UNIVERSAL
LIBRARY
OU_148591
UNIVERSAL
LIBRARY

OSMANIA UNIVERSITY LIBRARY

Call No 330.1/K26T. Accession No. 29355.

Author Kirschbaum B.S.

Title Theory of Economic Change. 1948

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

THE THEORY OF ECONOMIC CHANGE

THE THEORY OF ECONOMIC CHANGE

BY

B. S. KEIRSTEAD

William Dow Professor of Political Economy in McGill University

TORONTO
THE MACMILLAN COMPANY
OF CANADA, LIMITED

1948

Copyright, Canada, 1948
By
THE MACMILLAN COMPANY OF CANADA LIMITED

All rights reserved—no part of this book may be reproduced in any form without permission in writing from the publisher, except by a reviewer who wishes to quote brief passages in connection with a review written for inclusion in a magazine or newspaper.

Printed in Canada
by
The Hunter-Rose Co. Limited, Toronto

PREFACE

This is a pretentious book, I fear, in the sense that it seems to pretend to be saying something new, whereas it is really not saying anything new at all, but merely arranging what has been said before in a new form. I am also painfully aware that this is scarcely a book at all, in the usual sense, but rather a collection of more or less related essays. Indeed, I suspect that my preface must consist of a defence of my decision to publish in this form, rather than to rewrite the whole thing as a shorter, integrated thesis. I believe I could have done this, and have produced a much neater, sharper argument. Aesthetically it would have been more satisfying, and the result would probably have been both easier and more agreeable to read.

For two reasons I have not done so. One reason is a lazy proclivity which makes me reluctant to undertake the laborious task of rewriting when I might be going on to more interesting things. The other is that I believe there is some value for the serious student in this kind of "work in progress" book, with all its hesitations, digressions and circuitous development. After all, I have been working with these ideas for seven years, and I must assume, as a teacher, that a record of my thoughts, though difficult, devious and diffusive, should be interesting and provocative to a student.

The student I have in mind is not, of course, the beginner. This book is for the honours student and the graduate student. Because I have students in mind, I have let stand, against the advice of many friendly readers, the introductory chapters on method and the digressions of Part II and of the passages on the demand function, the production function and the review of the literature on the theory of the firm. "Everyone knows this," one friend commented on Chapter X. But he was a learned man. Quite a few people, I hope, who are reasonably competent in economics will nevertheless be grateful to have this literature reviewed for them. It is for such students, rather than those like my friendly critic, that I am writing. They are, I suspect, the majority among honours and graduate

students in this country. Because I have them in mind I have tried, also, to keep a promise I made to my own graduate students to keep all mathematics out of the text. This has resulted in some laborious circumlocutions, but, whatever may be said of the footnotes, I think the promise has been kept, or very nearly so, in the main text. It is, then, to the advanced student rather than to the undergraduate on the one hand or the professional economist on the other, that this book is directed.

McGill University,
Montreal,
November 1947.

B. S. KEIRSTEAD.

ACKNOWLEDGEMENTS

I have to acknowledge the assistance given me by my friends and colleagues at McGill and by my graduate students. I have also had the kindly advice of Dr. Harold Innis, of the University of Toronto, of Dr. H. G. Littler of Montreal, and of Mr. Nathan Keyfitz of the Dominion Bureau of Statistics. To all these I express my gratitude. To many, the extent of my indebtedness is indicated by footnotes.

I wish further to express my thanks to the various editors and publishers who have waived copyright and permitted me to quote extensively or in brief from copyright works.

Specifically I wish to acknowledge with thanks, permission to quote, from the works of the following authors:

To Dr. L. Richter, and the Dalhousie Institute of Public Affairs, for permission to reproduce in large part Chapter III from my *Economic Effects of the War on the Maritime Provinces of Canada.*

To Macmillan & Company Limited, of London, and to The Macmillan Company of New York, for permission to quote from the works of Alfred Marshall, W. S. Jevons, Joan Robinson, and shorter passages from N. K. Smith's translation of Kant's *Critique* and from J. M. Keynes's *General Theory.*

To the University of Toronto Press and the editors of the *Canadian Journal of Economics and Political Science*, the *Canadian Historical Review*, the *Transactions of the Royal Society of Canada*, and the *Canadian International Journal*, for permission to quote *in extenso* passages from articles of mine published in these journals.

To W. W. Norton and Co. of New York, for permission to quote from *Fiscal Policy and Business Cycles*, by A. H. Hansen.

To Staples Press, Ltd. for permission to quote at length from *The Pure Theory of Production* by Sune Carlson.

To the University of Chicago Press to quote from *The Theory and Measurement of Demand* by Henry Schultz, and to the Press and the editor of the *Journal of Political Economy* to quote from

"Realism and Relevance in the Theory of Demand" by Frank Knight in No. 4, Vol. 52 of the *Journal*.

To the Oxford University Press for permission to quote from *Kant's Theory of Knowledge* by H. A. Prichard.

To the Harvard University Press for permission to quote from *The Theory of Economic Development* by J. A. Schumpeter.

To Ginn and Co. for permission to quote from *Social Causation*, by R. M. MacIver.

To the McGraw-Hill Book Company for permission to quote from *Business Cycles* by J. A. Schumpeter, and also from *A History of Mechanical Inventions* by A. P. Usher.

The source of all quotations is fully indicated in footnotes.

May I finally express my warm appreciation of the patience and careful assistance of the Editorial Department of the Macmillan Company of Canada Limited.

B. S. K.

CONTENTS

PAGE

PREFACE v

INTRODUCTION 1

PART I—SOME FUNDAMENTAL CONCEPTS

CHAPTER

I CAUSALITY 11
1. The Problem of Method 11
2. The Idea of Cause 15
3. "Meaningfulness" as an Element in Social Causation 22

II CAUSALITY. FREEDOM AND THE SOCIAL FUNCTION OF ECONOMICS 27
1. Causality and Freedom 27
2. The Social Function of Economics 32

III EQUILIBRIUM, PROCESS, TIME 38
1. Equilibrium and Process 38
2. Time 49

PART II—THE GENERAL THEORY OF ECONOMIC CHANGE

IV THEORIES OF GIVEN CAUSES 63
1. Given Causes 63
2. Some Theories Based on Given Causes 68
Adam Smith 69
David Ricardo 77
Karl Marx 85
J. A. Schumpeter 94

V THE GENERAL THEORY OF ECONOMIC CHANGE 99

PART III—MODELS OF THE GENERAL ECONOMY

VI INTRODUCTION TO PART III 107

VII Population Changes 113
Model I 113
Model II 129

VIII Innovations 132
1. Some Definitions and General Observations 132
2. The First Model 139
3. The Second Model 143
4. A Summary 145
5. Institutional Change and Welfare Effects 148
6. Relation to Short-Run Cycles 156

IX Population and Innovations 161
1. Secular Stagnation 161
2. The Demand Function 166
3. The Production Function 174
4. Synthesis 179

Appendix to Chapter IX—The Theory of Demand. 183
1. The General Theory of Demand 183
2. The Measurement of Changes in Demand 191
3. Conclusions 200

Part IV—MODELS OF THE FIRM

X The Traditional Theory of the Expansion of the Firm 209
1. The "Expansion Path" 210
2. Ambiguity in the Theory of the Expansion Path .. 226
3. The Envelope Curve 229

XI The Temporal Process of the Firm 239
1. Process under the Usual Assumption of Profit Maximization 239
2. Profit Maximization in Time and Senility of Firms 254
3. Conclusions 261

PART V—TEMPORAL SHIFTS IN LOCATION—THE CASE OF THE MARITIME PROVINCES

INTRODUCTORY NOTE 267

XII THE THEORY OF LOCATION EQUILIBRIUM 269

1. The Theory of Location Equilibrium 270
2. Technical Change and the Small Firm in the Maritimes 277
3. Location Equilibrium—The Maritimes and the Central Provinces 286

XIII THE COMPARATIVE RATES OF DEVELOPMENT OF CERTAIN MARITIME INDUSTRIES IN THE INTER-BELLA PERIOD .. 291

1. Some General Considerations 291
2. Number and Size of Establishments 293
3. Productivity and Wages 300
4. Effects of the Second World War—General Conclusions 310

PART VI—AREAS OF DECISION AND POLICY

XIV INTRODUCTION TO PART VI 317

1. A Summary of the Argument: Presentation of the Problems of Part VI 317
2. A Digression on Some Matters of Politics 326

XV POPULATION AND INVESTMENT 332

1. Population Transfer and Regional Development within a National Economy 332
2. Economic Change and the International Order ... 339

XVI THE BUSINESS CYCLE AND ACCUMULATED SURPLUSES 348

XVII MONOPOLIES AND CARTELS 359

XVIII CONCLUDING REMARKS 375

INDEX 381

INTRODUCTION

This book is concerned with the extension of analytical methods to problems of economic change. Traditionally, all sciences have their well-defined boundaries beyond which they do not push their enquiries, accepting rather, as so much given data, the conclusions of those subjects which lie beyond the frontiers, and using these as the basis for their own proper enquiries.[1] Sometimes, the boundaries of a science are widened, as when the need for a revision beyond the frontier becomes evident as the result of new discoveries within, so to speak, the metropolitan area, but in the case of economics, the trend has been an ever-increasing restriction of the field of pure economic inquiry and a constantly sharper delineation of the borders of history, politics and sociology. It is part of our purpose to inquire if this process of specialization and restriction has gone too far, and if there are not certain questions to which economic analysis ought to be extended but which lie beyond what is ordinarily regarded as its scope. In the first part of this work, we shall be concerned with the problem of the appropriate methods for such an extension of economic inquiry. In the subsequent parts we shall deal with the application of these methods to selected processes of the dynamic economy.

Of the two methods of approaching an economic problem, that of analysis is a sure and certain one, but severely restricted, capable of asserting certain general relations which must obtain if given assumptions are true, but often incapable of demonstrating that the assumptions are true. Thus, to choose an illustration later to be discussed, if production functions are linear and of the first degree, the total produce, it can be shown, must be exhausted by the distributive shares, but it cannot be shown, by ordinary economic analysis, that production functions are linear and of the first degree. This work of limitation, of

[1] One must always remember Hartman's warning, in *Kategorielehre*, against mistaking "the boundaries of one's own science for the boundaries of human knowledge in general."

which the stultifying effects have been grossly exaggerated by some critics, is nevertheless sufficiently restrictive to prevent the development of the pure theory of economics into a predictive science. For that, a criterion of the validity of the assumptions in the real and temporal world would be necessary.

The method of economic history, by contrast, is naturally designed to relate causal sequences in time and hence to provide a basis for value and policy judgments. The purists of economic analysis who have divorced the solution of the problem of allocation—how, in fact, scarce resources are allocated among various uses—from the examination of the welfare effects of the various institutions of the allocative or distributive machinery, have done economic science the great service of freeing it from political preoccupation and special pleading, and they have won for its professors authority and prestige. Yet "the practical significance of economic theory", as Professor Knight has said, "is in the field of social action . . ."[2] Our studies, however rarefied, are aimed ultimately at informing social policy. This demands of economics that it become, in a certain sense, a predictive science. No divorce of economics from policy can ever be complete or permanent, and, if it could, it would be unhealthy. Economic policy would then be unenlightened and unstudied, and for economic studies such an enforced intellectual celibacy would be barren in results and would lead to a niggling, fussy, old-maidish life, devoted to the elegant arrangement of unused and unusable furniture. Vitality in economics, as in the other inquiries of the human spirit, comes from the insemination of humanistic values. The illumination of the mind and the ethical progress of the race are coincidental, and, in purpose, conjoint. The vitality of the classical and Marxian economics in the past, and of Keynesian economics today, is testimony to the need for economic inquiry to be concerned with welfare, to be directed ultimately to the policy judgment. A question which is implicit in this present work is whether the various techniques of the economic analyst and the economic historian can be combined to this end.

[2] F. H. Knight, "Realism and Relevance in the Theory of Demand", *Journal of Political Economy*, Vol. 52, No. 4, Dec., 1944, p. 311.

Whereas the method of the economic analyst is rigorous, if restrictive, that of the economic historian is apt to achieve freedom only by the sacrifice of rigour and certainty.[3] Although some historians talk of an "historical determinism" that is reminiscent of Calvin, as though there were a rigid determination from a single reason that ran through the course of history, it is doubtful if the historical approach, by itself, can ever produce generalizations adequate as a basis for policy. Hegel, indeed, did teach that there was a necessity in the process of history; but surely no one would seriously claim for the triads of his dialectical system the same necessity that is observed in a deductive logical or mathematical system. The followers of Marx and Engels, also, have talked of an "economic determinism" observable in history, but they do not mean to designate a determined sequence, in the proper sense of the term, at all, but rather a simplified technique of interpreting historical events in terms of probable cause. Marx and, even more, Engels, were careful to show that they were aware that they were dealing with hypotheses of historical cause and were not following Hegel in developing a pseudo-deterministic system.[4] Nevertheless, though fruitful, the Marxist hypothesis has had too much claimed for it by the dogmatists. "In all the voluminous writings of Karl Marx," says Professor MacIver,[5] "there is nowhere any attempt to test the doctrine of the 'materialistic interpretation of history'. Constantly we are bidden to view events and situations in the light of doctrine. We are told, for

[3] " . . . the effectiveness of the application of mathematics varies in the two fields (economic analysis and engineering) and may make for a divergence rather than convergence in the study of economic phenomena as a whole." H. A. Innis, "On the Economic Significance of Culture", in "The Tasks of Economic History", Supplement IV, of the *Journal of Economic History*, Dec. 1944, p. 84.

[4] Karl Marx seems to have been unsure of himself. He admitted the hypothetical and conjectural nature of "dialectical materialism," yet speaks of himself as setting the dialectic of Hegel "right side up," and often uses dogmatic and intolerant language in claiming for his economic determinism a complete and rigorous necessity. Marx was comparatively ignorant of logic and philosophy and the truth is he probably was unable to distinguish between strict determinism and probable cause. He recognized the value for his purposes of the Hegelian method and appropriated it, adapting it to his devices and then claiming—with an intolerance, characteristic rather of the religious leader than the scholar—an absolute authority for the maxim.

[5] R. M. MacIver, *Social Causation* (New York, 1942), p. 117.

example, that 'the ideas of freedom of thought and of religion express only the dominance of free competition within the sphere of knowledge' (Communist Manifesto). The analogy is ingenious, if left as such, the 'only' turns it into dogma. It forbids us to view a many-angled situation from any angle but one." In no other sphere of learning could this assertion of cause be accepted as anything more than an hypothesis. Certainly it cannot approach the rigorous demonstration of a determined sequence.

Even non-Marxist historians, claiming rather less for their method, are prone to assert cause on the slightest evidence.[6] The most interesting, and often the most useful, historical writing, is frequently that which proceeds boldly to sketch out a causal thesis and to interpret events in its light.[7] The multiplicity of events in history poses a very real problem of method for the historian. If he is to avoid mere chronology[8] he must select and order his processes in time, and this selection and arrangement require, to be intelligible, some causal hypothesis, some reason for relating events in a particular pattern. Yet,

---

[6] Compare, for example, the various "causes" of the American Revolution. Popular textbooks find the "cause" in British tyranny; one historian finds it in a complex of the "colonial mind"; another in the growing away of the colonies and a desire for equality of status; another, in the growth of an American bourgeois class; still others in various manifestations of eighteenth-century British imperialism and colonial reaction. It should not be difficult to reproduce for history a catalogue of alleged single causes of an event paralleling Professor MacIver's catalogue of the alleged (by sociologists) "causes" of criminality.

[7] Thus, Professor Croce's brilliant interpretation of the nineteenth century in terms of the growth of the liberal idea. See *History of Europe in the 19th Century*, tr. Henry Furst (New York, 1933).

An interesting experiment in the bold hypothesis is Professor J. U. Nef's attempt to relate the rate of industrial progress to national wars. See his "Wars and the Rise of Industrial Civilization, 1640-1740", *Canadian Journal of Economics and Political Science*, Vol. 10, No. I, Feb. 1944, pp. 36-79. He speaks significantly in reference to historical method, of the "task of arranging such materials as are available into a durable pattern . . . " "The significance of economic history in all this," H. A. Innis writes, "is shown in its concern with long-run trends and its emphasis on training in a search for patterns rather than mathematical formulae". H. A. Innis, "On the Economic Significance of Culture", in "The Tasks of Economic History", Supplement IV, *Journal of Economic History*, Dec. 1944, p. 96.

[8] History, as a mature discipline, Bury says, could only come with the idea of human development: "This is the great transforming conception, which enables history to define her scope". He traces the idea to Leibniz but does not discover it as a "governing force" until the nineteenth century. See J. B. Bury, "The Science of History" in *Selected Essays*, ed. H. Temperly (Cambridge, 1930).

it is clear, to claim for history the capacity to prove determination is absurd; the proof of cause is difficult, and seldom really attempted.

Social scientists in other fields, notably in sociology, are becoming increasingly preoccupied with problems of method, and they are, as a by-product of their labours, so to speak, producing questions for the economist, questions which are a challenge to those who believe that a political economy is a possibility. Economists, it is true, have themselves refined their concepts and clarified their method, but it must be confessed that they have done so with the especial intent of defending the abstract method against wrong-headed criticisms and of removing misunderstanding about what the economic theorist was really trying to do.

The challenge of the sociologists,[9] however, is of a different order. They are aware that the techniques of the economist are perfectly valid as instruments for the analysis of a static order. They are, perhaps, not aware how valuable sound analysis of a static order may be, even for the comprehension of change, but they are certainly dubious of the ability of "equilibrium" theory, based on more or less rigorous determination, to explain (in terms of cause-and-effect) dynamic shifts of equilibrium. Of course, all this is not entirely new to economists who will regard with suspicion all that smacks of historicism,[10] its ill-advised adventures in the field of factual research, and its easy seduction by the vampires of government and business. Yet it does seem that if economic analysis is to avoid the fate of aridity, some return must be found to the old traditions of political economy, some attempt made to relate the findings of analysis to the problems of policy which rest in a temporal and dynamic matrix. That this attempt is already being made is witnessed by the current output of articles and books attempting to measure demand and production functions over time,

---

[9] See, for example, P. A. Sorokin, *Sociocultural Causality, Space, Time* (Durham, 1943); R. M. MacIver, *Social Causation* (New York, 1942); Felix Kaufman, *Methodology of the Social Sciences* (New York, 1944).

[10] Cf. Karl Popper, "The Poverty of Historicism: A Criticism of Historicist Methods", *Economica*, Vol. II, No. 43, Aug. 1944, pp. 119-137.

the development of an applied theory of location, and the contemporary interest in general equilibrium theory of a temporal sort.[11] All these extensions of economic analysis to problems of policy in a temporal order have been characterized by a loss of rigour and precision. Thus the now popular and exciting "general equilibrium" and fiscal theories of Lord Keynes and Professor Hansen which have been directly applicable to problems of war and post-war policy, are open to serious objections, particularly when presented by enthusiastic disciples as demonstrated beyond reasonable doubt. Insufficient attention has been paid by Keynesians to questions of method. An analytical technique has here been extended to temporal sequences without adequate and sophisticated inquiry as to the nature of time. Consequently, in the use of tax and national income formulae in the problems of policy, an improper optimism is generated, based on the failure to allow for other temporal and spatial phenomena, especially the mediacy of all equilibrium adjustments, regional leakages and rigidities and the effects of innovations. As Professor Myrdal says, perhaps too unequivocally, of Professor Hansen's calculations of the "inflationary gap" in the United States and the volume of building necessary to maintain full employment, ". . . these calculations are worthless. The analysis as a whole is altogether static. It is not at all concerned with how the whole procedure is to be brought about . . . by developments over a period of time."[12]

---

[11] Awareness of the need for a methodological clarification is apparent in the following quotation from one of the more eminent of contemporary American economists: "Turning from theoretical to practical difficulties, it is actually quite unrealistic to think of the functional (cause-and-effect) relations between economic variables as remaining unchanged through a real change in any independent variable. There is always a significant 'lag' in time; and in the meantime other things will not remain 'equal' . . . But the conceptual interpretation of the lag of consumption change behind price change involves a causal sequence bearing some analogy to the action of a force against a resistance in which inertia is combined with friction (opposing the redistribution of a flow among alternate paths). The nature of the concepts required for interpreting such phenomena calls loudly for investigation . . . This line of argument suggests a basis for a pure theory of economic fluctuations." F. H. Knight, "Realism and Relevance in the Theory of Demand", *Journal of Political Economy*, Vol. 52, No. 4, Dec. 1944.

[12] Gunnar Myrdal, "Economic Developments and Prospects in America". Address before the National Economic Society of Sweden, March 9th, 1944. Translated and distributed by the National Planning Association, Washington.

