explicitly delegated to the Court a task of *economic* judgment. Private rules of the game are permissible if certain defined consequences would follow from their absence and, in the view of the Court, these would damage the public interest more than would the evil effects which Parliament must be presumed to think in general follow from restrictive agreements. I do not see how the Court can discharge this function without taking a view on both the nature and the magnitude of the damage which normally results from restrictive agreements. This in turn requires a view of the nature of the competitive process.

I have tried to provide such a view of the nature of the process, albeit in the crudest of outline forms. My empirical material does not establish, nor do I know, how seriously the sorts of restrictions practised in the United Kingdom have interfered

with the market mechanisms I have described. I believe, however, that the framework of thought I have presented offers scope for further work which would not only refine the outline but also begin to shed some light on the size of the problem I have been examining.

The analysis has much less relevance for the residuary problem of monopoly with which the reconstituted Monopolies Commission is now concerned. Monopoly can be defined only in terms of products and markets, whereas my classificatory principle here has been the technique and the industry. It is perfectly possible, though perhaps unlikely, that a world in which all firms were monopolists should have, on my definition, neither concentration nor restrictive agreements. The moral I draw from this is not that the model is inadequate for its purpose but that a further model is needed for thinking about monopoly.

Appendix

A NOTE ON SOME FURTHER CORRELATIONS

THE economic research worker invariably spends a great deal of his time in discovering the absence of relationships. In the earlier stages of my work I tested for correlation between a number of variables which I then thought likely to show some. I failed to find any and, indeed, should now be seriously embarrassed had I succeeded. Nevertheless, I think the negative results should be recorded. Absence of correlation may sometimes be as illuminating as its presence and, at least, I may so spare others a fruitless traverse of avenues I have already explored.

2. The following quantities will be referred to by numerical subscripts as follows:

(1) Net output of a firm in 1935
(3) Ratio of net output of a firm in 1948 and 1935
(4) Net output per head in a firm in 1935
(5) Net output per head in a firm in 1948
(6) Mean of net output per head in 1935 and in 1948 corrected for price changes

3. The coefficients of correlation calculated are shown in the Table. There are gaps in the table because various lines of enquiry were suspended when testing had shown them to be fruitless.

Industry	*13*	*14*	*15*	*16*	*34*	*35*	*36*
Dyes	...	...	−·06	−·05	−·27	...	−·07
Rubber	−·14	+·08	+·11	+·11	−·26	+·22	+·10
Glass	−·17	+·13	−·06	...	−·18	...	−·07
Cotton	−·38	+·62	−·11	+·17	−·26	+·29	+·12
Batteries	−·12	+·13	−·07	+·11	−·61	−·25	−·46
Cement	−·16	+·46	—	—	−·51	+·18	—
Wires and cables	—	—	—	—	−·61	−·24	—
Margarine	−·09	+·84	—	—	−·21	−·18	—
Boots and shoes	−·17	+·16	—	—	−·50	+·41	—

Industry	*13*	*14*	*15*	*16*	*34*	*35*	*36*
Vehicles . .	−·17	+·05	—	—	−·19	+·66	—
Canvas sacks .	−·24	+·28	—	—	−·26	+·42	—
Textile machinery	−·11	+·29	—	—	−·23	+·44	—
Wool . . .	−·09	+·14	—	—	−·32	+·15	—
Leather . .	−·18	+·40	—	—	−·46	+·51	—
Blast furnaces .	−·18	+·03	—	—	−·28	+·17	—

... means less than ·01.

Chapter XVII

CONCLUSIONS

I ENDED Chapter I with a quotation from Keynes on the nature of economics. It is because economics is as he described that few books on economic subjects have a very lengthy chapter on conclusions. This book is no exception.

I cannot pretend to have moved my subject very far towards the "received ideas" end of the spectrum of opinion. Moreover, I have in one sense been overtaken by events. The Restrictive Trade Practices Act became law while this book was being written. By introducing the Bill the Government acted as if the undesirability of private rules of the game were a received idea, and Parliament, by legislating, confirmed that the idea was in fact widely received. It might seem then that the time for argument and analysis is past.