Such inadequacy is probably inevitable. It comes from the method of "comparative statics", which is a way of studying the disturbances of equilibrium that result from a shift in one of the parameters of the model. Now equilibrium is a "tool" concept. It is one of these abstract ideas that we use because it gives us a way of looking at things and relating them. It can be a very useful tool. When we are interested, for example, in periodic instability of the economy, it is a sound device to ask what governs the conditions of stable equilibrium. That is why the equilibrium concept gives results in the analysis of industrial fluctuations. When, however, our point of departure and our focus of interest are different, we may require other analytical tools. Now in this book I am interested in long-run structural changes in the economy and their effects on welfare. My point of departure is closer to that of the classical economists than to that of many of my contemporaries. My inquiry is into what changes the "wealth of nations", using wealth as Adam Smith used it. Moreover, since I believe that the concentration of economic power and the balance of that power are the most significant dynamic forces in the evolution of our social and political life, I want to see the structural changes in the economy which accompany any process of change. I do not find that a consideration of the changes in the aggregates of income, investment, spending and employment sufficient for my purpose. Hence the preoccupation with the structural changes, especially those in the structure of the market.

To make this kind of study, if it can be done at all, requires as a prolegomenon a consideration of the possibility of establishing firmly and exactly our methodological concepts. The analytical concept of ex-temporal determinism must be modified in favour of explanation in terms of cause-and-effect, but, if possible, a more precise and rigorous treatment of causality must replace the rather loose and hypothetical notion common in historical writings. Also, since we wish to observe sequences of events in time, one must try to establish a time concept that will have more content than the purely formal dimensional notion of time characteristic of a truly abstract system; but we cannot, as we shall see, permit an entirely causal notion of

time; if time is to be regarded as concrete rather than abstract, if it is recognized not formally as succession but in terms of its contents, of the events that succeed one another, economic time will have to be defined in terms of the nature of the events of an economic sort that are regarded as constituting time in their succession.

Hence the pattern of the present work. Part One is the prolegomenon we referred to, the discussion of the concepts of cause, time, equilibrium and change, the basic concepts of a theory of economic change. Part Two is a review of the general theory of economic change and a statement of the problem; Part Three is an attempt to develop a general theory of change in terms of the major causes of change; Part Four reviews the theory of the process of the firm and attempts to integrate this with the general theory of change; Part Five deals with the theory of changes in location equilibrium and includes a case study of a portion of the Canadian economy; and Part Six is devoted to some of the political and social implications of the foregoing argument.

PART I

SOME FUNDAMENTAL CONCEPTS

CHAPTER I

CAUSALITY

1. THE PROBLEM OF METHOD

The study of economic change is the meeting-place of the theoretic and historical methods. The historian studies changes as they have occurred in a particular time and place. To make sense of these events, to render them, as we say, intelligible, he has to select his events, arrange them in an understandable order and postulate a causal relation between them. In order to do this he must have formed explicitly or implicitly a *theory* of change. If he does not have some kind of a theory he cannot select and interpret events, he can only chronicle them. There is neither virtue nor significance in mere facts. "We shall not be able to go straight in economics by easy expedients, such as addressing ourselves 'directly' to the 'facts'. Surely we ought to know by now that facts without understanding are meaningless; and understanding is a matter of perspective and pattern, that is to say theory."[1] The theoretical economist must help in this interpretation of the past by providing the historian with good theories. By this I mean that vague, "common-sense" and sometimes emotionally biased interpretative theses must be replaced by conscious, refined and rational hypotheses.

To some extent this is done by the traditional economic analysis, which, though confined to static models, nevertheless reveals certain general relationships of economic quantities, the knowledge of which may be of importance in the understanding of real processes. Further, the economic theorists have developed a method which I shall identify by the phrase "comparative statics". This method consists in setting up a series of like models in the parameters of which small alterations are

[1] C. E. Ayres, *The Theory of Economic Progress* (Chapel Hill, North Carolina, 1944), p. vi.

made, and the consequent modifications of the positions of equilibrium subsequently noted. This method contributes to the understanding of certain real events when what we want to know about the system of real events is how stability is disturbed. When, for example, we are trying to understand industrial fluctuations, and our concern with these phenomena derives from the sense of instability they engender, we properly focus our attention on the conditions of general equilibrium in the system. For the analysis of such a problem the method of comparative statics is fruitful. When, however, we are interested in the long-run effects on welfare of structural changes in the economic system, or of certain basic forces acting on that system, the appropriateness of equilibrium analysis, either economic statics or comparative statics, is more doubtful. Like Adam Smith, we are inquiring into the causes of the wealth (and changes in the wealth) of nations. Now, in economic statics, cause is excluded from the models. The typical model rests on a series of assumptions which are its logical premises, and the study of the model is simply the drawing out by deduction of all the implications or *consequences* which are implicit if the premises are regarded as being true without contradiction. Quantities are *determined* within the model, but cause is excluded. Thus, in economic analysis the market situation of monopoly may be defined in terms of the elasticity of the average revenue curve of the firm. This definition, which in effect is a generalization of market conditions in which no competitor is able to substitute a rival product in the event of changes in selling price, forms the major premise in the logic of economic analysis. The minor premise is the common economic assumption that the *economic rationale* will not be satisfied by any position short of maximum profits. Any position of output and price which did not yield maximum profits would be an unstable one, because it would leave the entrepreneur an incentive to alter output and price. Only the output which satisfies the minor premise, the maximization of profit, may be regarded as stable or in equilibrium, in the sense of providing no incentive to change. Since the price-output point which satisfies the minor and major premises is the equilibrium point, it is the point which, under the con-

ditions assumed, the economist believes monopolies will in fact tend to approach. This point, with all logical propriety, is regarded as rigorously determined, and the demonstration is intended to have the universality and necessity of a theorem in Euclid.

Economic history, for its part, concerns itself with the quite different question of how monopoly comes about. What led Elizabeth to grant monopolistic charters to the chartered trading companies? Was it the demands of the Spanish war on the Privy Purse? Was it the personal avarice of the bourgeoise Queen? Was it the interest and influence of Cecil? Was it a desire to strengthen the Fleet by the encouragement of foreign trading? One school of economic history plunges behind these questions of personal motives and speaks in terms of broad historical forces, to which some writers refer in language which is suggestive of a necessity in history, independent of personal motive, impersonal, imperative and determinate. The Spanish monopoly of the new world "forced" English merchants to seek alternative outlets and to organize counter-monopolies. Similarly the later "rationalization" of British industry was "impelled" by the necessity of achieving scale economies to compete with German, Dutch and American trusts.

This "impulsion", however, differs, on analysis, from the logical necessity of a price theorem. Euler's theorem, borrowed from mathematics, is true and necessary in a sense that any historical "determinism" is not. Thus, if the production function is of the linear type $\lambda P = F(\lambda A, \lambda B, \lambda C, \ldots)$, when P is product and $A, B, C, \ldots$ the factors of production, the desired "exhaustion of product" equation of marginal productivity analysis $P = \left(\frac{\partial P}{\partial A} \cdot A + \frac{\partial P}{\partial B} \cdot B + \frac{\partial P}{\partial C} \cdot C + \ldots\right)$ follows at once.[2] The necessity of this proposition is that referred to by Kant when he speaks of the necessity of *a priori* propositions. "Experience," he says, "teaches us that a thing is so and so, but not that it cannot be otherwise . . . Experience never confers on its

[2] P. H. Wicksteed, *Coordination of the Laws of Distribution*, London School of Economics Reprint, No. 12 (London, 1932).

judgments true or strict, but only assumed and comparative universality, through induction. We can properly only say, therefore, that, so far as we have hitherto observed, there is no exception to this or that rule. If, then, a judgment is thought with strict universality, that is in such a manner that no exception is allowed as possible, it is not derived from experience, but is valid absolutely *a priori*."[3] The Euler theorem, for example, is applicable to the problem of distribution, and the consequent equation holds strictly and absolutely, and is true independent of all judgments based on experience, once the premise is admitted that production functions are homogeneous and linear functions of the first degree. If one wishes to deny—as one well may[4]—that typical production functions are of this general type, one simply decries the applicability of the theorem to the problem of exhaustion of the product; one does not deny the strict validity and universality of the theorem itself.

On the other hand the so-called "determination" discoverable in historical events is established by experience and rests on a single reading of events. It lacks even the degree of probability inherent in the usual "laws" of experimental science, which are based on repeated experiments under controlled conditions and checked by a series of readings. The determination that is alleged is no more than a hypothetical cause, probably sufficient at best, and not infrequently inadequately supported by evidence.

In comparative statics an alteration in the equilibrium may be imputed to an alteration in a parameter, and the latter may be called the cause of the former. This is not a study, however, of genuine causal relationships, or so it seems to me, because one does not study the succession of causal events, asking why they succeed one another in a definite way and in no other way. Rather one says, let us suppose a certain alteration to occur, what will be the necessary readjustments to a new position of equilibrium? This is, as I understand it, the method of Lord Keynes, in his *General Theory of Employment, Interest and Money.*

[3] Immanuel Kant, *Critique of Pure Reason*, Norman Kemp Smith, ed. and tr. (London, 1939), pp. 43-44.

[4] See Ch. IX.

Lord Keynes asks, in effect, what assumptions are necessary to define a general equilibrium system at full employment. He then shows the alterations in the assumptions which permit equilibrium of the system at other than full employment, and is consequently able to identify the differentiae of the two systems. Of course, this is an inquiry into change—the change in an equilibrium model—as far as it can be inferred. It does not, however, involve the examination of a dynamic causal system.

It is the latter kind of inquiry that I hope to attempt in this book. The problem, once posed, suggests the difficulties. Is there any possibility of a theoretic approach which is not that of comparative statics? Is anything possible other than intelligent and patient historical inquiry, which may succeed in isolating and even in generalizing about some unique processes? What, anyway, are we to mean by a causal system, what is the nature of economic causation and what can we identify as causes of economic change? These questions, I think, must be answered before we can know what we are trying to do and how we are to set about it.

2. The Idea of Cause

In the physical sciences the idea of cause, accepted commonly until very recently, was Hume's doctrine of invariable sequence. This notion of causality is empirical and carries no implication of necessity. In effect, we say if, when A is present, B is present, when A is changed, B is changed, when A is eliminated, B is eliminated, we have a causal relation. Our assertion of cause rests on repeated observation, and, in natural sciences, repeated tests under controlled conditions. The relationship is invariable so far as our observations go, and, on empirical grounds, we suppose it to be generally so; but we see no reason why it must be so; it lacks strict necessity and universality. This is the notion of causality, associated with the name of David Hume, characteristic of the methods of the empirical and mechanical sciences of the nineteenth century and held, by Karl Pearson in *Grammar of Science*, to be the only notion of cause pragmatically acceptable. We shall call this the doctrine of "*mechanical*

causation". It corresponds to the "why of invariant order" in Professor MacIver's scheme,[5] and, fairly closely, to the Aristotelian notion of "efficient causation".[6]

Now it is a vulgar error to suppose that it is this concept of causality that is carried over to the analysis of any historical complex. One discovers event B to follow event A and one infers a causal connection which would appear to rest on an observed *order* of the two events. But this is not strictly true. There are three important differences between the observed order of events in history and in nature. The first is obvious. In nature one is able to repeat the sequence, to observe under conditions so controlled as to exclude other possible or contributory causes. One is able to simplify and to repeat so as to to obtain invariability. In history even apparent repetition is illusory because of the inability to control conditions and because, further, the active elements in history being human agents, the fact of a similar event having occurred in the past influences human reactions to the similar pattern which is being "repeated". So, in history, one never has invariability or certainty in the observed sequences.

Again, in history, causality is teleological in the sense that it includes, usually, some idea of purpose or intent. When we say the competition for furs was the cause of the westward expansion of the whites in Canada and of the Anglo-French rivalry among the aboriginal tribes about the Great Lakes of the North American continent, our use of the causal concept implies a purpose (profit-winning) in the reaction of the human agent to the physical environmental challenge. There is no purposefulness in the reaction of gaseous volume to changes in pressure. Thus in history, and, indeed, in all cases where *social causation* is involved, we have present the active human principle of *motive*, which distinguishes the social from the physical causal nexus. When we ask *why* Great Britain went to war in 1939 our question is on a different level of discourse from the question "why did this gas expand in volume?" In the latter case the

[5] R. M. MacIver, op. cit. p. 24.

[6] Cf. Descartes: "The whole class of final causes is of no avail in physical (or natural) things". IV *Meditation* (Everyman ed., London, 1912), p. 113.

answer is a comparatively simple one: because, temperature being constant, pressure was reduced, and the explanation rests on an observed invariability of sequence. But in the former case, the sequence is a unique one and involves an action by an active and responsive, not a passive, agent. We answer, Great Britain went to war because His Majesty's government was bound by a treaty obligation to Poland, or because the British people saw that their own independence and security were involved with the successful resistance of the remaining Continental powers to German expansion, or because of rival economic imperialisms, and so on. Whatever our answer—and the fact that several answers may seriously be offered on more or less equally valid grounds is an indication of the difference between the causal hypothesis in science and history—it is clear that no answer is possible that does not impute a motive to a corporate human agent or a group of individual human agents.

Now, except on the crudest of mechanical behaviouristic hypotheses, the presence of motive in social causation gives, at one and the same time, a kind of necessity to social causation which is lacking in the concept of *mechanical cause* and yet weakens the general applicability of the concept and makes its application far more problematic than in the natural sciences. Thus we can say that if human motives are of thus and such an order, the expected subsequent action would be of such and such a nature. This hypothesis can then be extended to a sequence and if the order of events in the sequence corresponds to the expected order of the hypothesis we allege that we have a causal relation. Let us note that this order of cause is that of *sufficient reason*, which implies an element of necessity, in the sense that consequences are deduced from premises involving human motive patterns. The element of necessity, nevertheless, is lacking strict logical universality when applied to historical events, because it rests on postulates of simple motivation, whereas in history all motivation is exceedingly complex, thus the discovery of sufficient cause is never *proof*, in the strict sense, that event A, and event A only, is the *cause* of event B. That is why rival causal hypotheses can exist side by side as the explanation of a single historical sequence.

This brings us to the third point of difference. Mechanical causation is a continuous cause-effect sequence in a single line of like causal nexuses. It is a simple chain of exactly similar links, and it is an endless and unbroken chain, implying inevitably the Kantian antimony of postulating somewhere an uncaused cause, or of thinking of an infinite series of cause, itself without cause. Historical causation is of a more complicated pattern. For any historical event the causal nexus involves (a) probably various and differing human motives, the relative assessment of which may be impossible, plus (b) a series of physical causes which may be the challenges, to use Professor Toynbee's term, which evoke motives as responses. The historical chain is thus a complex one involving a twofold and interwoven series of physical events (mechanical causes) and human responses directed purposefully to human ends (which can be assessed only in terms of sufficient cause).

In neither case do we have an approach to the strict principle of necessity and universality characterizing "*final cause*". Kant, anxious to find an escape from the dilemma left by Hume —viz., that, if cause is simply the perception of invariable sequences, it would be impossible to distinguish the apprehension of a sequence of perception from a causal sequence of events in nature[7]—attempted a solution on the grounds that, as even Hume admitted, we do succeed in distinguishing between the sequence of events in nature and a succession of perceptions, and, as a condition of this sequence the necessity of causality must be assumed. "We have, then, to show, . . . that we never, even in experience, ascribe succession . . . to the object, and so distinguish it from subjective sequence in our apprehension, except when there is an underlying rule which *compels* us [ital. mine] to observe this order of perceptions rather than any other,

[7] "Holding, with Hume, that in all cases of perception what we are directly aware of is a succession of perceptions, he (Kant) contends that it is necessary to explain how in certain cases we succeed in passing from the knowledge of our successive perceptions to a knowledge of a succession in what we perceive. How is it that we know, when, as we say, we see a boat going down stream, that there is a succession in what we perceive and not merely a succession in our perceptions of it, as is the case when, as we say, we see the parts of a house? Hume, according to Kant, cannot answer this question . . ." H. A. Prichard, *Kant's Theory of Knowledge* (Oxford, 1909), p. 277.

nay, that this compulsion is really what first makes possible the representation of a succession in the object."[8] For Kant the "experience of an event is only possible on the (this) assumption" that "the subjective synthesis of apprehension" is only possible "by reference to a rule in accordance with which the appearances in their succession, that is, as they happen, *are determined by the preceding state.*"[9] This is held to be in contradiction to "what has hitherto been taught"[10] that "only through the perception and comparison of events repeatedly following in a uniform manner upon preceding appearances are we enabled to discover a rule"[11] of cause.

Now, though the mechanical sciences of the nineteenth century avowedly accepted the Humeian view of causality, in fact, both the sciences and vulgar usage held always implicit a sense of impulsion and necessity in the notion of cause that was closer to the theory of Kant than to that of Hume. When, in common parlance, we say with relief, during our long-drawn Canadian spring, that the heat of the sun will soon cause the snow to melt we intend to mean more than that we have repeatedly observed that the snow in March melts in the sunlight. We intend to mean that there is a necessity in the order of events, that the stronger sunlight radiates a greater heat on the surface of the earth and that snow *must* melt in heat. It is not simply that snow does melt in heat; from the nature of snow and heat it follows that snow will not continue to exist in the presence of heat. Similarly the astronomer, who, gazing into the heavens, observes an aberration in the procession of the celestial bodies, predicts the presence, density and path of some hitherto undiscovered planet not on the basis of repeated observations, but on the grounds that such a body *must* exist if the movements of other stars are to be rendered intelligible. Thus the assumption of intelligibility, the supreme but justifiable arrogance of the knowing human mind, is implicit in the scientific practice of applied causality. That there is an intelligible order in the

[8] Kant, op. cit. (ed. Smith), pp. 223-24.

[9] Ibid., p. 223.

[10] Meaning Hume.

[11] Ibid., p. 223.

external world, or at least an order that can be rendered intelligible, is a condition of all science. This implies a necessity in the order of events and it is this principle of necessity that Kant knows as cause.[12]

This necessity seems akin to, though distinct from, the necessity inherent in a logical system. Certainly mathematical systems do "fit" natural systems, though why they do so, we may not be able to say. Kant would imply, of course, that the systems of natural events, being understood through the categories of the understanding, can only reveal themselves in relations that conform to those categories; but the unfortunate dichotomy of the Kantian ontology makes the epistemological argument difficult to accept, because if the apprehended relations are the relations of representations only, how, then, are we to account for the alleged necessity and irreversibility of causal sequences? Presumably the correspondence of events in a causal sequence and the apprehension of them through the categories of the understanding requires the postulate of a Cartesian God or a Leibnizian Harmony.

However this may be, we have the pragmatic sanction of the physical sciences for operating *as if*[13] the necessary relations between events which we intend to mean when we speak of cause do correspond to the necessary relations of logical and mathematical systems.

Actually there has been a revolt in modern science against the causality concept in its mechanistic form, and even in the

---

[12] We avoid here the epistemological problems that Kant's theory of causality involves him in. Clearly for Kant the admission of an external order in things-in-themselves was inconsistent with his principle of the undetermined and hence unknowable nature of the noumenal world. "Kant, by his doctrine of the unknowability of the thing-in-itself, has really deprived himself of an object of apprehension—Representations (apprehensions, that is) have to be related successively as our representations or apprehensions—and, at the same time, successively or otherwise, as the case may be, as facts of the object apprehended, viz., a reality in nature. . . . A representation, however, cannot be related by a rule to another representation, for the rule meant relates to realities in nature, and, however much Kant may try to maintain the contrary, two representations, not being realities in nature, cannot be so related . . . . The result is he (Kant) is unable to justify the very distinction, the implications of which it is his aim to discover, and he is unable to do so for the very reason that would have rendered Hume unable to justify it." Prichard, op. cit., pp. 280-82.

[13] Cf. H. Vaihinger, *The Philosophy of As If*, V. F. Ogden (London, 1924).

modified Kantian form. Mach is reported to have said: "I hope that the science of the future will discard the idea of cause and effect, as being formally obscure; and in my feeling that these ideas contain a strong tincture of fetishism, I am certainly not alone."[14] Mach's hopes have been partially realized, though a very weighty body of modern scientific opinion still holds firmly to the causal concept. The basis for the attack on causality is the failure of microscopic particles to conform individually to causal laws. "We have lost cause where we lost substance, in the atom."[15] If, however, it is impossible to predict, on the basis of causal law, the exact behaviour of the microscopic—or ultra-microscopic—particle, it is still perfectly possible to predict, on the basis of probability theory, the behaviour of any group of particles, provided the number of individuals in the group is sufficiently large. It is no longer possible to assume, in the words of Professor Max Planck, formerly Director of the Kaiser Wilhelm Institute in Berlin, "that the course of a process can be represented by means of an analysis of it into its spatial and temporal aspects. . . . It is thus the concept of wholeness which must be introduced as well into the field of physics, as into that of biology, in order to enable us to understand and formulate the laws of nature."[16]

This modification of the causal concept by the introduction of the idea of probability is more difficult to understand logically than to use pragmatically. Judged in this latter sense it amounts to no more than this; that within the sphere of usual physical phenomena we can behave as though the causal laws operate without exception, and we shall find all our anticipations justified by the results. It is only when we deal within the atom with particles like the electron that we cannot apply ordinary causal concepts to individual instances with pragmatically justifiable results. Judged, logically, however, the application of the theory of probability seems to throw us straight back to Hume. Certain "wholes" or groups of phenomena, looked upon within a system as units, may appear to be

[14] Quoted by J. L. Stocks, *Time, Cause and Eternity* (London, 1938), pp. 42-43.
[15] Ibid., p. 44.
[16] Quoted by Stocks, op. cit., pp. 44-45.

connected in "causal" sequences, in the sense that, as far as we know, one "event" always follows upon another, and we can say, with Hume, it is exceedingly probable that it always will. But, in probability theory, there is no reason to believe that it must. Indeed, the essence of probability theory is the assertion that there is no *necessity* in individual behaviour, it is arbitrary or whimsical[17] and in that sense non-causal, and it is only if you have a sufficient number of individuals that you can assert that the group, *as a whole*, *does*, *in fact*, behave in a specific way; but from the very nature of your premises you are precluded from saying that it must do so.

We find, then, that in one direction in the physical sciences the concept of causality has been carried far from the mechanistic theory of the early nineteenth century towards a rigid determinism closely allied to that of mathematics, while, at the same time, in another direction, the introduction of probability has thrown the causal concept back to the uneasy scepticism of David Hume. Mechanical causation continues to be used in the fields of study where it yields useful results, but in the more abstract and mathematical branches of science, such as macrophysics, it has been replaced by the concept of *necessary reason* and in the micro-physics by the *calculus of probability*.

3. "Meaningfulness" as an Element in Social Causation

We return to the social sciences where we saw the causal concept to be further complicated by the existence of human *motive*. In the physical sciences when we say that A is the cause of B, either we intend to say that there is a high probability that B will follow A, or that there is an impulsion, in the sense that the idea of A is not tenable without the consequence B, as when we cannot, knowing the laws of planetary motion, entertain the idea of planet A deviating from its projected course unless an, as yet unknown, planet B, is supposed to exist. But we cannot intend to mean that A, as an active

[17] This is not to say that all behaviour we call "whimsical" is in any sense uncaused. If we knew enough about it, we might be able to assert cause. This question remains one of controversy in modern science.

agent, participating consciously in the event, brings about *B*, according to plan.[18] Yet, in all the social sciences—to a greater extent in political science than in sociology, and in sociology more than in economics or human geography—the nexus of motive is what we have to deal with along with the unconscious causal relations. They exist side by side, are interwoven, and, so to speak, set one another in motion. We have what we might describe as a socio-physical event which evokes a human response.[19] The response is purposeful and takes form through another socio-physical event, perhaps an institutional change. Neither the first socio-physical event nor the human motive can be regarded separately and independently as cause; it is the complex of the two which results in the final event.

Thus climate may lend to a larger expenditure of real effort in the provision of fuel and shelter in a northern than in a southern latitude, but the responses of human agents are purposeful and the institutions for the provision of fuel and shelter may differ widely among communities in the same general latitude and climate. Habitually we refer to these motive relations as causal, saying that Britain's policy of "divide to rule" is the cause of Indian disunity or that a desire for liquidity is the cause of high interest rates. Again, we speak of very mixed and complicated relations as being causal in nature, as when we say that an increase in bank credits is the cause of higher prices. We have, in these instances, three varieties of causal relationships—and there are many others,[20] all of which are commonly met with in the social sciences. In the first instance the cause, climate, was the passive causal agent of the

---

[18] Even if we believe the universe to represent "design" which is implicit of intention on the part of some Divine Being or First Cause, the design is of a different order from human purpose. If the Universe is of God's design, the design is implicit, actual and immediate in the order of the Universe, for we cannot think of God as being imperfect and striving to achieve values external to and transcending Himself. Human values only exist as arising from the imperfections of human nature as humanity strives to realize something better. Achievement is the improvement of a present state of being. It is within this context of striving that human purpose exists.

[19] It is at this point that the problem of freedom is involved. We return to this problem in Ch. II.

[20] See MacIver, op. cit.

physical sciences, entering into a relationship with human agents in a form to elicit some general type of response from them; but the actual institutional pattern of the effect would not be completely determined by the physical cause, but would be partly influenced—we shall use no stronger word—by human motives responding to the causal physical event.

In the second instance human motives, in policy form, played the causal role, eliciting a designed human and institutional "effect". Back of operative policy intention there may be a complex of motive, into which we need not now inquire, but some of which, such as the response to the limitation of physical resources in a small island like Great Britain, may have had physical causes behind them to which they were the responses.[21]

In the final instance, which is very typical of the sort of causal relation familiarly treated in institutional economics, we speak of two events linked causally as though they were physical events in the world of passive natural objects. The events are what we call institutional, and our language about them suggests that we regard institutional events to follow one another without the intervention of human agency or plan. What we are really doing here is to introduce the calculus of probability. In dealing with human aggregates, as when with the aggregates of physics, we discover certain normal responses in like situations, and these responses can be depended upon with a high degree of probability. It is that which makes the generalizations of the social sciences possible. In the present illustration, human action, purposeful in nature, does, of course, intrude as a part of the complex causal relation. Prices go up because, having more money to spend, people do spend it because they want to buy. Normally their behaviour pattern —or their motives—can be so depended upon to conform to a generalized norm that we can afford to neglect it and speak of more money causing higher prices. In time of war, however, or under other circumstances of unusual effects on human motives, issuance of an increased monetary supply may be accompanied by increased saving—what with Victory Loan

[21] Cf. the "challenge response" thesis of Professor A. Toynbee in his *Study of History* (London, 1934), Vol. I.

campaigns and so forth creating other than usual human motives—and then the effect does not follow. The causal relations have been broken, not because the causal concept does not apply, but because it applies only through the motive-response situation, and not passively as in the physical sciences. In Professor Sorokin's language a "causal system" of the "pure" sort, viz., that which obtains in the biological and physico-chemical worlds, is "characterized by the existence of a tangible one-sided or mutual dependence of its parts upon one another, of the whole upon its parts, and of the parts upon the whole,—provided that this dependence is due to the physico-chemical or biological properties of the whole and of its parts," while "socio-cultural systems are characterized by the dependence and interdependence of their parts upon one another and upon the whole—; but this independence is not entirely due (and is sometimes due not at all) to the physico-chemical and biological properties of the parts, but principally, and sometimes exclusively, to the fact that the parts are the vehicles of the same system of meanings."[22] It is the meaningfulness of events, institutions and policies to which people respond, and the causal sequences with which social sciences have to deal are motive sequences, with purposeful responses appearing as the *effects* of meaningful institutional or physical situations and in their turn serving as the causes of planned or intended institutional or physical effects.

The possibility of treating causal relations of this type scientifically must depend on our ability either to group individual responses so as to apply the techniques of probability analysis or to simplify the situations sufficiently to permit certain fairly simple empirical generalizations about motive responses, as "people tend to buy in the cheapest and sell in the dearest markets". Economic theory has come to follow the second of these courses, and, in most recent times, has developed a vastly complex system, resting on a few essentially simple, highly-general propositions about human motive-responses in economic situations. In the consequent system, which we know

[22] Pitirim A. Sorokin, *Sociocultural Causality, Space, Time* (Durham, N.C. 1943), pp. 12-13.

as the corpus of economic theory, the truly *motive-causal* events are all assumed under a group of propositions concerning economic behaviour. These propositions or *laws* are the simplifying assumptions which define the universe of economic analysis. From them and the law of non-contradiction the complex system of equilibrium theory is deduced. We believe that it should be possible to build up in a somewhat similar way simplified models of change, that is dynamic models, in which human responses could be inferred on a basis of probability. If that be allowed, we should be studying truly causal situations and though our conclusions could never be regarded as strictly determinate or as rigorously demonstrated, we could claim for them a high degree of probability within the conditions defined by the simplifying assumptions. The body of theory so developed would not, of course, describe the real world more than static theory describes it. It should, however, "fit" certain actual sequences and so afford history and policy with hypotheses which might prove fruitful. The relationship, however, of this kind of inquiry to the formation of policy requires some further examination. If causal relationships can be established in economic events on the basis of probability, how can policy be effective in guiding economic development? What freedom of action is there, and in what way can the study of economic change throw light on the problem of the policy makers? Though to answer these questions requires something of a digression, our argument has certainly involved them, and they are, I think, of sufficient importance to merit answer. So to these questions and to others that will be shown to emerge from them, I devote the next chapter.

CHAPTER II

CAUSALITY, FREEDOM AND THE SOCIAL FUNCTION OF ECONOMICS

1. CAUSALITY AND FREEDOM

The argument so far has been intended to show the lack of a study of causal relationships in economics and the need for an understanding of the nature of these relationships. We have argued that the cause-effect relation in a system of social events is not a simple one, but involves human response to the meaningfulness of the challenge of the physical and institutional situation. Because the responses of aggregates of human beings in like situations are likely to be similar it is possible to generalize and to speak as though a simple cause-and-effect relationship obtained, and to do so with a high degree of probability. Such simple situations depend, however, on certain assumptions about the system of values held by the society, that the ends of social action are understood and agreed upon and that clear-cut economic motives operate. We have now to ask what distinguishes the situation in which a conscious social decision involving basic value judgments is made.

Now the argument we are developing involves, it is clear, the problem of human freedom. Let us push our questions a little further. When social or economic changes occur, are they inevitable? If they are not, if an element of free election as between courses of action is admitted, in what sense may we speak of cause?

It is surely today as unnecessary to insist that objective events exercise a governing influence on the conduct of human affairs as it would have been for a Greek of the heroic age to urge that the Gods controlled Man's fate.[1] We are all determ-

[1] Though man's fate was generally regarded by the early Greeks as beyond his own control, the gods themselves disposed of man according to recognizable

inists to that extent. The significant question is the extent to which we can find "free decision" and assimilate that with our concepts of causality. Our position is that the physical-institutional world, "objective reality", determines the limits within which decisions are made, that human responses to that objective reality are limited within this "area of decision" but within the area free decisions are taken. These decisions obey causal laws, but they are not physical or mechanistic causal laws: rather they are laws derivative from the prevalent system of values or *mores*, and cause is consequently purposive. The rational and purposive judgment is surely what we must mean by free decision.

Let me illustrate this. If we consider the corruption of the freely competitive market we may observe that technological innovations requiring larger capital investment and at the same time leading to expansion of output for a period at decreasing unit costs must have a twofold effect. There is a necessity here, implicit in the full meaning of the process. Decreasing unit costs mean, for a period at least, unstable competition with

purposeful laws. Zeus wished the Acheans to pay in disaster for Agamemnon's rebuff to the renowned Achilles. The rational Greeks modified the primitive idea of a supernatural control of human destiny by making their gods human and rational. The great Ulysses was not above scheming to circumvent the gods, and Zeus himself says: "Mortals say that evils come from us, but they even of themselves, by their follies win sorrows beyond the claims of fate." (*Od.* I, 32-34). With growing civilization the Greeks moved far from the primitive concept of destiny. Though the Fates appear in the classic tragedies to thwart and confuse man, the tragedy very often lies, as it always does for Shakespeare, in man's own incapacity or interest to rise above circumstances. Aristophanes, the great rationalist, pokes fun in *The Frogs* at the idea of the *deus ex machina*, and Plato in the *Republic* writes, "This is the word of Lachesis: No divinity shall cast lots for you, but you shall choose your own fate: . . . The blame is his who chooses; God is blameless." *Republic*, Book X (Everyman ed. London, 1906), p. 344. Similarly the great Hebrew prophets took the primitive idea of God's displeasure and transmuted it to show that the sufferings of Israel were the consequences of no arbitrary destiny or capricious Jehovah, but of their own sins. "Thy way and thy doings," cries the prophet Jeremiah, "have procured these things unto thee. This is thy wickedness, because it is bitter, because it reacheth into thine heart" (Jeremiah IV, 18). Indeed, it is a single characteristic of the primitive and unsophisticated mind to believe in supernatural or supernal Fates or "inevitable forces" that fix the course of human events beyond the altering or control of the conscious mind. The modern babble about inevitability of history is simply the clothing of an ancient superstition in modern dress Causal laws, of course, operate in human affairs, but their operation, as we shall show, involves the active participation of the human agent. A highly civilized society has always known its power over its own fate.

each competitor chasing the next in a destructive race still further to increase output and lower costs.[2]

There is no incentive to operate to reduce output and hence raise the effective demand price. At the same time the new capital equipment requires physically and financially larger plant units.[3] If we rule out, for the purposes of this illustration, the possibility of an expansion of demand sufficient to take the new output at a price to cover costs, we see that this process leads inevitably to the growth of larger and fewer plant units. So far the process appears an entirely objective one, the causal relations very like those which obtain in the physical world. Actually human motives and decisions are involved at each step, but so definite is the process that little choice is left the human agent. It is true an entrepreneur might decide not to make use of the new technique, not to increase output, not to lower price; but if any entrepreneurs did so decide, their decisions would not affect the process. It would go on in spite of them and they would simply go bankrupt and shut down, thus unwillingly contributing to the same effect of concentrating the industry.

When we turn, however, to the question of deciding what should be done in the social interest about the resultant condition of monopolistic competition which, if the process continues, will ultimately obtain in the industry, we find a different situation. If the resultant situation is not recognized as being hostile to the public interest, nothing may be done. Let us consider this for a moment. It is a situation where human purpose fails to make itself declared in social action. The institutional process continues according to well-known laws. Some kind of price understanding will be achieved by the firms and a monopolist's surplus will be wrung from the consuming public. There will be excess capacity and social waste.[4] In

[2] Cf. Ch. XI.

[3] This is an illustration and is not intended as a complete explanation of the causes of industrial concentration and financial amalgamation. There are, of course, contributing causes such as the realization of special gains from stock manipulation, etc.

[4] Cf. Chamberlin, op.cit. For an estimate of excess capacity in monopolistic trades, see L. Reynolds, *The Control of Competition in Canada* (Cambridge, Mass., 1940).

such an event it would be accurate to view the entire temporal process as governed by physico-institutional causal law, and to speak of it as "inevitable" and to deny freedom to the human will. But the process is inevitable, the will is not free, because of a failure of knowledge or because of a value system that does not recognize this sort of situation as socially evil, as, for instance, in a society which takes more delight in the pride of possessing "industrial tycoons" of great wealth, power, and news interest than in the material advantages of more and cheaper goods. The lack of freedom is not imposed, in our illustration, by the technical process; its source is moral.

If, on the other hand, monopolistic competition is viewed as socially undesirable, if it is regarded as contrary to the socially accepted system of values, there are several things that can be done about it. It would be possible to do as the Socialists urge and take over the firms involved; it would be possible to bring them under some public regulatory commission; by an ingenious scheme of taxation they might be induced to operate at an output considered to be in the public interest.[5] There are certain other things, however, that probably cannot be done about the situation. The technical trend that has created monopolistic markets cannot be reversed. Although it might be possible by regulation to force the firms to simulate the conditions of a perfect market it would not be possible to force a reversal to a large number of small-scale competitive firms. Thus the area of free division is defined by the physico-institutional forces; the nature of the actual social judgment made depends on the active human agent and his scheme of values.

Many have argued that this scheme of values, itself an objective social fact, is institutional and determining. So it is, but not in the same way as the physical and institutional forces such as we have been discussing. At any time the scheme of values commonly accepted by a society will have an influence on men's judgment with respect to any policy problem. But judgment is active, conscious and rational.[6] The value scheme

[5] See B. H. Higgins, "Post War Tax Policy, Part I", *Canadian Journal of Economics and Political Science*, Vol. IX, No. 3, Aug. 1943, pp. 408-28. Cf. also C. W. Hazlett, *Incentive Taxation* (New York, 1936).

itself is quasi-rational. It enters into the thinking of the deciding agents, but there is always some revaluation made in the light of present circumstances. The value scheme is never wholly accepted, is always subject to criticism and re-assessment. This is the very kernel of all liberal thinking, that by taking thought man can progress,[7] not necessarily in a material sense, but in the sense of working towards improved and more rational value judgments. "For man . . . walks up the stairs of his concepts".[8] Thus it is that implicit in the liberal faith are the two propositions that values are rational and that free discussion

---

[6] This is basic to liberal philosophy. Professor Hobhouse, a great modern defender of the liberal tradition says, "it (any body of social knowledge) is social knowledge or social thought, not in the sense that it exists in the mind of a mystical social unit, nor in the sense that it is the common property of all men, but in the sense that it is the product of many minds working in conscious or unconscious cooperation, that it forms a part of the permanent social tradition going constantly to shape the thought and direct the efforts of fresh generations of learners . . . " L. T. Hobhouse, *Social Evolution and Political Theory* (New York, 1911), p. 95. Later Professor Hobhouse argues, "Any developed unity in the social mind rests on a consciousness, first of some special relation of each individual constituting it to his fellow members, and secondly of the group, society, institution itself as a whole" (p. 99).

[7] "Progress has consisted in the realization of the conditions of full social cooperation, and *in the extension of the rational control of life*", [italics mine] L. T. Hobhouse, op.cit., p. 156. Graham Wallas wrote, "The function of social judgment is the guidance of human action", Graham Wallas, *Social Judgment* (London, 1934), p. 30.

The liberal position derives ultimately from Locke: "Prejudice is not insurmountable, moral knowledge is just as possible as physical knowledge", as one contemporary writer summarizes Locke's position, "(and) Locke was convinced that his sensationalist psychology demanded an individualist ethics", W. Stark, *The Ideal Foundations of Economic Thought* (London, 1944), pp. 4, 6. Locke developed a rationalistic utilitarianism that implied the control of judgment over destiny: "If to break loose from the conduct of reason, and to want that restraint of examination and judgment which keeps us from choosing or doing the worse, be liberty, true liberty, madmen and fools are the only freemen . . . This is the hinge on which turns the liberty of intellectual beings in their constant endeavour after . . . true felicity, that they can suspend this prosecution in particular cases, till they have looked before them, and informed themselves whether that particular thing which is then proposed or desired be in the way to their main end, and makes a real part of that which is their greatest good", John Locke, *An Essay Concerning the Human Understanding* (Pringle-Pattison ed., Oxford, 1924), Book II, Chap. XXI, paras. 50-52, p. 147. As Professor Gibson says, *Locke's Theory of Knowledge and Its Historical Relations* (Cambridge, 1917), p. 106, Locke's acquaintance with Newton and Boyle led him to understand the nature of causality in science, and he carried over the causal concept to the study of man, but he developed the fundamental liberal theorem that man's judgment and will enter actively into the causal process.

[8] John Steinbeck, *The Grapes of Wrath* (New York, 1939), p. 179.

in the long run leads through error to a clearer vision of truth. The entire liberal credo is based not on an indeterministic whimsical theory of human behaviour, but on a concept of cause which applied to policy problems involves the careful analysis of the forces defining the area of decision and the reassessment of accepted value systems in the light of the special circumstances of that "area" and in the light of rational criticism based on previous experience.

2. The Social Function of Economics

When we turn from the implications of this argument for political theory to those affecting the method of economic inquiry, we must see at the outset that the primary function of political economy must be to define these areas of decision. The old political economy did exactly this. The quasi-historical laws of capital accumulation (Smith), of diminishing returns and rent (Ricardo), of population growth and production (Malthus, Ricardo), of comparative advantage and foreign commerce (Ricardo, Mill) and of tax incidence and shifting (Smith, Ricardo, Mill), and the scientific socialism of Karl Marx and Freidrich Engels, were all intended to this purpose. They informed or pretended to inform the voter and the legislator of the physical and institutional laws that would determine the sequence of events and that fixed the possibilities of social judgment. Malthus, for example, taught that the condition of the poor was inevitable and showed the Speenhamland system to be a mere toying with "fate", but indicated an area of decision, positive action to check population growth. Smith laid down the conditions of economic progress and showed a whole range of possible policy to achieve this progress. Ricardo's disillusioned pessimism made him more "scientific" in the sense in which modern economists use the term, but he was still concerned to justify his pessimism by showing how the physical forces of population growth and diminishing returns rendered any considerable economic progress unlikely and policies directed to the improvement of wages nugatory. He rules out that kind of reform as without the area of rational decision and

directs the legislators rather to the possibility of a gradual improvement in the lot of the masses by a free commercial policy based on the law of comparative advantage. His whole rent theory is a moral apology for disregarding the case of the landlord for protection. The present writer believes that the proper interpretation of Marx's argument is that the processes leading to the disintegration of the liberal competitive capitalist system are inevitable, but that there is no inevitability in the coming of socialism. On the contrary, Marx's whole appeal is to the workers to unite to create the better socialist society; if they fail to unite, competitive capitalism may give place to a brutally authoritarian and exploitative system. On this interpretation, the Marxian faith in inevitability is tempered by the belief that there is a conscious choice the workers may make as to what is to succeed bourgeois capitalism.[9]

In our own times the work of Lord Keynes and of the writers who follow him seems closer to the old political economy in the respect of this function of defining the area of free decision than is the work of those more traditionally minded economists who have been too impressed by the arid methodology imposed on economics by Jevons and the Austrians.[10] Professor Hansen, for example, sets out clearly the possible choices open to a society which wishes to maintain a high level of employment.[11]

The definition of the areas of decision, however, demands of economics that it be able to generalize about human motives under certain like circumstances as well as about purely physical or what we might call "physico-institutional" laws. This is the second implication of our previous argument. It is impossible to develop any theory of causal change—for example, that which brings about industrial concentration—without making certain generalizations about human behaviour, in this case that

---

[9] Cf. Sidney Hook, *Towards the Understanding of Karl Marx* (London, 1933), especially Chapters XII and XIII.