I think that this is not so. The law is what the Courts say it is, and the process of judicial interpretation and application of the will of Parliament is only about to start. Parliament has defined the restrictive agreements which it has prohibited, but the translation of words into working rules will be made by the practice of the Restrictive Practices Court. Moreover, Parliament has explicitly delegated to the Court a task of *economic* judgment. Private rules of the game are permissible if certain defined consequences would follow from their absence and, in the view of the Court, these would damage the public interest more than would the evil effects which Parliament must be presumed to think in general follow from restrictive agreements. I do not see how the Court can discharge this function without taking a view on both the nature and the magnitude of the damage which normally results from restrictive agreements. This in turn requires a view of the nature of the competitive process.

I have tried to provide such a view of the nature of the process, albeit in the crudest of outline forms. My empirical material does not establish, nor do I know, how seriously the sorts of restrictions practised in the United Kingdom have interfered

with the market mechanisms I have described. I believe, however, that the framework of thought I have presented offers scope for further work which would not only refine the outline but also begin to shed some light on the size of the problem I have been examining.

The analysis has much less relevance for the residuary problem of monopoly with which the reconstituted Monopolies Commission is now concerned. Monopoly can be defined only in terms of products and markets, whereas my classificatory principle here has been the technique and the industry. It is perfectly possible, though perhaps unlikely, that a world in which all firms were monopolists should have, on my definition, neither concentration nor restrictive agreements. The moral I draw from this is not that the model is inadequate for its purpose but that a further model is needed for thinking about monopoly.

Appendix

A NOTE ON SOME FURTHER CORRELATIONS

THE economic research worker invariably spends a great deal of his time in discovering the absence of relationships. In the earlier stages of my work I tested for correlation between a number of variables which I then thought likely to show some. I failed to find any and, indeed, should now be seriously embarrassed had I succeeded. Nevertheless, I think the negative results should be recorded. Absence of correlation may sometimes be as illuminating as its presence and, at least, I may so spare others a fruitless traverse of avenues I have already explored.

2. The following quantities will be referred to by numerical subscripts as follows:

(1) Net output of a firm in 1935
(3) Ratio of net output of a firm in 1948 and 1935
(4) Net output per head in a firm in 1935
(5) Net output per head in a firm in 1948
(6) Mean of net output per head in 1935 and in 1948 corrected for price changes

3. The coefficients of correlation calculated are shown in the Table. There are gaps in the table because various lines of enquiry were suspended when testing had shown them to be fruitless.

Industry	*13*	*14*	*15*	*16*	*34*	*35*	*36*
Dyes	...	...	−·06	−·05	−·27	...	−·07
Rubber	−·14	+·08	+·11	+·11	−·26	+·22	+·10
Glass	−·17	+·13	−·06	...	−·18	...	−·07
Cotton	−·38	+·62	−·11	+·17	−·26	+·29	+·12
Batteries	−·12	+·13	−·07	+·11	−·61	−·25	−·46
Cement	−·16	+·46	—	—	−·51	+·18	—
Wires and cables	—	—	—	—	−·61	−·24	—
Margarine	−·09	+·84	—	—	−·21	−·18	—
Boots and shoes	−·17	+·16	—	—	−·50	+·41	—

Industry	*13*	*14*	*15*	*16*	*34*	*35*	*36*
Vehicles . .	−·17	+·05	—	—	−·19	+·66	—
Canvas sacks .	−·24	+·28	—	—	−·26	+·42	—
Textile machinery	−·11	+·29	—	—	−·23	+·44	—
Wool . . .	−·09	+·14	—	—	−·32	+·15	—
Leather . .	−·18	+·40	—	—	−·46	+·51	—
Blast furnaces .	−·18	+·03	—	—	−·28	+·17	—

... means less than ·01.

INDEX