[10] Professor Frank H. Knight has surely failed to relate method to function in his contention that it was Jevons and the Austrians who put economics on the right and proper path. See his "Ricardian Theory of Production and Distribution", Part I, *Canadian Journal of Economics and Political Science*, Vol. L, No. 1 (Feb. 1935).

[11] Alvin H. Hansen, *Fiscal Policy and Business Cycles* (New York, 1941).

entrepreneurs will act to maximize net gains. Moreover, it would make nonsense of any theory of economic change if one were unable to assume a certain consequence in economic behaviour. To say that there is an area of free rational decision for the determination of policy, to admit that value judgments are slowly modified, and to claim, consequently, that there is sense in the liberal belief in discussion and reason as the basis for political judgment, is emphatically not to say that there is no regularity or generality about human responses in like situations. In a political situation complex value judgments must be made, situations do not repeat themselves, men learn from experience; it may consequently not be possible to generalize about the kind of political judgment that will be made in an area of decision. In an economic situation, which is comparatively simple, constantly repeated in the essentials of economizing scarce means to satisfy as highly as possible an unlimited hierarchical system of wants involving the most elementary human impulses, it is perfectly possible to generalize on the basis of probability theory about how human agents will behave. It is necessary to do this if any analysis of the complex causal chain of economic phenomena is to be possible.

Since Jevons and the Austrians, economic theory has consisted of the elegant elaboration of market models based on a few axiomatic generalizations about human motives and human behaviour in highly simplified economic situations. The great strength of modern economic analysis is the rigour and high degree of generality obtained in these models. The weakness is that the restricted nature of the model confines it to the explanation of behaviour in a "given", i.e., static, context, and does not enable the inquirer to explain how changes in these situations may occur. Thus this kind of analysis provides one with a part of what is needed to explain economic development—i.e., generalizations about human motives, in given economic situations—but it cannot provide the other necessary analytical tool, the concept of the cause of change in a temporal order. In order properly to define areas of decision and fulfil the function of informing policy it is necessary, as the classical political economists saw, to explain change in a temporal order

and the consequent changes in the institutional situations in which economic motives operate.

When we talk about economic change we are in danger of speaking vaguely; we may be referring to changes in the operation of the fundamental causes of economic movement, as when population structure, spending habits, incentives and motives or techniques are changed; we may refer to changes in the economic or even the social institutions, by means of which economic systems function, as when we discuss changes in banking law and practice, or, more significantly the legal institutions enshrining property rights;[12] or we may be referring simply to a process of adjustment within a model, the parameters of which are regarded as fixed, as when we study the process

[12] In English law, for example, the continental doctrine of abuse of rights is scarcely recognized as relating to property rights. "No use of property, which would be legal if due to a proper motive, can become illegal because it is prompted by a motive which is improper or even malicious" (Bradford v. Pickles, A.C. 1895, 587). In this leading English case Counsel for the appellants cited the rule of Marcellus: "Si non animo vicini nocendi, sed suum agrum meliorem faciendi", but Lord Watson in giving judgment against the appellants held that "the noble and learned Lords who took part in the decision of Chasemore v. Richards held that the doctrine (of Marcellus) had no place in the law of England". "If the Act," held Lord Macnaughton, "apart from motive, gives rise to damage without legal injury, the motive however reprehensible it may be, will not supply that element." The case, we might add, had to do with the right of one Pickles to divert streams flowing underground across his land, and so to interfere with the water supply of the City of Bradford, in order to force the city to buy him out.

In a similar French case (L'arrêt Clement-Bayard, Cassation, 2 Aôut, 1915) a certain Coquerel was held to have abused the right of use of property when he tried to force M. Clement-Bayard, dirigible manufacturer, to buy his property by erecting stakes on it so as to endanger the dirigibles in mooring. The French Court held the abuse to be established and held that an act, otherwise legal and proper as attached to the right of property, might become illegal if the means employed were illegitimate and inspired only by malicious and mischievous intent.

The effect of a change in English attitudes towards property when reflected in the law would have profound economic consequences. The use of property in a manner detrimental to the public interest would be held abuse. This might well change the whole status and practice of trade associations, combines, cartels, etc. In turn the economic theory of pricing under monopolistic competition would have to be modified. Statutory limitation of rights is not without precedent: many will spring to the legal mind. Perhaps most relevant to the question of property rights is the restriction on the proprietary right of patent imposed by the Patent and Designs Act (6, Edward VII, Ch. 29) which provided for the loss of patent for non-use. In section 27, 2d, the Act makes use of the concept of abuse of rights in the following phrase: "the monopolistic rights under a patent shall be deemed *to have been abused* in any of the following circumstances . . . " It then recites the circumstances of non-use which constitute abuse.

by which the firm moves to the equilibrium output.[13] These categories are illustrative, not rigid and mutually exclusive. Most studies of economic "development" commence with changes in selected basic causes, such as Malthus's preoccupation with population growth or Professor Schumpeter's concentration on innovations as the prime cause of "development",[14] and then trace out the full consequences of these causal changes in the system of economic institutions and perhaps even in the analytical models as well. The intention of these theories is to find a unifying cause to explain the long-run temporal process known as development. Most "process" studies on the other hand are concerned not with basic causal changes in the temporal order but with adjustments to minute changes within an extra-temporal model. Business cycle analysis is concerned again with a comparatively short-run change, usually conceived as occurring in rather more complicated models, and the changes, as we have seen, are usually of the order of adjustments, observed by the method of comparative statics, to temporal fluctuations in the rates of spending or saving and investment. Some theorists, of course, look for an explanation of the cycle to more fundamental phenomena as when Jevons suggested that sun spot cycles lead to cyclical variations in trade or when, like Spiethoff, Schumpeter and others, they find the cause of the cycle to lie in the periodic introduction and adaptation of new productive techniques.

We are not ourselves gravely concerned to attach a precise and definite connotation to the phrase economic change. If there were a particular process that, teleologically conceived, could be regarded as development, we might, indeed, be concerned to identify this process and to try to explain it in terms of a single cause. It is our contention, however, that economic change is a complex phenomenon, that it has many causes, often inter-related, and that it must be studied in terms of this

[13] See Ch. X.

[14] J. A. Schumpeter, *The Theory of Economic Development*, tr.Opie (Cambridge, Mass., 1934). Professor Schumpeter, it is scarcely necessary to add, holds in the forefront of his readers' attention that the "social process is really one indivisible whole", op.cit., p. 3.

complex causal structure.[15] We shall show that the selection of a single "given" cause of change introduces a bias and that the assumptions with respect to causes regarded as given have determined the conclusions, indeed the entire systems of economists who have treated economic change. No important cause of change, if I properly interpret the implications of the previous argument, can be regarded as exogenous, that is as external to the system of economic events. It is acted upon by those events and by other major causes. We may make a distinction between external and internal causes, but the former must be limited to that category of events which English common law ascribes to "acts of God". They are single, individual events which impinge on the economic order, but which, being non-repetitive in any intelligible pattern and beyond the influence of social action or organization in any way, are not susceptible of generalization. Major causes of economic change, like scientific progress, population growth or geographic discovery, though they may involve non-economic forces, are, nevertheless, acted upon by the course of economic events and all such "causes" must be regarded as within the order of economic causation. But all this can be studied more effectively with representative theories of change before us as illustrations of the methodological concepts which are important to us because of our preoccupation with processes that are essentially temporal in nature.

[15] Purposive changes by means of policy decisions on the part of the society are, of course, included in our system. We regard them as *ad hoc*, however, and as emerging, as we have argued, within the framework set by the operation of other causes. We cannot conceive of any *long-run* developments, planned as such in advance either by man himself, or working themselves out purely according to some blind natural law. For us the process is always one of mutuality between the physico-institutional causes and the purposes of human agents.

CHAPTER III

EQUILIBRIUM, PROCESS, TIME

1. EQUILIBRIUM AND PROCESS

We have already said that the idea of "equilibrium" is what is called a methodological concept, a way of looking at things, and, in this case, a way of looking at certain forces which tend towards stability in the economic system, and of ignoring forces which create change. The idea of "process" in economics, and of such ideas as "stationary flow" and "flow equilibrium" are all extensions of the equilibrium concept to take account, by what we have called the method of comparative statics, of certain types of changes creating particular instabilities. As such these concepts help us to understand some kinds of economic motion, the adjustments of an "equilibrium" system to a given shift in a parameter, for instance, and to assess thereby some causes of fluctuation. While this method has value in business cycle analysis, I seriously question that "process analysis" tells us much about the adjustment of the firm. In any case these equilibrium concepts lack, as we shall show, much significance for the theory of long-run change. They focus attention on stability and the short-run adjustments of essentially stable models to particular shifts in parameters, and ignore the causes of change, the structural alteration of the economy, the long-run dynamic motion and the welfare effects. They imply, also, a particular and peculiar concept of time, which necessarily distorts the real nature of the temporal process of the economy.

There have been several distinguishable uses of the word "equilibrium" in economic literature. In the Walrasian sense it has been used with reference to the economic system as a whole; in the Marshallian with reference to an individual firm, and, sometimes, to a single industry.[1] In either case it

[1] Under perfect competition equilibrium of all firms requires as a condition equilibrium of the industry.

has sometimes designated a condition which is "stable", sometimes one that is "determinate", though usually, not always,[2] these are taken to be identical.[3] Again there is the important difference, noted by Professor Robbins,[4] between the concept of equilibrium in the stationary state, the equilibrium of the classical and Marshallian schools, and the "static" equilibrium of Professor J. B. Clark. Finally, as Professor Pigou shows[5] there is a difference between the firm's equilibrium position of maximized profits and the position of optimum or least-unit cost, a distinction which appears only when competition is non-perfect and which, therefore, was neglected by most writers prior to the publications of Mrs. Robinson[6] and Professor Chamberlin.[7]

For equilibrium of the system to exist in the Walrasian sense, "every household and every firm in the domain is, taken by itself, in equilibrium . . . For the firm this means that, under existing circumstances . . . no firm feels able to increase its revenue by transferring any element of its monetary resources

---

[2] In the case of "multiple equilibrium", either under perfect or non-perfect competitive conditions, there may be more than one "determinate" point of equilibrium between the cost and revenue curves. But of these only one is apt to be stable. Compare Joan Robinson, *The Economics of Imperfect Competition* (London, 1933), pp. 57-59, and A. Marshall, *Principles of Economics* (ed. 2, London, 1891), pp. 491-92. Marshall suggests that more than one position may be stable, though others will be unstable. Somewhat similarly for Walras, a series of determinate general equilibria might exist, not all of which, however, need be stable. We shall not treat this case in the course of this discussion.

[3] Thus under duopoly, or under monopolistic competition with decreasing costs, where prices and quantities are indeterminate, the market is said to be unstable. In Mr. Harrod's language, if two firms are in imperfect competition and one of them operates under decreasing costs, restrictions of output on the part of the firm with increasing costs will require even greater restrictions on the part of the other, and they may thus "chase each other indefinitely" (R. F. Harrod, "The Equilibrium of Duopoly," *Economic Journal*, Vol. XLIV, June, 1934, pp. 335-37). Similarly two decreasing cost firms may chase each other into bankruptcy with a series of increases of output.

[4] L. Robbins, "On a Certain Ambiguity in the Conception of Stationary Equilibrium", *Economic Journal*, Vol. XL, June, 1930, pp. 194-214.

[5] A. C. Pigou, *Employment and Equilibrium* (London, 1941), Ch. IV.

[6] Joan Robinson, *The Economics of Imperfect Competition* (London, 1936).

[7] E. H. Chamberlin, *The Theory of Monopolistic Competition* (Cambridge, Mass., 1933).

(capital) from the factor it is actually spent on to any other factor."[8] But, if full equilibrium requires that each firm be in equilibrium, it also requires that they stand in certain relations to one another, to the factors they employ, and to the consuming public as a whole,[9] so that it is not necessarily true that, if all the firms are in equilibrium, in the Marshallian sense, this totality will constitute an equilibrium of the system. This point is important because we mean to argue that there is one condition of equilibrium for firms which is incompatible with a condition of general equilibrium of the system.

Generally speaking, it has been assumed that the determinate output and price mark the equilibrium position of the firm and, because, on the usual level of abstraction, this is a stable position, this position has been identified with stability.[10] We must probe, however, the possibility that stability might occur at alternative positions. It has been argued that under conditions of both perfect and non-perfect competition, stability, in one sense, may be at points other than determinate ones, and that certain points of determinacy may be unstable. The stability that may come at points other than the determinate ones is, however, the result of frictions and is thus not an equilibrium position in any rigorous theoretical sense of the term. The incentive to change is present but the possibility of change is blocked, whereas, in a true equilibrium as long as consumers' preferences, the supply of the factors and techniques remain

---

[8] J. A. Schumpeter, *Business Cycles* (2 vols., New York, 1939), Vol. I, pp. 41-42.

[9] "What we want to learn before anything else is whether or not the relations known to subsist between the elements of the system are, together with the data, sufficient to determine these elements, prices and quantities, uniquely . . . The values of prices and quantities which are the only ones, the data being what they are in each case, to satisfy these relations, we call equilibrium values. The state of the system which obtains if all prices and quantities take their equilibrium values we call the state of equilibrium" (Schumpeter, *Business Cycles*, Vol. I, p. 41).

[10] Thus Edgeworth says, "Economic equilibrium may be regarded as *determined* [italics mine] by the condition that the advantage of all parties concerned, the integrated utility of the whole economic system, should be a maximum" (F. Y. Edgeworth, *Papers Relating to Political Economy*, 3 vols., London, 1925, Vol. II, p. 295). Edgeworth clearly believes that the determinate position maximizes advantages so that no party concerned would have an incentive to modify his behaviour, on the given data, and that it would be, therefore, stable.

unchanged enterprisers' behaviour is stable because, things being as they are, there is no incentive to change.[11]

But as Mr. Kaldor has shown,[12] there are elements of indeterminateness even in the classical case of equilibrium under perfect competition. These arise, however, because the classical notion of a long-run equilibrium under stationary conditions is the one Mr. Kaldor selects for analysis. This interpretation of stationary equilibrium is that which implies that, given the state of the arts, population, and capital supply in a stationary state, and given perfect competition in the factors and consumers' goods markets, the firm's equilibrium will be reached at the point where average unit costs, including rents, are just equal to price, a no-loss, no-profit point. Mr. Kaldor shows that, because the short-run purchases constantly alter the supply, and at the same time partially satisfy demand, the series of adjustments which are supposed to lead to long-run equilibrium do, in fact, continually alter the conditions of that equilibrium, so that its postulation in the long run is incompatible with conditions of perfect competition. This is true of certain agricultural commodities, and is the explanation of such phenomena as the hog and poultry cycles. In the general case, however, Marshall would probably reply that enterprisers, from experience, soon learn to anticipate market behaviour and to behave themselves as though the full supply and all purchasers were brought simultaneously together. Indeed, in his illustration of the corn market, he may have anticipated this objection.[13]

However this may be, there are certain other difficulties with the Marshallian concept of long-run equilibrium. It is

---

[11] Compare Professor Schumpeter's footnote on the distinction between a stability brought on by frictions and that which satisfies the "economic rationale" (Schumpeter, *Business Cycles*, Vol. I, p. 42 n.).

[12] Nicholas Kaldor, "The Equilibrium of the Firm", *Economic Journal*, Vol. XLIV, 1939; March, 1934, p. 60.

[13] "The price (of 36s.) has thus a claim to be called a true equilibrium price: because if it were fixed on at the beginning, and adhered to throughout, it would exactly equate demand and supply; and *because every dealer who has a perfect knowledge of the circumstances of the market expects that price to be established*" [italics mine]. (Alfred Marshall, *Principles of Economics*, ed. 2, London, p. 392).

a position toward which prices and quantities are supposed to move over a long-run period, there being no "dynamic" changes over the period. Now, in fact, the relations of the market which we wish to observe free of frictions, technical changes, and the other phenomena which are the inevitable processes of a time period, can be equally observed as functional relationships which are "out of time"; adjustments may be regarded as taking place immediately. This is the static abstraction, one which does violence to reality, no doubt, but no more violence than the notion of a stationary state which is imagined to exist in time, but in which none of the changes usually associated with the passage of time is allowed to have occurred. It has been contended (by the late Professor Henry Schultz, among others) that, when we proceed to the examination of these temporal changes, what we are doing is more clearly seen if the fundamental functional relationships of revenues and quantities, and costs and quantities, are treated as immediately related to one another, out of time. Thus a demand schedule is a series of potential prices for potential quantities at a moment of time. On this level of abstraction Mr. Kaldor's objections do not apply, and the equilibrium position of the firm is clearly determinate at the point of equivalence between marginal revenue and marginal cost: but this static equilibrium is, as Professor Robbins says,[14] a conceptual device for analysis, not a position toward which in the long run actual prices and quantities are supposed to approach. In the language we shall use, it is a point of "potential equilibrium",[15] a point, admittedly, from which entrepreneurs would have no

---

[14] Robbins, *The Conception of Stationary Equilibrium.*

[15] " . . . the various outputs for which marginal cost and marginal revenue points are shown by the marginal curves must all be considered potential outputs at the selected moment of time. Actually the firm does not increase its output through those various points, and is not considered as doing so. They are potential in the moment" (B. S. Keirstead, "Technical Advance and Economic Equilibria", *Canadian Journal of Economics and Political Science*, Vol. IX, Feb. 1943, p. 57). This use of "potential" in the logical sense of imminent must be distinguished from the sense of the word as used by J. B. Clark, for example, when he speaks of "potential" competition which producers of substitutable commodities offer a monopoly, a usage which suggests the recognition of a probability.

incentive to move, and stable in that sense, but not a point to which, over time, prices and commodities would necessarily approximate, because over time we should have to allow for temporal change. We return to the case where the market is imperfect. Here we find two points of determinacy,[16] one the point of maximum profit and the other the point of minimum cost, and these points do not, as under perfect competition they do, coincide. The first point, which on the basis of minimum theory could be used under perfect competition to define the firm's equilibrium, fails to satisfy the condition, when competition is not perfect, of no incentive to change. The second point, that of maximum profit, marks the point where, for a single firm, there is stability in the sense of there being no incentive to change, and, on the static assumption, would appear to be the point of potential equilibrium.[17] We should observe, however, that, if this point is selected as that of the firm's potential equilibrium, the condition of firms' equilibria is now incompatible with equilibrium of the system as a whole. This is because in the non-perfect market enterprisers are able, by the very fact of restricting production to less than the optimum cost point, to realize abnormal and unequal profits. As long as inequality of profits exists there will be an incentive for enterprisers in the less profitable industries to move to the more profitable ones. This can be, and is, resisted by closure of entry, which is maintained by all sorts of

---

[16] Professor Pigou expresses the same distinction in somewhat different language: "There are still two senses of equilibrium between demand and supply that need to be distinguished . . . the first is equality between the quantity, which, at the ruling price, demanders would like to buy and the quantity which, at that price, suppliers would like to sell; the second is the equality between the quantity which, at the ruling price, demanders would like to buy, and the quantity which would make price equal to the marginal cost of production. Under perfect competition these two senses of equilibrium come to the same thing" (Pigou, *Employment and Equilibrium*, London, p. 30). I find Professor Pigou's language here a little troublesome, but he is clearly referring to the intersections of the price curve with the average and marginal cost curves. The point is the same under perfect competition for all firms only if rent is included, and it is then the least cost point.

[17] We exclude the case of duopoly and such other cases of oligopoly as are indeterminate. These only go to increase the element of indeterminacy and the imperfection of equilibrium in the system as a whole, which we intend to show arises from non-perfect markets.

devious devices, a sort of jungle law of threats, buyings-off, and secret and semi-secret understandings. Closed entry may be maintained and the position may then appear to be "stable", but we must see that it is "stable" in a sense exactly analagous to the friction-held "stability" under perfect competition to which we earlier referred, and, on the level of the static assumption, this is not equilibrium at all. It is in no way a determinate position, its stability does not conform, in Professor Schumpeter's phrase, to an "economic rationale", it is held by no exact balancing of economic forces as in the Walrasian equilibrium of the system under perfect competition, but is maintained by force, by the exercise of power in an institutional structure in which the power relationships are essentially unstable. Such a condition then, even if stable over a period, is not one of equilibrium, and we conclude that the equilibrium position of the firms under non-perfect competition is incompatible with any equilibrium, in the strict Walrasian sense, of the system. The inherent instability of duopolist and certain oligopolist markets goes to increase the instability of the system in an economy of non-perfect competitive markets. We shall refer to this condition of the system, when the firms are in potential equilibrium and the system as a whole is consequently subject to disequilibrating forces held only temporarily in check by the balance of power, as a condition of "imperfect equilibrium".[18] We note also an implication of welfare significance. Walrasian equilibrium, under perfect competition, was a social optimum. It was one ideal solution of the economic problem of allocating resources so as to maximize satisfactions at a minimum total of real costs. In imperfect equilibrium no such social optimum appears. It is necessarily a position of excess capacity. There is thus, if one may introduce a non-economic consideration, a social motive to change under imperfect equilibrium which is not present under perfect general equilibrium with perfect competition.

[18] This should be distinguished from "imperfect equilibrium" as defined by Professor Schumpeter, for whom imperfect equilibrium is the condition which "without satisfying ligamina exactly, is as near to perfect equilibrium as it will go, and (which) will not move from that position unless some event impinges upon it" (Schumpeter, *Business Cycles*, Vol. I, p. 44).

In short, I conclude that the concept of equilibrium is of value in the analysis of the firm's behaviour, but, if it is used in the static sense, it fails to explain in any way how firms come to the potential equilibrium point; if used in the stationary sense, it does explain how equilibrium is approached but only by ignoring much that is essential in the time process. Our whole theory is distorted when the concept is extended to the system as a whole. The emphasis is on stability and the apparent inconsistencies between general and partial equilibria are ignored, as are all the elements of "indeterminacy", and it is these very elements that are ignored which we hope to show are significant both directly to welfare and indirectly, through the changes in the structure of the market, for the balance of power in society. This is enough to justify the rejection of the concept of equilibrium, in any static or stationary sense, for the study of economic change.

Could we, however, make use of some concept of "dynamic equilibrium", in spite of the dangerous conjuncture of terms involved? "It may be doubted if such a conception of dynamic equilibrium is of great importance in interpreting economic change, for there is no 'natural' tendency of society to conform to it."[19] There have been, however, concepts of equilibrium over time and these have been used to discover adjustments to shifts in a single governor of the system. There is, for example, the notion of long-run equilibrium in Marshall. The Marshallian concept is that of an equilibrium position toward which, over a sufficient period, prices and quantities are believed to approach and at which, no changes in techniques or other changes of a temporal sort having occurred, they come to rest. This, of course, is simply an extension of the stationary equilibrium concept, and it has the weakness, observable in Marshall's

[19] K. E. Boulding, *Economic Analysis* (New York, 1941) p. 769. I might add, for what it is worth, that I nevertheless experimented with abstract models of change or flow in which certain defined changes were allowed to occur and, a continuous adjustment being assumed, a "dynamic equilibrium" was postulated. Other investigators may have better results and I do not want to discourage them, but I feel it may be helpful to record that I found the models I thus created sterile and formal, and even indeterminate except under assumptions that both violated all common sense and disguised the real nature of the temporal process.

treatment of diminishing returns, of tending to confuse an historical process with purely functional relationships of quantities.[20] Since it is these very temporal changes, excluded from operation in Marshall's stationary state, which we wish to study, it is clear that we should receive little benefit from the Marshallian concept.

Similarly we must distinguish Professor Pigou's "flow equilibrium"[21] and Professor Schumpeter's "stationary flow"[22] as non-dynamic concepts. In each case the concept is one of a static equilibrium projected over a period of time and held constant by reason of the assumed immutability of population, capital supply, state of the arts, etc. Thus, for Professor Pigou, flow equilibrium entails a constant rate of purchase and sale, of hiring and letting. "For the economic system as a whole to be in flow equilibrium obviously means that all rates of demand and corresponding rates of supply embodied in it are in this type of equilibrium. Plainly this condition cannot be satisfied except in the classical stationary state. The existence of such a state implies, of course, unchanging tastes and technique. It implies, too, a definite relation in every industry between selling price and marginal prime cost . . . Strict flow equilibrium implies, moreover, a stationary population and a fixed stock of capital equipment."[23]

For Professor Schumpeter stationary flow is "an unchanging process which flows on at constant rates in time and merely reproduces itself."[24] Thus in both cases the concepts are analytical devices postulating the continuation of a stationary or static equilibrium over a period of time from which ordinary temporal changes have been abstracted, and designed for the study of the disequilibrating forces that may upset that equilib-

---

[20] Thus Marshall in one place states the law of diminishing returns: "The application of increased capital and labour to land will add a less than proportionate amount to the produce raised" (Marshall, *Principles of Economics*, ed. 8, London, 1920), p. 153.

[21] As defined in Pigou, *Employment and Equilibrium*, pp. 32-33.

[22] As defined in Schumpeter, *Business Cycles*, Vol. I, pp. 35-42.

[23] Pigou, *Employment and Equilibrium*, pp. 32-33.

[24] Schumpeter, *Business Cycles*, Vol. I, p. 35.

rium. Both the Pigou and Schumpeter concepts are part of an accepted analytical method. They say, in effect, "these are the conditions of a sustained equilibrium over time. Let us set them out and see then how and in what directions this equilibrium is upset as we admit the operation, one by one, of the disequilibrating forces." This is to say that the method of comparative statics is to be used. There is nothing wrong with this for the purposes Professor Schumpeter has in mind, but, as we have already argued, it precludes the possibility of a general study of an inter-related causal system operating in time. We are interested in a process of change, in which effects in turn become causes of still further effects and the sum of economic effects is the cause of alterations, in turn, in what was taken as the original cause. The process and the causal links are continuous, and the cycle of causation is regarded as a closed cycle.

This idea of a process differs fundamentally from what the word has come to signify in economic theory. Ordinarily in economic literature, the attention given to the idea of "process" differs from the notion we have been elucidating. The analysis of process generally means the theory of adjustment from one potential equilibrium to another, the only equilibria envisaged are the static equilibria positions; something is regarded as disturbing one of the parameters defining the system; the system then adjusts itself to the new conditions approximating another point of static equilibrium. Time is thus conceived as constituted by moments of disturbance alternated with moments of equilibration, and process is the sequence of approximations to potential equilibrium created by firms moving along their marginal cost and marginal revenue curves until they "hit upon" what for the time appears as their highest profit.[25]

---

[25] Thus Joan Robinson: "It is true enough that no monopolist will hit upon the exact point at which his net revenue will be greatest unless he has an enlightened and accurate system of cost accounting and a good knowledge of the market conditions in which he has to sell. But *if the conditions of demand and supply remain constant over a fairly long period* [my italics], the monopolist will be able to hit upon the exact monopoly output merely by balancing marginal receipts against marginal cost. We need not imagine that he is able to plot the demand and cost curves throughout their length, but merely that he can see whether selling a little more of his product than he does at present will increase or decrease his net gains" (*The Economics of Imperfect Competition* (London, 1936) pp. 56-57).

Dr. Karl Frisch's "expansion path" concept, as cited and developed by Dr. Carlson[26] appears similarly to conceive of adjustments to be in the form of experimental increments of "inputs", adjusting the combination of factors, according to the law of varying proportions, to suit the economy of increased outputs.

We are going to examine this whole theory of the process of the firm in some detail in Part Four. In the present context I wish only to suggest that it is a formal presentation of the deductions that are possible from the profit-maximizing assumptions,[27] and that it has nothing to do with "process" in a temporal sense at all. It deals with "instantaneousness" (in Professor Whitehead's sense), that is, it "deals with logical entities . . . at an instant, where an instant is conceived as deprived of all temporal extension." As a result "process analysis" does not tell us how a firm moves in time to approach its equilibrium position, it simply further elucidates the properties of that position as they are implicit in its definition. "Equilibrium analysis is reduced to a purely formal level, since it gives no explanation of the equilibrium."[28] All process analysis, Dr. Dingwall concludes, "associated with the traditional equilib-

---

The distinction between this concept and the one we are trying to achieve is similar to the distinction Professor Whitehead draws between "simultaneity" and "instantaneousness" (A. N. Whitehead, *The Concept of Nature*, Cambridge, 1920, pp. 56-58). "Simultaneity is the property of a group of natural elements which in some sense are components of a duration . . . a duration retains temporal thickness. . . .Instantaneousness is a complex logical concept of a procedure in thought by which constructed logical entities are produced for the sake of the simple expression in thought of properties of nature. Instantaneousnesss is the concept of all nature at an instant, where an instant is conceived as deprived of all temporal extension." The equilibria of rigorous economic analysis have instantaneousness. It is therefore questionable if one can put them into time and ask how temporal adjustments are made from one position of equilibrium to another, as each instantaneous position is conceived as containing within itself all the elements of adjustment. What we are searching for is a concept of simultaneity, whereby a temporal process can be conceived in such terms as to permit of generalization about the process of temporal adjustments.

[26] Sune Carlson, *A Study on the Pure Theory of Production* (London, 1939), pp. 34-35.

[27] Dr. James Dingwall contends that Dr. Carlson's argument reveals only one of many possible expansion paths to maximum profit. See his "Equilibrium and Process Analysis in the Traditional Theory of the Firm", *Canadian Journal of Economics and Political Science*, Vol. X, No. 4, Nov. 1944, pp. 448-63.

[28] Dingwall, op. cit., p. 455.

rium theory of the firm has been shown to be illegitimate. On the other hand, when stripped of such propositions, the equilibrium theory ceases to have explanatory content. In place of explanation must be substituted assumptions which reduce, in effect, to the assumption of an undescribed process which will in fact cause the firm to establish itself at the position labelled as equilibrium."[29] Dr. Dingwall finds the cause of this "sterile formalism" to be methodological and asserts the necessity of formulating "an explicit assumption" with respect to the nature of time. This is also our contention, viz., that economic processes are temporal and that what we are to mean by time must be explicitly stated and clearly defined.

2. Time

Our problem, here, is to show how time may be conceived so as to permit of the treatment, in terms of cause and effect, of an economic process. The methodological weakness of traditional process analysis in the theory of the firm, to which we referred in the previous section, is that formal equilibrium theory deduces, by *a priori* argument from a given set of limiting assumptions, which among other things *exclude* a causal sequence in time, a determinate equilibrium position. The firm, in the formal arguments, is not considered to reach this point through a series of adjustments over time; the formal argument simply defines the point (it is a typical, maximizing to a limit, process) which would be an equilibrium under the assumed rationale. If one wished to extend the formal concept further, the only plausible extension would be that real firms could achieve such points only by *immediate* adjustments. Since such immediacy is clearly impossible in the real world, it follows that any attempt to extend the formal theory to the study of process in the traditional sense must miss out the actual nature of the real adjustments which constitute a true temporal process. One cannot, surely, define the equilibrium position as that which is determined by marginal costs and marginal revenues within a model from which all temporal changes in the way

[29] Dingwall, op. cit., p. 463.

of alterations in the technical coefficients of production, or tastes or desires of the consuming public, are excluded, and in which all the functional relationships between output and costs and output and revenues are regarded as immediate, one-sided and irreversible as to sense, and at the same time attempt to say that the reason this equilibrium is achieved is because firms, in time, operating in a motive-cause nexus, will gradually move towards it. There are two different universes of discourse involved. The equilibrium concept is purely formal, is perfectly valid and precise, and is established by *a priori* argument that is quite independent of how firms, in fact, do adjust their outputs. That the concept is of great value we do not deny, but it is only of value for static analysis; it is not set up in a form that permits of immediate extension or application to the problems of firms in a continuous process of adjustment.

The concept of time is one of the most difficult of all ideas in the history of human thought. We all know what we mean by "time" in any given context. What time actually is, we do not know. We ordinarily think of it in terms of the units in which we measure it. Thus we speak of "the time" as being five o'clock, or we refer to a duration of so many hours, or days, or years, or even light-years. In all such cases we are talking of units of measure and not of what we measure; and though we commonly speak as though the artificial unit of measure were identical with pure time itself, we do even in common speech make a distinction between that which is measured and that which we measure, as when we say, "the time seemed very long but when I came to look at the clock (i.e., to measure it) it was only twenty minutes."

Sometimes, too, we refer to time in terms of the events that transpire in time, as when we say, "Oh yes, I remember, that was the time Johnny fell into the cess-pool", or "in good King Charles' glorious time". Once again, we may note, the events which transpire in time are capable of being distinguished from that in which it is that they transpire. What, to the Vicar of Bray, was "Good King Charles' glorious time" was to the Puritans and the Whigs a time of reaction and persecution and of national shame and humiliation. It was also Cardinal

Richelieu's "glorious time" and William of Orange's "glorious time", less "glorious", however, to John and Cornelius de Witt. When we speak of the "time" of Charles II we intend, then, to refer to a selected group of events; our reference is, in itself, a selection and limitation, we do not intend to refer to the abstract medium in which events flow, subsist or have their being, or even to the totality of such events in any period. Thus time to us is always distinct from the events which exist in it, as it is from the units in which it is measured; yet, for all practical purposes we do not bother with these distinctions, but use, as suits our convenience, the abstraction of units of measure, or the concrete reality of events and we let these notions stand for our idea of time.

These two common uses of the word "time" stand as fair prototypes for the more sophisticated notions of time used by the sciences and history. For the modern physicist time is a dimension. This concept of time is an exceedingly handy one to use, if one is versed in the techniques of modern physical science, but it is difficult, considered in abstraction, apart from its use, to understand. The concept of a dimension itself is easy enough to comprehend. We all know what we intend to mean when we say a thing has the dimensions of length, breadth and depth. Modern physics adds that an event has duration. To the mathematical physicist this is perfectly comprehensible, and is pragmatically justified by the ability it gives him to explain phenomena otherwise inexplicable. The concept is, thus, straightly derived from Aristotle's definition of time as the number or measure of motion;[30] in itself it is a formal and empty abstraction to which it is impossible to attach a concrete meaning, but as soon as units of measure are devised in terms of motion, the evolution of the earth upon its axis, the circumvalence of the earth by the moon or of the sun by the earth, time as dimension assumes the concreteness and obviousness of the length or breadth, in physical terms, of tangible masses. The formal concept, in itself, however, in the hands of the mathematician, rapidly loses any relation it may have to physical

[30] Aristotle, *Physics*, p. 223, 1.15.

(eventual) reality and becomes a mere dimension in a manifold system.[31] There is no reason, *a priori*, why that system should be restricted to one of four dimensions. This formal concept assumes the role of a methodological fiction. "The concept on which the whole of mechanics is based of *empty time* as a firm and lasting construction, a form as Kant also assumed it to be, is a fiction based on an abstractive and one-sided isolation. But it is instructive that empty space and empty time should be indispensable fictions for mechanics and for the theory of knowledge."[32] Formal time thus becomes so purely formal that it is simply a dimension, indistinguishable, in its formal sense, from any other. It may exist as a variable in a manifold of which the other dimensions are length, depth and breadth; it may exist in a manifold of which the other dimensions are prices and quantities of goods; or as one variable in a system of which the others may variously be simply x, y, and z. True

[31] There are more uses than one of time, even within the field of mathematical physics. Thus in dynamics time is a fundamental variable with only one sense. In quantum and relativity theory the "point of reference" is relative rather than absolute, and time may, for certain purposes, be regarded as non-continuous. Indeed, here, as elsewhere, the criterion for the methodological concept appears to be pragmatic.

[32] Vaihinger, *The Philosophy of "As If"*, tr. Ogden (London, 1924), p. 51. We might note that Kant's formal concept, as basic to his theory of knowledge, made time a purely *subjective* mode of apperception. "Time," he says, "is not an empirical concept that has been derived from any experience. For neither coexistence nor succession would ever come within our perception, if the representation of time were not presupposed as underlying them *a priori* . . . We cannot, in respect of appearances in general, remove time itself, though we can quite well think time as void of appearances. Time is, therefore, given *a priori* . . . Time is not a discursive, but a pure form of sensible intuition . . . " He concludes that "time is nothing but the form of inner sense", and defines this as "the immediate condition of inner appearances (of our souls), and thereby the mediate condition of outer appearances." (Kant, op. cit., pp. 74-77). By "sense" Kant here refers to "inner sense"; not the "external sense" of the affection of our physical organs by the phenomenal world, but to the affection of our mind by "things-in-themselves". When the mind perceives itself it can only be aware of itself as active, and requisite to its perception of activity is the sense of order or succession. It is thus this internal dimension of succession which is the "mode of apperception" called Time. The psychological implication of this view is foreign to the formal concept of time in modern mathematical theory, though some mathematical physicists, Eddington, for example, hold a somewhat similar notion. But there is no necessary reason in mathematics to regard time as subjective, or as an objective property of things. It may simply be regarded as a device or fiction of analysis, another axis in graphical representation, a further variable in functional relationships in a multi-dimensional manifold.

it has the property, according to Sir Arthur Eddington, of being irreversible, of having, in the meaning of a vector, only one *sense*, which implies, one would imagine, that it must always be the independent variable in any functional system. Ordinarily in physics and mechanics it is so treated. Whether or not this is so in pure mathematics is a question on which the present writer is not competent to pronounce, but there does not appear to be any necessary reason that it should be so.[33]

For the historian who is conscious of the phenomenon of time, it is a concept of concrete eventuality, corresponding to the second of our common notions. This stands in clear antithesis to the mathematical concept from which in its most formal presentation all eventuality has been abstracted away. "For him who is at home in becoming," writes Bergson, "duration appears as the very life of things, as the fundamental reality."[34] Time is no longer an independent mathematical variable capable of being viewed as a series of infinitely small increments. Such increments or "moments" do not endure; the "thread which bound them to time" has been cut and they have no reality; they are artificial fictions[35] implicit of contradictions. Duration is "lived time" and is the very stuff of reality. To abstract time from the reality of the temporal process is to find it nothing. Professor Whitehead suggests a similar line of approach. He is, in his own language, "in full accord with Bergson",[36] and finds that "the relations of other

[33] Mr. J. W. Dunne, *An Experiment with Time* (London, 1927) and *The Serial Universe* (London, 1934), as interpreted by Professor Stocks (op. cit.) appears to argue that the multiplication of dimensions in a physical manifold must always be by way of a multiplication of time itself, which he regards as being not a simple dimension, but as having the property of being raised to any power of itself. When you perceive time as the fourth dimension of a spatial system, you perceive it from an observational point in time. "The perception of these four dimensions will have its time", and this time, in turn, can become a fifth dimension of the system. This can go on indefinitely; but, however many dimensions within the system, there is always a further temporal point of reference or observation without it. But this view of the properties of the dimension of time cannot be regarded as characteristic of mathematics.

[34] Bergson, *Creative Evolution* (Paris, 1909), p. 343. As translated in Stocks, op. cit., where cited.

[35] See also Vaihinger, op. cit., who points out (p. 52) the fictional nature of the incremental treatment of time and the contradictions to which it leads, as, for example, the Zeno paradoxes.

[36] A. N. Whitehead, *The Concept of Nature* (Cambridge, 1920), p. 54.

events to this totality of nature form the texture of time".[37] Professor Whitehead, however, is able to assimilate this view of the reality of time to the mensurable and objective nature of serial time. What Professor Whitehead calls "passage", the passing of one duration into another which is the essential process of nature in his ontology, is not measurable "except as it occurs in nature in connection with extension",[39] but is rather a quality "which extends beyond nature" and relates the known to the knowing and hence to "ultimate metaphysical reality".[40] "The measurableness of time," on the other hand, "is derivative from the properties of durations . . . We shall find that there are in nature competing serial time-systems derived from different families of durations."[40] Thus there is a unity in the nature of real time as passage, and a whole group of serial times, which are abstractions depending on the properties it is desired to measure. The choice of formal or abstract time would seem to be pragmatically determined.

That all historians are Bergsonian we do not pretend, but, in effect, the historical view of time is that it is process in duration and the reality is the eventual process, *in extenso*, itself. That is to say, the reality of history is the living event or sequence of events and it is this enduring eventuality which constitutes time. This "psychological"[41] view of time, of the consciousness of events in duration, is clearly suited to the needs of the historian. We must, it seems, recognize that each science selects the concept which is adapted or adaptable to its requirements.

Now the economist has never been happy between these two views of time. He has jumped from one to the other, as suited his need, though generally he has taken the historical view, sometimes especially redefined for his purpose. Sometimes economists have developed a concept of time as a dimension,

---

[37] Ibid., p. 53.

[38] Ibid., p. 55.

[39] Ibid., p. 55.

[40] A. N. Whitehead, *The Concept of Nature* (Cambridge, 1920) p. 55.

[41] Cf. Sorokin, op. cit., p. 164. Sir Arthur Eddington in *Nature of the Physical World* (Cambridge, 1932), p. 40, also distinguishes between objective physical time ("Astronomer Royal's time") and subjective or "experienced" time.

but usually only in connection with one particular concept, usually the dimension of capital. Thus Jevons makes time a dimension of capital investment,[42] though elsewhere in the *Theory of Political Economy* time is either excluded from the functional processes[43] or is treated as a utility.[44] Time, invested with certain metaphysical properties relative to value, is treated as a dimension of capital also by Böhm-Bäwerk in the *Positive Theory of Capital.* In some modern works of analytical rigour time is banished so as to permit of mathematical treatment of demand and supply functions.[45] In effect the assumption here is that time consists of frictions and the determinate model is consequently removed from time to permit the immediacy of adjustments required to satisfy the functional equations.

The historical views of time adopted by economists fall into two categories. Most "pure" economic theory, following Marshall, is concerned with the concept of a "period of time", a period in which *nothing happens but adjustments* to various equilibria positions defined as being in the "short run" or "in the long run". The second view is that of cycle theorists whose concept of time is also historical, but in this case consists of a time which is constituted of disequilibrating forces, generally regarded, sometimes with an almost Aristotelian acceptance of rhythm as the stuff of history, as essentially cyclical.

The problem of the time period is a problem of selecting the period suitable to the sort of adjustments one wishes to observe. "Given the system of wants and the stocks of goods, their

[42] W. S. Jevons, *Theory of Political Economy* (4th ed. London, 1924), pp. 224-38 Time is also suggested as a dimension of labour (p. 179).

[43] W. S. Jevons, op. cit., pp. 100-01. His critical comment on Mill (p. 101) and his analogy to the theory of the lever both imply a static assumption, a removal of the model from time, in the sense that all adjustments are regarded as immediate. Here also the concept of time is mathematical and Jevons refers to the fact that he is "really considering" rates per unit of time. In this Jevons is following Cournot, whom he admits to have studied "but not mastered" because of the difficulty of some of the mathematical analysis. In brief, the mathematical economists have largely taken over, more or less unconsciously, the mathematical concept of time; the others have not.

[44] W. S. Jevons, op. cit., p. 176.

[45] See, for example, J. R. Hicks and R. G. D. Allen, "A Reconsideration of the Theory of Value, Part II. A Mathematical Theory of Individual Demand Functions", *Economica*, New Series, Vol. I, No. 2, May, 1934, pp. 147-219.

disposition will vary according as a short or long period of time is under consideration . . . the period of time for which one economises must be defined in order that conduct may be unequivocally determined".[46] Thus for Marshall "the short run" is a period which allows for the perfect adjustment of the price of a given stock of goods, regarded as being unalterable within the period, to the system of wants. The "long run" is a period in which the supply is adjusted to the demand so as to equate average costs of production to demand prices.[47] The mathematical theorists doubtless achieve greater precision be defining the model out of time altogether, but, to all intents and purposes, the Marshallian treatment achieves substantially the same model in a universe less drastically remote from the real world. Hence Marshall's concept of time is simply another analytical device for equilibrium analysis from which *change* is explicitly excluded.

In distinct contrast to this historical time in which the stationary state lives, moves and has its being, is the historical time of some cycle theorists. For them time is the cyclical process of the economy, the "ceaseless ebb and flow" of the cycle of prices, investment and employment. With few exceptions, until recently, cycle theorists were never concerned to relate the cycles of disequilibrium which they studied to the equilibrium concepts of "pure theory". The two sorts of theory existed in quite separate, watertight compartments, and bewildered students were one year taught that the fundamental forces of the economy were always tending to bring it to a position of rest or equilibrium in time, and were only prevented from achieving a perfect adjustment by certain "temporal

[46] P. N. Rosenstein-Rodan, "The Role of Time in Economic Theory", *Economica*, New Series, Vol. I, No. 1, Feb. 1934, p. 78. Mr. Rosenstein-Rodan's paper contains an interesting review of the literature on the temporal concept in Economics.

[47] "The nature of the equilibrium itself, and that of the causes by which it is determined, depend on the length of the period over which the market is taken to extend. We shall find that if the period is short, the supply is limited to the stores that happen to be at hand; if the period is longer the supply will be influenced, more or less, by the cost of producing the labour and the material things required for producing the commodity", A. Marshall, *Principles of Economics*, 8th ed. (London, 1925), p. 330. Cf. also R. Opie, "Marshall's Time Analysis", *Economic Journal*, Vol. XLI, June, 1931.

frictions" and "institutional rigidities"; another year, however, they learned that the essential and fundamental temporal process of the economy was cyclical and that it was in a state of chronic disequilibrium. The two notions were not necessarily contradictory, but the failure to relate them was confusing. Mr. Harrod makes a conscious effort in his *Trade Cycle*[48] to assimilate the two systems in one temporal universe, showing certain forces to be "stabilizers" and certain others to be "destabilizers", and to explain the trade cycle as the resultant of the two opposing sets of forces, all seen in the same time series. It is, however, the work of Lord Keynes[49] which has most succeeded in bringing economists to recognize that the equilibrium of a market in a "stationary state" is but a partial equilibrium set in a system that moves, in time, through a whole series of possible positions of general equilibrium in the sense of satisfying the economic rationale with given rates of investment and employment, other things remaining unchanged. The method of Lord Keynes is to take a series of "instants" or "periods", each one of which is static in itself and to compare the various equilibria so obtained and so to infer the causes of shifts or adjustments in equilibria. Thus Lord Keynes, in effect, achieves his synthesis, by adopting what is essentially a static method, a method of comparative statics, and a Marshallian view of time. In each of his instants or periods nothing else happens but the adjustment to equilibrium. Now it is of the very essence of time that "other things" cannot remain unchanged. The increase (or decrease) of income occasioned by any level of employment brought into adjustment with a given rate of investment causes an increase (or decrease) of the marginal propensity to save, with consequent effects on the rate of spending, the size of the multiplier and, in turn, on the rate of new investment. Moreover, all such shifts in general equilibrium cause shifts in the terms of employment of the factors of production, the combination and allocation of the factors, the supply of goods and the structure of consumers' demand, thus causing

[48] R. F. Harrod, *The Trade Cycle* (Oxford, 1936).

[49] We refer, of course, to the *General Theory of Employment, Interest and Money* (London, 1936).

shifts in all partial equilibria and setting up frictions and disequilibrating forces that generate a continuous dynamic. It was such considerations that led Professor Myrdal to say of Professor Hansen's Keynes-like calculations of the volume of building necessary to maintain "full employment" in the United States, "these calculations are worthless. The analysis as a whole is altogether static. It is not at all concerned how the whole procedure is to be brought about . . . by developments over a period of time."[50] It is not the value of the Keynesian or Hansenian analysis that is under criticism. It is the attempted application of this analysis to problems of policy without appreciating that it is essentially a static analysis and that in the real temporal order dynamic forces are present which are outside the analysis.

Indeed, I suspect there are some serious confusions in Lord Keynes's use of the concept of time, and that some of the difficulty in understanding him arises from a lack of agreement between his definitions and his usages. We shall later refer to the fact that he defines income, expenditure, savings and investment as primitive functions of time, but treats them as rates in time. The ambiguity in the concept of the marginal propensity to consume comes from doubt as to whether it is a first or second derivative of income, and Lord Keynes himself admits that the effects of a changed marginal propensity to consume on investment and income cause further changes in the marginal propensity to consume.

Again, as Dr. Littler has pointed out,[51] whereas in a stationary period, the quanta of real savings must equal the quanta of real investment, over a dynamic period, the rates of monetary savings and investment will not be equal, and must in fact differ by the amount of the change in the rate of consumption multiplied by the number of time units in the period. This is to say that the dynamic disequilibrium of savings and investment is part of the concreteness of time, and is, indeed, quantitatively related to the magnitude of the time dimension.

[50] Gunnar Myrdal, "Economic Development and Prospects in America", address before the National Economic Society of Sweden, March 9, 1944.

[51] H. G. Littler, "A Pure Theory of Money", *Canadian Journal of Economics and Political Science*, Vol. X, No. 4, Nov. 1944, p. 432.

However this may be, we believe that the selection of the time in which an event is to be studied must depend on what that event is. Our only contention is that the traditional economic "times" are not the proper ones to select for the analysis of economic change and that the attempt to carry through process analysis within the limits of these assumptions as to the nature of time is bound to become confusing and ambiguous. As we saw, in Section I of this chapter, process analysis of the firm is impossible within the limits set by the static or stationary assumptions. The analysis of changes in institutions, of economic change in the broadest sense, is likewise impossible within the temporal framework of the Keynesian system. (No intention, however, exists to offer solace and comfort to those critics of economic analysis who deny price equilibrium analysis or general equilibrium theory as being "unreal" or "impracticable".)

It would appear that the judgment of the proper concept of time is pragmatic. The philosophic problem of time seems insoluble in its own terms, for the more formal and abstract time becomes, the more it approaches the working concept of mathematics. The sciences have developed the concrete notions of time suitable for the analysis of their own proper problems. The real issue is not what time absolutely is, but rather the kind of time that we are dealing with. Our belief is that for the treatment of static equilibrium the mathematical concept of time is the most exact for the kind of economic analysis involved, but that when one comes to the study of economic change a more concrete and historical concept of time is necessary. Here the question is one of selection. Of what sort of process may time be said to consist? Is time an eight-to-ten-year cycle of the money flows of income and investment? Is it a sequence of short-period adjustments from one equilibrium to another during which nothing else is regarded as happening? Is it the full content of events, economic, cultural, and political which the general historian observes?

The answers to these questions depend, we believe, on the intention of inquiry.[52] Abstraction depends on selection, and selection, in turn, on purpose. Since our purpose here is to

inquire into certain types of comparatively long-run economic changes, we must define our time in terms of the pattern of change we wish to observe. Time for us shall consist of processes of change, the units in which it may be measured being purely arbitrary; and where we observe only one selected process, other events being disregarded, we shall have "partial time", capable of being treated as a dimension measurable in its own functional units; and when we combine the processes we shall have full economic time, to be regarded as the sum total of economic processes, and measurable only chronologically as calendar time.

For our purposes this time must be treated as continuous. The processes we wish to observe flow on in an unbroken series and the time which contains them is itself a continuous flow which may, however, for special purposes be broken up into arbitrarily determined convenient periods. Moreover, since we wish to select certain of these processes and to regard them for analytical purposes as the sole constituents of their time, and since we wish to reach certain generalizations of a quantitative nature about them, we shall want to view time as a continuous variable in a functional system. We believe we are justified on these grounds in electing to treat time as a continuous variable. Discontinuous variation may be the character of certain of the processes we wish to observe. In such cases, if our pragmatic view is accepted, we shall be justified in shifting to a concept of time changing discontinuously throughout the process.

Thus in the consideration of certain long-run changes in demand we shall find a continuous variation and a high degree of generality in the form or pattern of change. In the theory of production changes we may well find a discontinuous process of innovations followed by adaptations. When we return to the theory of the firm we may be able to treat time once again as continuous and to observe in a highly simplified temporal model a generalized process of change with important consequences to the structure of the market.

[52] Compare P. A. Sorokin, *Sociocultural Causality, Space, Time* (Durham, N.C. 1943) Ch. IV. Professor Sorokin's book came to my hands only after this chapter was written. The similarity of view is gratifying to the present writer. Cf. also A. N. Whitehead, op. cit.

PART II

THE GENERAL THEORY OF ECONOMIC CHANGE

CHAPTER IV

THEORIES OF GIVEN CAUSES

1. GIVEN CAUSES

The distinction we have made between causes which are internal to the economic order and those which are external to it corresponds with the distinction often made between exogenous factors or causes and endogenous factors or causes.[1] Certain of the causes we classify as external to the economic order, such as the illustration of the Tokyo earthquake used by Professor Schumpeter, are clearly accidental and non-recurring and hence not susceptible of scientific treatment. Others, like wars, however profound their economic effects, similarly defy systematic treatment because of the complexity and lack of uniformity of the non-economic forces entering into their composition, although in these cases the causes are not entirely non-economic in origin. Wars, indeed, are often in part effects of causes inherent in the economic process.

There is finally a group of causes acting on the economic process in a direct and primary way, which in economic literature has always been accorded a special significance. Whether these causes are or are not external to the economic order we shall leave for the moment undecided. They consist of causes constituted of physical or biological or natural phenomena of a

[1] Cf. Professor Schumpeter's footnote on his distinction between "external" and "internal" factors. "It is permissible," he says, "to draw a line between the phenomena directly incident to the working of the economic system and the phenomena produced by other social agencies acting *on* the economic system . . . " In the footnote Professor Schumpeter says that this distinction is not precisely the same as that drawn by Spiethoff between endogenous and exogenous factors (in cycle theory), "but the principle of analysis underlying all these and similar distinctions is the same. They are meant to express the fact that what we are faced with is an economic process disturbed by events not inherent in it. What precisely is inherent in it will, of course, depend on how we delimit it . . . ", J. A. Schumpeter, *Business Cycles* (New York, 1934), p. 7.

certain regularity of behaviour, having profound economic consequences. They stand to the theory of economic change in much the same relation as the physical law of diminishing returns or the psychological or physiological law of diminishing utility stands to static economic theory. In other words, these phenomena are susceptible of certain broad generalizations, from which, in turn, conclusions about economic consequences flow. The economist does not inquire into these particularly, how they come to occur, why they are as they are; ordinarily he has taken them as given, and we shall consequently call them "given causes of economic change". We should note, as a matter of great consequence to the method of analysis later to be used, that there is one important difference in logic between the relation of these phenomena to the theory of economic change and of such phenomena as diminishing returns to the theory of economic statics. Whether the law of diminishing returns is empirically true or not, the deductions of economic theory are true and certain within the framework set by the assumption of diminishing returns. The theory is logically sound, and that is all it pretends to be. "Every economic theorem may be expressed in the form of a syllogism having as its major or minor premise the hedonic hypothesis, and for its other premise some matter of fact, which may be a truth borrowed from other science . . . The demonstration of the truth of these premises pertains to the science to which they respectively belong . . . "[2] Now the conclusions which follow from postulating the existence in some defined form of "given causes" of change do not follow *logically*, i.e., with strict determinability, from the postulated conditions. They are, rather, inferences, of various degrees of probability, as to the consequences which, in empirical fact, will flow from the given cause. Thus the method for their analysis cannot be the strictly rigorous method of static inquiry. It must either be a purely historical method, the inductive study of various historical processes in which these causes have been present—an unsatisfactory method by itself because of the mixed causation always present in an historical process and

[2] M. Pantaleoni, *Pure Economics* (London, 1898), pp. 3-4.

because of the impossibility of generalization[3]—, or it must be a method of setting up simplified models of change in which given causes are postulated and the probable consequences inferred rather than demonstrated. It is this latter method which will be attempted in this work.

Causes which have figured prominently in the literature as given causes of economic change are technological innovations, changes in population and geographical discovery. These causes are ordinarily taken as given in some form or state, that is, population is increasing, or innovations are appearing in investment-creating form, and certain economic consequences are then inferred to flow from them. Few economists have been content, however, to regard these given phenomena as themselves external to the order of economic causation. Ricardo believed, for example, that the rate of population growth, after creating certain economic effects, would in turn be affected by its own consequences. The rate of innovations can scarcely be regarded as being independent of causes within the economic order. To use an illustration, if we have a process represented by A, B, C, . . . M, and A is given as having changed, we may then inquire how M is changed. But when A is changed it may effect B or C, and, in any case, one can scarcely claim to have established a general theory of economic change if the changes in A are left unexplained. We have also the problem of the interaction of causes, the problem of how changes in A may affect B, C, etc., and how, in consequence, the sum effect on M will manifest itself.

The general theory of economic change must then explain (a) how changes in given causes affect the economic order, (b) how given causes react upon one another and, in sum, on the economic order, and (c), if the given causes are to be admitted to be not entirely external to the economic order, how, in turn,

[3] The uniqueness of economic situations has always been considered by competent economists as rendering the historical method inappropriate. "In economic or other social problems", Marshall wrote, "no event has ever been an exact precedent for another", A. C. Pigou (ed.) *Memorials of Alfred Marshall* (London, 1925), p. 166. Cited by E. R. Walker, *From Economic Theory to Policy* (Chicago, 1943), p. 142.

changes in the given phenomena come to occur as a result of the economic process itself.

It has been said, and with great authority, that this kind of inquiry is beyond the boundaries of theoretic economics. Professor Schumpeter protests against "the metaphysical treatment of social development", and believes that it is doomed to be "unscientific" and "mystical". He believes further that "it is not possible to explain *economic* change by previous *economic* conditions alone. For the economic state of a people does not emerge simply from the preceding economic conditions, but only from the preceding total situation".[4] To enter this field of inquiry, Professor Schumpeter consequently urges, is to abandon the general scientific rule that each science traces back the causal sequence to the point where the next step in the sequence would carry it to a cause external to the proper order of the science in question. Thus he argues, "if economists, nevertheless, have always had something to say on this theme (i.e., the study of the basic causes of economic change in their full historical context) that is only because they did not restrict themselves to economic theory, but—and indeed quite superficially as a rule—studied historical sociology or made assumptions about the economic future".[5] This severe indictment should, no doubt, be a sufficiently deterring warning to any investigator contemplating the attractions of this forbidden garden, but Professor Schumpeter himself opens the way to temptation. In the same footnote, from which we have already quoted, he allows that Karl Marx did indeed construct a genuinely "economic" theory of economic change, finding "an *internal* economic development and no mere adaptation of economic life to changing data".[6] This can surely also be said of Ricardo, and, I strongly suspect, of Professor Schumpeter's work itself. However clearly we try to define the boundaries of economic science, so profound and pervasive are economic forces in the total social matrix, and so sensitive is

[4] J. A. Schumpeter, *The Theory of Economic Development*, tr. Opie. (Eng. ed., Cambridge, Mass., 1923), p. 58.

[5] J. A. Schumpeter, op. cit., pp. 59-60 fn.

[6] Ibid.

the "purely economic" in response to "non-economic" stimuli, we are obliged to take note of the effects the economic process produces in the social evolution, and of the latter as a cause of economic change. We shall inquire in this chapter how theories of change from given causes have developed a theory of "internal economic development", bringing the causal factors within the compass of economic explanation.

We shall also inquire how far preoccupation with a single given cause has precluded attention to what we have called the interaction of causes. Whereas business cycle theory has, at any rate of late, taken proper notice of the complexity of economic causation in short-run fluctuations, there has been a tendency in the general theory of economic change to forget that "as the motion of every body in the solar system affects and is affected by the motion of every other, so it is with the elements of political economy".[7] As Professor Schumpeter has said, any economic state proceeds from the social totality of preceding states. Moreover, it is in turn a member of a social order, a distinguishable member, it is true, which is in a continuous movement of which all the members take part, interacting with one another. The distinctions which analysis requires of phenomena, which like bread and cheese are "separable only in thought", must always do violence to the delicately complicated pattern of historical reality. Yet it is surely clear, even if one can never compass in thought or words the full complex of historical reality, that there is an order of causal relationships, that some causes are more dominant than others, and that, by proper selection, the process may be rendered intelligible. There, indeed, is the paradox of all historical scholarship. Intelligibility, which is the principle of explanation, can only be achieved by a kind of violence to "reality". Yet "reality", of itself, is unintelligible and consequently, within the human universe, unreal.

The economists we are about to examine have all given some attention to the complexity and interaction of causes, but have emphasized one given cause, that which was most clearly op-

[7] A. Marshall, "Review of Jevon's 'The Theory of Political Economy' ", *The Academy*, 1st April, 1872.

erative in the economy of their day. We shall strive to show that this concentration was faulty, and that a general theory of economic change must provide an integrating factor, explaining the interaction of the causes on one another, and their position within the economic order as a whole.

2. Some Theories Based on Given Causes

(I should like to express my indebtedness to Mrs. Ruth Albert for the freedom with which I have used her thesis, "Some Theories of Economic Change" in preparing this section. I am never quite sure how far instructors are justified in using the work of graduate students. I know the practice is often abused. I hope that is not the case here. Mrs. Albert has been working under my direction and we have repeatedly discussed the authors here studied. It is hard to be sure what ideas are hers and what mine. She has permitted me to read the MS. of her thesis, which I have found most helpful, and many parts of this section are based directly on it. It is indeed a case of joint authorship, except that I have exercised the prerogative of final decision, and so, on doubtful points of interpretation, must assume single responsibility.)

We propose in this chapter to examine certain theories of economic change based on certain selected given causes. These theories are all more or less systematic not only in the sense that they afford a comprehensive explanation of the economic consequences which flow from an existing given cause or causal state, but also in the sense that they comprehend, too, the cause of change itself and relate its origins to the economic process, so that the entire process is explained in economic terms.

It is to the classicists that one naturally turns for comprehensive theories of economic change, because one of the achievements of the Austrian revolution was to limit pure economic theory, i.e., systematic thinking about economics, to the logical analysis of certain market situations based on two or three highly general and simplifying assumptions. The success of this revolution, partly a matter of reaction of sophisticated theoretic minds to the errors of the German historical school,

cut economics off from the old interest in the progress of society in wealth and welfare.

ADAM SMITH

It was this very question of what contributes to the progress of society in welfare that formed the kernel of Adam Smith's work. His purpose is clearly stated in his title, *An Inquiry into the Nature and Causes of the Wealth of Nations.* Whereas modern economists usually accept Professor Robbins' statement[8] that economics is the study of the disposal or allocation of scarce means to a multiplicity of ends, Adam Smith was in the first instance concerned with the problem of how the scarce means could be increased in numbers and efficiency. Such intent required a study of change, for progress is a temporal concept—the accumulation of capital stock is a phenomenon of the time process, a change in the conditions of the economy, not an adjustment to the conditions of a static equilibrium. In his Introduction, Smith defines as the subject of Book I, "The causes of (this) improvement in the productive powers of labour, and the order, according to which its produce is naturally distributed among the different ranks and conditions of men in the society."[9] He regards the size of the labour force, "its skill, dexterity and judgment", its allocation and the division of labour all as playing a part in governing the size of the national product, and any changes in these factors, making for an improvement in the productive power of labour, are causes of progress.[10] Smith does not believe, however, that these factors are important as *independent* causes of change. He does not regard them, that is, as changing independently of the economic

[8] L. Robbins, *An Essay on the Nature and Significance of Economic Science* (London, 1932).

[9] A. Smith, *The Wealth of Nations* (Everyman ed. London, 1911), p. 2.

[10] As we shall shortly see, these are not "exogenous" factors for Smith, but are regarded by him as coming properly within the sphere of economic speculation. Except for a characteristic disregard for demand forces Smith thus includes as phenomena "in regard to which change or the possibility of change must be studied" (F. H. Knight, *Risk Uncertainty and Profit*, Boston, 1921) the same factors as Professor Knight lists (op. cit., p. 147).

situation, or according to some "natural" law of their own.[11] For him, writing as he was in the generation of the great power revolution in Britain's basic textile trades, technical innovations were the independent given cause, and their introduction forms the integrating concept in his theory. Technical innovation is the cause and condition of more and more productive division of labour and the development of more specialized skills (indeed in Book I, the division of labour is almost indentified with innovation, in the broadest sense of the term),[12] it is, as we shall show, practically indentified with capital accumulation, and is the cause of population increase. Thus the given (and, for Smith, exogenous) cause of progress in wealth is innovation, and the other factors governing economic development are included in the range of systematic economic inquiry.

Wealth (measured by Smith as annual per capita consumption) is increased when the productive power of labour is increased. Smith is thus not interested in market or general equilibrium and his interest in exchange and market price is quite clearly only an incidental sub-thesis introduced to show that the total money value of the national produce is an unreliable measure of the wealth (real national income) of the society. We may well today prefer to distinguish wealth, as the stock of goods, from real income, but modern economists will not dispute Smith's insistence on the size of the real national product as the criterion of welfare.

The basic and essential cause of the improvement in welfare was the increasing division of labour which we may identify with innovation. Innovation occurs in conjuncture with, and can be understood only in relation to, capital accumulation. It is hard to see in Smith's work any clear distinction between the notion of capital accumulation, as such, and the notion of

[11] It is true that in the first paragraph of Ch. II (Of the Principle which gives Occasion to the Division of Labour) he finds the origin of the division of labour not in "human wisdom", but in the "propensity in human nature . . . to truck, barter and exchange".

[12] In spite of the assertion referred to in footnote 11 that the division of labour originates in exchange, this statement is correct, because Smith shows exchange to be conditioned by extent of the market, the use of money, and innovations of a mechanical sort. " . . . everybody must be sensible how much labour is facilitated and abridged by the application of proper machinery" (op. cit., p.9).

qualitative improvement in the capital. He views the process of improvement as a gradual, continuous, and almost necessary, process.[13] Capital accumulation is a condition, indeed, of the division of labour, because, for Smith, a surplus must be accumulated before a man can be employed producing one article or part thereof only, rather than the full variety necessary to support life. His concept of saving implies that accumulation is the first stage of specialization, the division of labour, and advance in welfare. This very accumulation, however, is made possible, or at least very much easier, by technical improvement. "As the accumulation of stock is previously necessary for carrying on this great improvement in the productive powers of labour, so that accumulation naturally tends to this improvement . . ."[14] Thus the accumulation of capital, real investment, is, for Smith, attended by technical advance, and is the occasion of the general progress in welfare.

What is the nature of this capital accumulation, which, allied with innovation, is the basis of economic progress? For Smith it was not, as it was for Marx, an increase in the ratio of fixed to variable capital, or, as the casual reader might assume, simply an increase in the number of machines. Rather it was the process of real saving, or new investment in all productive factors.[15] Now the improvement in productivity that Smith observed would follow naturally from innovations, but it would follow from new investment only if the new investment was in superior durable goods. Smith, therefore, assumed that investment occurred because of good investment opportunities created by innovation, an assumption which brings his theory of development very close to that of Professor Schumpeter.

[13] I qualify the "necessary", because Smith did observe and comment on "mature economies" in which the process of capital accumulation had come to a halt. Like Professor Hansen of our own day, Smith believed the rate of new capital accumulation to govern welfare, not the total quantum of already amassed goods. Since these mature economies failed to score further advances, Smith realized that the progress of society was neither necessary nor inevitable, yet his spirit was so unshakeably optimistic that the general tone of his writing was to suggest, at the least, that there was no necessary or foreseeable limit to innovation in the west.

[14] Adam Smith, op. cit., p. 242.

[15] Smith ordinarily speaks of capital as "stock", and obviously has a mercantile concept in mind. This was very natural in a period when the merchant combined with the function of trade that of manufacturer as well.

He never seemed, however, to observe that increased productivity per man might bring unemployment. Say's law seems to have been implicit in his mind, for he certainly assumes that increased total production would result in increased consumption, and that no disadvantage from what we shall later call the "aggregative effects" would tend to offset the positive "real effects". That is to say, he did not envisage the possibility of a failure of aggregate demand to take the increased production at prices which would stimulate, or even maintain, investment, a failure that might lead to a periodic cessation of investment and unemployment, which might from time to time offset the progressive welfare development of increasing productivity.

This view is further supported by his argument on distribution. Smith believed that real wages improved with capital accumulation and innovation, but that the long-run effect on profits was to diminish the rate at which they were earned on investment. "It is the great multiplication of the productions of all the different arts, in consequence of the division of labour, which occasions, in a well-governed society, the universal opulence which extends itself to the lowest ranks of the people."[16] This is because the demand created by increased capital stock for the employment of labour leads to higher money wages. "The demand for those who live by wages, therefore, necessarily increases with the increase of revenue and stock of every country, and cannot increase without it. The increase of revenue and stock is the increase of national wealth . . . It is not the actual greatness of national wealth but its continual increase which occasions a rise in the wages of labour."[17] That higher money wages lead only to higher prices, a fallacy revived today in the newspapers of America as an argument against the demands of trades unions, Smith resolutely denied, and pointed out that the very people who alleged this to be true also complained of the luxury and feckless spending of the workers. "The common complaint that luxury extends itself to the lowest ranks of the people, and that the labouring poor will not now be contented with the same food, clothing and lodging which satisfied them

[16] A. Smith, op. cit., p. 10.

[17] Ibid., pp. 61-62.

in former times, may convince us that it is not the money price of labour only, but its real recompense, which has been augmented."[18] And so it should be, Smith believed, because the increased productivity of labour also lowered the real price of the commodity manufactured. He was aware, too, of the stimulating effects of higher wages which he believed would stimulate further production. "The liberal reward of labour, by enabling them to provide better for their children, and consequently to bring up a greater number, naturally tends to widen and extend those limits (of the market)."[19]

Smith believed, on the other hand, that profits tended to fall with increased investment. Smith's argument here seems to turn on his concept of capital as increased investment in productive labour. Thus any employment afforded to capital is employment afforded to labour. The increased competition of capitalists (with increasing capital accumulation) leads, then, to more employment of labour at higher wages, and to lower profits. Smith cites the low interest rate in Holland (he never distinguishes between interest and profits) as empirical proof of this proposition. Clearly Smith not only ignores the possible cyclical effects of innovations if there should be any break in their rate of introduction, he also ignores the relationship, later stressed by Professor Schumpeter and Professor Robertson, between the possibility of enhanced profits and the adoption of an innovation.

In a similar manner Smith, when he considers population, sees it as reacting passively to changes in the rate of capital accumulation and does not allow population change to act as an independent cause of economic change. The facts, so far as they were known, of English population growth in the eighteenth century were undoubtedly explained by his hypothesis. Capital accumulation, increased productivity and wealth led to a higher standard of living and an increased population. Smith

[18] Ibid., p. 70. Smith was also sceptical of another argument not uncommonly heard today. "What improves the circumstances of the greater part can never be regarded as an inconveniency to the whole. No society can surely be flourishing and happy, of which the far greater part of the members are poor and miserable" (p. 70).

[19] Ibid., p. 71.

also does observe the stimulating effects of increased demand (from increased population) on production,[20] but, generally speaking, he treats this as following either the effects of innovation in the domestic economy or the extension of the market by extending the frontier by foreign commerce, geographical discovery or conquest.

Thus Smith's theory of development hinges on the steady introduction of innovations with capital accumulation, the whole depending on a certain natural "propensity" of mankind. This process, once initiated, leads through higher real wages and improved productivity, to a widening of the limits of the market by reason of population growth and mercantile adventuring and discovery, and is so accelerated. The cumulative process is seen as a steady development without recessions, accelerating in young and rich countries as the original cause is strengthened by the growth of population and expansion of commerce.

To this Smith added a consideration, to which we attach great emphasis, viz., that the institutions of human society conditioned and limited or stimulated the entire process. Smith was aware of many societies which, in spite of the natural propensity towards innovation (division of labour) and of great natural wealth, had failed over centuries to achieve economic progress. He attributed that failure to the stultifying effects of their political and economic institutions. He believed, and the whole orientation of his argument and distribution of emphasis is to bring out forcefully this point, that England's progress in wealth and welfare depended on a free market and a commerce free from crippling, hampering and frustrating taxes and restriction. The process was a "natural" one only in a society which gave the propensities of men free play. It was not inevitable in the sense that it would happen anyway. Progress was pos-

[20] Professor Alvin Hansen says that Adam Smith "regarded growth of population as at once a consequence and a cause of economic progress" ("Economic Progress and Declining Population Growth", *American Economic Review*, March, 1939). This is true, but the emphasis is faulty. Population growth was treated as a consequence, which, in turn, acting as a contributary cause, accelerated the rate of progress, which, however, was originated and sustained by capital accumulation.

sible only in a society with sufficient wisdom to permit the free play of the initiating and competitive faculties which generate progress. For Smith human purpose and rationality entered directly and fundamentally into the complex structure of social causation.

Now this is a comprehensive and integrated theory of economic change. There are, however, certain modifications which must be made in the light of subsequent experience.

In the first place, Smith fails to explain the existence of his given cause in economic terms. Professor Schumpeter, if we may believe the footnote we have cited (p. 66), would not agree that this is a weakness. He would say that Smith bases his given cause on the psychology of man and that to trace it further would be to press his inquiry beyond the proper confines of political economy. We must protest, however, against this view. The "given cause" is not given. The rate of innovation, as Professor Schumpeter himself has urged, is definitely affected by economic circumstance and cannot be treated as external to the order of economic causation. This can be more clearly seen if we trace two corollaries that follow from Smith's position. The first of these is that the division of labour and innovation develop steadily at a more or less constant rate, so that progress proceeds in a straight line. The inappropriateness of Smith's theory to business cycle analysis and his consequent complete neglect of the cycle, a phenomenon already clearly apparent in the economy of his time, follow from this. A second consequence of this failure properly to explore the origins of division of labour and innovation is the rather silly theory of profit that Smith entertained in the days when, as Horace Walpole said, "Such fortunes are made and lost every day as are past belief. Our history will appear a gigantic lie hereafter, when we are shrunk again to our little island."[21] Only if one ignored the relation between profit opportunities and the rate of investment, and supposed the latter to continue independent of all economic circumstances could we maintain that the rate of profit was falling and must continue to fall. Some decline in the *rate of*

[21] To Sir Horace Mann, July 19, 1769. Horace Walpole's *Selected Letters* (J. M. Dent ed., London, 1926), p. 416.

interest had no doubt taken place with the reduction of risk. The opportunity of enhanced profit remains the lure that leads enterprisers to introduce new inventions or techniques and hence to turn an invention into an innovation. I am not impressed with Mrs. Albert's argument that the facts of eighteenth century agricultural development conformed to Smith's views. She says, "It is only when you have imperfect competition that the line of profits makes a great incentive to innovation. Yet wealthy landowners (of the eighteenth century) were imbued with a spirit of change, and 'high farming' was a great fashion. They really saved and invested their savings in improvements that resulted in a complete revolution in methods. Aside from profits it was 'curiosity, a new form of life and leisure, a fashion, a love of progress for progress' sake' (Professor Heaton's *Economic History of Europe* is here cited) that made these results possible."[22] Though there was this element of fashionable scientific curiosity in, for example, Lord Townsend's experiments, the real pressure for agrarian technical progress was economic and came from the development of the woollen trade, on the one hand, which reduced acreage under tillage and, on the other hand, the growth of the mercantile and manufacturing towns which reduced the agricultural labour supply and at the same time increased demand for foods.[23] Even though perfect competition rapidly wipes out excess profits, the lure of profit operates; temporary surpluses may be achieved, and, in any case, if a competitor fails to keep abreast of others technically, he will soon be surpassed and driven beneath the margin.

We must therefore conclude that Smith's theory of profits is inadequate and its inadequacy derives from his failure to perceive the economic relationship between the development of the economy and the introduction of innovations.

The second main objection we should take to Smith's general position is his failure to give population (and demand forces) any independent status or originating function in economic

[22] Ruth Albert, "Some Theories of Economic Change", McGill University Thesis, Redpath Library, McGill, p. 7.

[23] Cf. A. H. Johnson, *The Disappearance of the Small Landowner* (Oxford, 1909); P. Mantoux, *The Industrial Revolution in England in the 18th Century* (London, 1928); and E. Lipson, *Economic History of England* (London, 1931).

development. His treatment of population is, as a result, confused. On page 70 (J. M. Dent ed.) he points out that poverty and large families are correlated. On page 71 he argues that higher wages mean larger families. He is bound to try to show that population changes follow and support the operation of the given cause. He is caught up by the fact that population sometimes behaves in an independent and opposite fashion and seems, as it were, recalcitrant. A corollary of this over-simplification, and of his general neglect of demand, is Smith's preoccupation with "real effects", i.e., with the real income effects of increased productivity, and his neglect of the effects of alterations in aggregate money demands. We shall later argue that any theory which fails to comprehend and relate the two modes through which effects of economic causes are expressed throughout the economy cannot be a comprehensive theory of economic change.

DAVID RICARDO

If Adam Smith laid too little emphasis on population changes as an independent cause of economic change, David Ricardo, writing after the Napoleonic period, disillusioned by the social wreckage that had attended the first stage of the industrial revolution, impressed by the monstrous growth of the new industrial towns and in contact with Malthus, gave to population the major place in his theory of economic change. Moreover, Ricardo was as pessimistic as Smith was optimistic.[24] It is interesting to speculate on the effects of the prevailing long-run trend on the attitudes of the different economists. Smith and Professor Schumpeter (*Theory of Economic Development*, 1906) were writing in the upswing phases of long-run trends. Ricardo,

[24] I do not mean to suggest that Ricardo's pessimism was derived from Malthus. Indeed it is a popular mistake to regard Malthus's population theory as enshrining either a philosophical or an historical pessimism. Malthus scorned the easy optimism of Condorcet and Godwin, believing that material and spiritual progress were difficult of attainment and limited, the one by the niggardliness of nature, the other by the evil propensities of human kind. Nevertheless he did believe in the idea of progress, his acceptance of the necessity of evil had more of the cheerful rationalism of the eighteenth century than of the gloomy Christian doctrine of original sin in it, and he placed great confidence in the potentiality of "physical discovery". His population theory did not take the form of Ricardo's, and he

Marx and now Professor Hansen have written in the downswing phases.

In Ricardo's day the increased rate of industrial development since 1776, the aftermath of a world war and a profound revolution, had combined to produce serious economic dislocations. Over-production from the simple cause of lack of accurate anticipation of the market was common. There was inflation in the chief countries of commerce. Machine-breaking riots were rife, along with straight political discontent. While weavers and some other skilled artisans earned extravagant wages, and while new industrial fortunes were coined, the masses of the people lived in poverty and misery which, if no more acute than had previously been known, was both more squalid and more discernible. Ricardo had little reason to believe that the increased division of labour and its increased productivity naturally brought about a greater opulence shared by "the labouring poor".

Nor did prices indicate that goods made with less labour were cheaper for the people. Inflation at home and abroad, and the shutting off for years of foreign sources of corn, had, with a rapidly increasing population, reduced the margin of cultivation, increased the price of corn more than the increase in wages, and brought England face to face with the problem of maintaining her increasing population with her own agriculture, a problem ultimately to be solved by the repeal of the Corn Laws.

Politically, Ricardo was friendly to the Whig interest, though he sat on the government benches. His chief political opponents were the squirearchy who formed the voting strength behind the mixture of King's toadies and "old Whigs" who had recreated the Tory party. It is, perhaps, no accident that Ricardo's political economy should constitute a condemnation

---

did not believe in the inevitable return of real wages to bare subsistence level after any improvement under the pressure of increased population. Indeed he contributed in a positive way to the formation of a qualified and sober version of the theory of progress that came to dominate men's minds in the nineteenth century. Compare Joseph J. Spengler, "Malthus's Total Population Theory; A Restatement and Reappraisal," *Canadian Journal of Economics and Political Science*, Vol. XI, No. 2, May, 1945, pp. 246-55, and the authorities therein cited, especially A. O. Lovejoy, *The Great Chain of Being* (Cambridge, 1936) and J. B. Bury, *The Idea of Progress* (London, 1924).

of the agricultural interest, showing the landlord class to exist on an unearned and socially unnecessary surplus, and prophesying a dreary outlook for both labour and capital, which could be ameliorated, his argument shows, only by the free importation of foreign corn. "Capital and labour wanted a return to cheap food from the outrageously high prices of the war: the landlords wanted a retention of those high prices. Thus was precipitated the struggle of the Corn Laws. Probably the major reasons for Ricardo's pessimism were policy ones. He wrote his book with a definite purpose in mind, to halt the reimposition of the corn laws. Siding with capital and labour in favour of cheap food, Ricardo wanted a large importation of corn, and the future that would ensue without it could not be painted too black."[25]

Ricardo, we have said, laid great emphasis on population as the governor of economic development and progress of society. But changes in population did not form, for Ricardo, a "given" cause of economic change in the sense in which we have been using that phrase. Ricardo believed, on the contrary, that the economic situation determined the direction of population change and that population, in turn, governed and limited the development of welfare. Population changes were internal to the order of economic causation, and Ricardo developed a completely self-contained theory of economic change in the same sense as, according to Professor Schumpeter, did Karl Marx. When technical progress was achieved so that goods were more plentiful and real wages rose, population increased, and this increase, pressing on the fixed factor of production, land, resulted in diminishing returns and a decline in real wages to the old level of a bare subsistence. The subsistence level was inevitable and a return to it could only be temporarily postponed, in any progressive society, by the extension of the frontier and the subjection to cultivation of new lands of good quality. Ricardo rejected the Malthusian view of population increase as the "main generator of social progress".[26]

---

[25] Ruth Albert, op. cit., p. 19.

[26] J. J. Spengler, op. cit., p. 263.

Ricardo's crude attempt to explain population movements purely in terms of economics requires today no extended criticism. The experience of the past century has shown its faultiness, and modern population theory has gone far towards a completely non-economic explanation of population phenomena. Indeed, as we shall later observe, it is today more important to rediscover and emphasize the element of truth in Ricardo's thesis, to show the degree to which population movement is still influenced by economic circumstance, than to belabour the dead horse of Ricardo's original thesis.

Given, however, an increase in population, are we to accept Ricardo's conclusion that real welfare must decline? We shall show in Part III that, given certain assumptions that Ricardo made, this is a perfectly valid conclusion. Given certain other assumptions, such as, for example, Professor Hansen makes, the conclusion of declining welfare is invalid and incorrect. The assumptions determine the conclusions, and, from the practical point of view, the society which approximates the conditions Ricardo assumed will experience a decline of welfare with population growth and an increase in welfare with population decline, while countries with conditions approximating Professor Hansen's assumptions will experience the reverse set of relationships. A close study of different inferential models is necessary for proper conclusions here, and about all we may at present observe is that the general theory of economic change requires study of that type if it is ever to get beyond the purely speculative stage.

Ricardo, of course, was interested not only in the effects produced on total wealth by an increase (or diminution) of population, he also studied the effects on the distribution of income. The crux of Ricardo's theory was that increased population necessitated a resort to inferior lands, so that rents increased. The increase in rent (corn rent) was accentuated by the increase in the money price of food relative to other prices, so that the landlord class benefited doubly. Money wages, too, might increase, but never more than in proportion to the price of food, so that real wages could not rise above the subsistence level, except for very short periods. The rise in money wages,

however, resulted in a diminution of money profits to all non-rent-receiving entrepreneurs, and, in so far as they bought foods at higher prices, their real income was still further reduced. There is a strong suggestion in Ricardo's writing that he regarded this falling rate of profit as his most gloomy conclusion, because it spelled an eventual failure of capital supply and still further retrogressive movements of the economy. One gains the further impression that he never really believed this, but held it up as a Cassandra-like warning against the policies of the landlord interest. The falling rate of profit he anticipated was not founded on analysis of the economic system as such, but rather as an aspect of population movement and prevailing food policy.

Ricardo's whole system depends on the simplicity of the model of "real" economics which he uses. His economic world functions according to Say's law of markets. Production is limited by physical capacity, and all that is produced is consumed. If we know (as from the crude application of the law of diminishing returns Ricardo thinks we do) the effects of increased population on production, we then know its effects on consumption. The secondary effects on relative prices and the distributive shares can be inferred according to classical value theory, as we saw above. This "real world", however, is a very formal one, a most "unreal" abstraction, and among the important things from which it has been abstracted are the very essences of capitalism, the function of entrepreneurship, incentives to change, and the shifts in aggregate income and outlay that govern changes in investment. The orderly and automatic adjustments of Ricardo's market leave no function for the entrepreneur to perform. He has no problem of anticipation, no real risk to bear, no serious charge is put upon his judgment. There are no incentives to new investment, other than for replacement, once all is settled by "real" and automatic forces. The fluctuations in aggregate demand which afford incentives to alter investment which, in turn, gives a stimulus to innovation of one form or another is likewise omitted from the system.[27] The flow and flux, the very life of capitalism, is absent.

Indeed, the complete rejection by Ricardo of Malthus's emphasis on effective demand, and Ricardo's uncritical acceptance of Say's law, diverted economic inquiry from what would have been most fruitful channels. Lord Keynes, in his generous acknowledgements to Malthus, has shown how penetrating and important his thesis was. "Malthus lived", as one of his most recent commentators has written, "in what, for a section of the European sphere, was the latter part of a period of transition; in a period when the industrial revolution was still transforming the productive basis of society from one in which the instruments of industry and commerce would predominate; in a period when the comparatively unlimited demand of the masses was beginning to outstrip the limited demand of the opulent few; in a period when the feudal polity and economy was being dissolved and converted into a polity that was more democratic, and an economy that was more non-agricultural, urban and free-enterprise, in character. In this new scheme of things the common man was far more important than in the erstwhile pattern of society. Upon his efforts depended the output of goods and services, upon his purse depended the opportunity to sell, and upon his soldierly skill depended the national security; . . . Malthus . . . recognized the new and growing role of the common man and gave expression to it."[28]

Population growth must be studied, as Malthus had seen, not only as it affects real production, but also as it affects aggregate demand and investment opportunities. Further, population growth cannot be studied as though it had no effect on, and was unaffected by, innovation. The entrepreneur

[27] Ricardo does devote a chapter to the question of over-saving. In Ch. XXI he finally concludes that savings are never too great, and that their accumulation, taken by itself, does not lower the rate of profit. "From the account which has been given . . . , it will appear that no accumulation of capital will permanently lower profits unless there be some permanent cause for the rise of wages." He quotes Adam Smith to the effect that capital accumulation raises wages but lowers profits and dismisses this by referring to Say's demonstration which "has most satisfactorily shown that there is no amount of capital which may not be employed in a country, because a demand is only limited by production", *Principles of Political Economy* (J. M. Dent ed., London, undated), p. 192. It is surely fair to say, then, that Ricardo, unlike Smith, completely ignored the "aggregative" aspects of population change as of innovation.

[28] J. J. Spengler, op. cit., p. 262.

always has an incentive, even if demand be unchanged, to introduce an innovation because of the possibility of temporary monopoly profits which he may earn when he is first in the field. How much greater is this incentive when population and aggregate demand are increasing! Again, if population is diminishing, there is an effect of a qualitative sort on innovations. Entrepreneurs tend to seek labour-saving, cost-reducing inventions.

Indeed, Ricardo treated all innovations as consisting of the substitution of machinery for labour. He records that he "changed his mind" about the effects of machinery, and the text of the third (1821) edition betrays this change of view. In Chapter XXXI of the *Principles* he first sets out the orthodox view that the substitution of machinery for labour increases productivity and that, as Adam Smith had argued, all classes of society benefited from this improvement. "As then it appeared to me," he wrote, "that there would be the same demand for labour as before, and that wages would be no lower, I thought that the labouring class would, equally with the other classes, participate in the advantage, from the general cheapness of commodities arising from the use of machinery".[29]

However, these opinions he was forced to modify. "These were my opinions, and they remain unaltered, as far as regards the landlord and the capitalist: but I am convinced that the substitution of machinery for human labour is often very injurious to the interests of the class of labourers . . . I now see reason to be satisfied that the one fund, from which landlords and capitalists derive their revenue, may increase, while the other, that upon which the labouring class mainly depend, may diminish,[30] and therefore it follows, if I am right, that the same cause which may increase the net revenue of the country, may at the same time render the population redundant and deteriorate the condition of the labourer."[31] David Ricardo, as one of his editors says, was "frank in acknowledging error and in admitting conviction",[32] and here he was modifying his

[29] David Ricardo, op. cit., p. 264.

[30] Mill's wages-fund?

[31] David Ricardo, op. cit., p. 264.

views in the light of certain known and observable facts of his day, but his theoretic system was incapable of assimilating the new view. He was stumbling, so to speak, on the threshold of an extension of theory to the field of the business cycle and the aggregative fluctuations that govern employment, but he was not destined to enter. Instead, he adumbrates some sort of wages-fund, makes little of it, and relapses into a "bald and unconvincing statement" of his favourite thesis of "redundant population" pressing on the means of subsistence. Sheer reliance on Say's law could not have permitted this admission—greater productivity would lead to greater consumption. A full understanding of the nature of innovation, of the different kinds of innovation and their differing role in economic development would have resulted in a complete revision along Malthusian lines of the population thesis as the basic cause of change. Ricardo leaves, however, simply an explanation unintegrated with his general theory, of labour-saving machinery reducing the demand for labour[33] and so diminishing labour's real share of the final product. As we shall later see, some innovations are causally related to long-run upward swings, others are cyclical phenomena, but are effects rather than causes of the cycle movement.

Ricardo, like Smith, was more aware of, or at least more willing to discuss, political institutions and policies and their relation to economic development than a later generation of economists. He, too, saw the importance of social direction and purpose in determining the future course and welfare of economic society. Indeed, like Smith, he meant, as far as he could, to influence by his writings the determination of policy. He believed that proper taxation, tariff and commercial policy could offset for a temporary but not inconsiderable period, the disastrous results of increasing population. He believed that

[32] F. W. Kolthammer, "Introduction" to the J. M. Dent edition of the *Principles*, p. 12.

[33] Capital is "stored-up" labour. But the amount of labour including the stored-up labour of the machine, is less than the former quantity of labour required before the introduction of the machine to yield the same final product. In fact, however, this depends on the nature of the innovation, its adaptability, the effect it has on the rate of obsolescence of old equipment, the investment opportunities it creates. See Ch. VIII.

the discovery and employment of new and fertile land would bring wealth not only to the colonists but relief to the mature economies of the metropolitan powers. Not only could populations be directly relieved by emigration, but, by freedom of trade and the unrestricted operation of the law of comparative costs, the margin of cultivation in the old lands could be raised so that all lands, through trade, could share in the abundance of the new production. Again, however, it is interesting to observe that Ricardo regarded this extension of the frontier not as a stimulus to production and the provision of investment opportunities, but primarily as a relief to the pressure of population on existing tillage, as an offset to the "real" effects of diminishing returns.

KARL MARX

Karl Marx is unique among the economists between Adam Smith and Professor Schumpeter in that his main design is to create a theory of economic development and social change rather than a static analysis of the market. Marx was in the tradition of the great religious prophets, perceiving the ills of contemporary society and foretelling the disaster and doom that would be the consequences if man did not mend his evil ways. It is this sombre touch of Hebraic prophecy which, added to the mixture of Hegelian philosophy and English classical economics, gives Marx his impressive stature. The prophet is always concerned with change, its causes, manner and mode. Thus, for Marx, change and development were the very essence of modern capitalism, the significant features which required understanding if man was to achieve mastery over his fate. The bastardized Hegelianism of Marx's philosophy of history somewhat clouds his genuine conviction, expressed in the revolutionary periods of the Manifesto, that man, by taking thought, can control his political and economic future. (Engels, later, made clear that he and Marx had never intended that dialectical materialism should be misunderstood to be a rigid economic determinism, and excused an over-emphasis, which might lead to this misconception, on the grounds that it was natural in pressing the merits of a new and controversial idea.)

Marx's system of political economy, however, though directed to the analysis of change in the capitalist system, is based on a highly abstract and more or less static sub-structure. Mr. Paul Sweezy has expressed the opinion that Volume I of *Das Kapital* was intended as a first approximation, involving a high degree of abstraction. This is close to, though not identical with, our idea that it is a kind of static analysis, designed to throw up the set of relationships obtaining in the capitalist economy which generate change, and which in the later volumes Marx studies as part of an historical process. For Marx, as Professor Schumpeter has observed, the causes of change lie within the economic system itself and are not external to it. Thus, though Marx studies the development of the economy from certain "given" causes, they are "given" only for that portion of his work. They emerge from within the economy and their analysis is part of the proper field of economic inquiry. They are certainly not "given" in the same sense as innovation and capital accumulation for Smith, or even population for Ricardo.

Marx's method is, however, open to an objection. He included capital accumulation in his first model, an inclusion possible for Professor Schumpeter because he treats one form of accumulation as compatible with continuing equilibrium ("stationary flow"), but not for Marx, who regards accumulation (from surplus value) as a basic cause of disequilibrium and change. It is therefore improper to include it as an element of a static and equilibrium system. The theory of value Marx achieves in his first model is "classical" in the sense that the labour theory of value is a form of the real cost theory. This emphasis on production and neglect of demand is carried over into his theory of development, which, as we have suggested, is based on his theory of value.

Essentially, Marx believed that everything that went into production "came out" in the final product, and nothing more. The form of the goods was changed, but there was no value surplus added. That which went into production was of two categories, labour power and fixed capital, and the latter was resolvable, from the point of view of value, into the labour that constituted it. Marx emphasized that fixed capital could not

change in value in production, and consequently could create no surplus in production. (In modern theory this line of thought might be expressed by saying that if the productivity of capital exceeds its price, the price will be bid up. The marginal productivity determines the price, and surpluses are earned only if there are imperfections in the market.)

Marx also held that labour did not create a surplus—the value of the labourer was equated with the value of what he produced. What is called "surplus value" was a surplus not between the value of labour and the value of labour's product, but between the latter and the wage the labourer received when he sold his labouring power at a price equal only to his subsistence. Thus "surplus value" for Marx is the difference between the value of the worker's subsistence and the true value of his labour power. It follows, therefore, that all increases from technical advance, the development of science, the improvement of skills and artisanship, which increase the worker's productivity, increase this surplus. Marx's notion of surplus value, then, depends on his acceptance of Ricardo's subsistence theory of wages, or something rather like it, and his own form of the labour theory of value in which he denies the possibility of an increase in value from the change of form which takes place in the productive process. We shall later (Chapter IX) have occasion to observe what we believe to be the true nature of surpluses in the productive process and the manner of their arising. At this juncture, however, we may agree with Marx that there are no true "surpluses", i.e., differentials between total costs, including opportunity costs, and total value produced, in a static and competitive model.

There remains then, as an essential condition of the notion of "surplus value", the subsistence theory of wages. Most of the earlier critics of Marx have dismissed his argument on this point with scant consideration. Their attitude is pretty well represented by Professor Davidson's blunt dilemma,[34] which may be summarized as follows: either the subsistence level is precisely defined as the exact quantity of nutrition necessary

[34] See John Davidson's *Bargain Theory of Wages* (London, 1898).

to maintain a given level of labour, in which case this quantity cannot be held to be the determinant of wages because it is both empirically false and theoretically unsound in its neglect of demand, or the subsistence level is an indefinite quantity necessary to some "standard of living", a vague, changing quantity, a useful notion to juggle with in political pamphleteering, but meaningless as a basis for scientific argument, because it has no definite meaning. Any theory based on such a notion has to explain how standards of living change, and hence has to admit that the indirect demand for labour plays a part in governing its price.

Later commentators on Marx, however, have attempted to discover in his subsistence theory an adumbration of later wage theories, which, properly understood, justify his position. Mr. Sweezy,[35] for example, argues that Marx substituted for Ricardo's crude population theory a concept of a "reserve of labour", and argues from this that the reserve of unemployed would compete with the employed to bring wages down to the subsistence level. This, however, not only offers no escape from Professor Davidson's dilemma, but also, if it is a true reading of Marx's view, is an improper explanation to offer within the logical confines of a static model, because Marx definitely explains the existence of unemployment as a dynamic phenomenon. In a static economy there would be no unemployment for Marx. Therefore, unemployment cannot be the explanation for subsistence wages, which, Marx holds, obtain under static conditions.

Mrs. Robinson has argued[36] that Marx bases his theory of exploitation on empirical evidence of unequal bargaining power and that his "surplus value" corresponds to the "monopsonist's surplus" of modern theory. This may well be fair and sympathetic reading of Marx's position, but it amounts to no more than saying that he had caught a glimmering of the truth, had correctly observed the empirical evidence, but had failed to explain it analytically. There is, under monopsonistic conditions, a differential between what labour produces and what

[35] Paul Sweezy, *Theory of Capitalist Development* (New York, 1942).
[36] Joan Robinson, *Essay on Marxian Economics* (London, 1942).

labour receives as wages. It is something to have perceived this, but to have explained it in terms of subsistence wage theory is not to have explained it correctly.

The distinction we draw here between the true notion of an exploited monopsonist's surplus and Marx's surplus value is an important one, because, for Marx, surplus value is the basis of profit, and the rate of profit, in turn, a dynamic force, an operative cause in his theory of change. In our system, too, as we shall show (see Chapter IX) the existence of surpluses is an important factor in economic change, but a true statement of the origin of surpluses shows them to arise from the exploitation of all factors, not of labour alone. The rate of profit for Marx is expressed as the ratio of surplus value to the capital, "constant" and "variable", employed. Thus if P^1 is the rate of profit, S is surplus value, C, constant and V variable capital,

$$P^1 = \frac{S}{C+V} \qquad ^{37}$$

In the Marxist system, it will be remembered, S is created only by V.

It is necessary here to notice Marx's distinction between consumers' goods industries and capital goods industries. Between these two sectors, or departments, of the economy an equilibrium is maintained. If we suppose Department I of the economy to be made up of the producers' goods industries, and Department II of the consumers' goods industries, and assume that throughout the economy surplus value (S) is equal to variable capital (V), then according to Marx we get the following circulation of commodities:

Department I

(in commodity units) $4{,}000C+1{,}000V+1{,}000S=6{,}000$

[37] This formulation, which makes profit depend on the capital structure of an industry, is obviously incompatible with Marx's concept of an "average" (in the sense of normal) rate of profit. In Vol. III, Marx explains his average as an arithmetic average. If he means to imply, however, that actual profits tend to approximate this average, he must be relying on some notion of competition and of mobility of capital, foreign to this present argument. On this point Mrs. Albert (op. cit.) has observed that technical determination of the combination of the factors has made any general theory of profits or wages problematic.

Department II

(in commodity units) $2{,}000C + 500V + 500S = 3{,}000$

In "simple reproduction", where surplus value is distributed to the capitalist as revenue rather than used to increase constant capital, the quantity of capital goods (4,000 units) needed for the continuation (at the same rate) of the productive process will be retained in Department I, from which end products (2,000 units) will be exchanged for 2,000 units of consumers' goods for the consumption of its workers and capitalists. Department II will have 1,000 units of consumers' goods left for the support of its workers and capitalists, and would have received 2,000 units of producers' goods to maintain the flow of production. In this "stationary flow" of simple reproduction an equilibrium is maintained between the two departments of the capitalist economy.

There is, of course, no conceivable reason for this equilibrium to be reached. Marx sets it up as an illustration, and then seems to treat it as a basis of analysis, ignoring the plain fact that the more producers' goods that pour from the production line, the more consumers' goods will eventually be produced. He goes on to argue that you can set up similar equations of exchange with capital accumulations, i.e., when the production of producers' goods increases, and the production of consumers' goods will fall off. Certainly such a relation between the two departments could never be explained purely from the production side, and Marx, no precursor of Lord Keynes or Professor Hansen, ignored the possibility of under-consumption.

Marx's theory of capitalist development, or of capitalist decline, we should say, involves these two concepts of the rate of profit and of capital reproduction (or accumulation). The dynamic driving force, in the Marxist system, is, however, innovation. The profit motive is the incentive to innovation, because, whatever the long-run consequences, in the short run innovations either promise profits to the foremost or are necessary to forestall failure for the laggards. As individuals, entrepreneurs must innovate or disappear. It is the social conjuncture of these individual acts that, Marx believes, leads to disaster. What profits—temporary only—result from inno-

vation Marx calls "super-profits", to distinguish them from the true profits residing in surplus value. These super-profits can never offset the tendency of profits to decline; they are temporary only, incidental in appearance and do not cause any increase in the source of true profits. Why, according to Marx, true profits must fall we shall shortly see.

For Marx the distinguishing aspect of an innovation was that it was cost-reducing. When we remember his theory of costs, this means that, like Ricardo, he saw innovation as the reduction in the amount of labour necessary to produce a unit of a commodity. It was the substitution of machinery for labour. This is, of course, a limited concept of innovation. It means, always, an increase in the ratio of constant capital to labour. That it meant, also, a reduction in values is a point Marx was not inclined to emphasize, an omission not to be held entirely to his scientific discredit—though it was no doubt convenient to his thesis to neglect this aspect of innovation—,because it was common to the classical school to be interested, not in relative values, but in physical riches. Ricardo thought it sufficiently odd to devote a chapter to show that values could decline while riches increased.

Since technical advance meant an increase in the ratio of constant capital to labour, Marx speaks of it as capitalist accumulation, and thus avoids the favourable "over-tones" that the phrase technical advance or innovation might bear. Now the theorem of a falling rate of profit (disregarding the temporary "super-profits") follows directly from the increase in the proportion of constant capital in the productive combination. Since S (surplus value) increases only with V (variable capital), and since

$$P^1 = \frac{S}{V+C},$$

if C is increased, while V and S remain constant, P^1 must decline.

This simple algebraic relationship, however, defies Marx's own recognition that innovation (which increases the proportion of fixed capital) also increases the productivity of labour.

I cannot see how this increase can be expressed in Marxist terms except as an increase in surplus value. In that case S is not a function only of V, but is a function of C also, with a positive sign. If this be admitted, as I think it must, when C increases, S also increases, and unless Marx has some way of showing that the rate of increase of S is less than the rate of increase of C, it is impossible to predict from the conditions here given whether P^1 increases, diminishes or remains unchanged. Marx may intend that the direct relationship between S and V, which also obtains, limits the increase in S because V is not increased, but without definite knowledge of the *rates of change*, it does not follow from this that S will increase less in proportion than C.

This is to say nothing of the empirical evidence of developments excluded from Marx's system. What about the possibility of innovations occurring so steadily that "super-profits" become, in fact, a part of expected "average profits", offsetting and more than offsetting the tendency of the latter to fall? What will be the effect of institutional changes—whose relationship to the economic development Marx strangely neglects in his theoretic system—the development of trades unions restricting exploitation of labour, the corruption of competition, making possible the retention of super-profits? A theory of economic change, on which such broad conclusions about policy as Marx's were to be based, should surely not neglect these important questions. Marx's argument, indeed, leaves it quite indefinite as to whether a falling rate of profit can or cannot be expected. Marx's empirical evidence does not help his case, because it has to do with the rate of interest, which he confuses with the rate of profit.

Changes, from innovation, in the general capital structure of industry, cause changes in the proportion of capital goods' industries to consumers' goods industries. There is a general "deepening" of capital, which, I think, would always follow from innovation as Marx understood it, and this deepening would be accompanied by the creation of "super-profits". Their existence would then upset the equilibrium between the two departments of the economy, and one might well expect

dislocation effects, but surely not a long-run or secular decline. It may be that, had Marx here looked to the effects on aggregate demand he might have developed a theory of the business cycle, that would fit in with his secular theory, but Marx is contemptuous of demand or "utility" analysis. Arrogance exacts its own price.

Marx was not, of course, interested in the decline in the rate of profits, *per se*, but in the effects of that decline on the capitalistic system of society, which, depending for its dynamic on profits, must, he felt, decline with profits. The capitalists, once aware of declining profits, would suffer panic. The psychology of depression would prevail, a crisis ensue. New investment would cease with consequent unemployment in the producers' goods sector of the economy, and a "slaughter" of capital values would be carried through on the stock market and in the bankruptcy courts, whereby capitalists would attempt to return to a low constant capital valuation in their capital structure, and so increase the proportion of surplus value. As soon, however, as production started again, competition would follow the increase in constant capital and once again profits would decline. This cycle, Marx believed, wiped out the super-profits of innovation and thus left the secular downward trend to prevail. The cycle would become sharper in contour, the crises more acute, the long-run downward trend more pronounced until, the system in bankruptcy, it could be overthrown by revolution. This is, indeed a theory of "secular stagnation", based, however, not on under-consumption, but on the notion of surplus value, the inevitable decline in the rate of profit, and the cyclical destruction of super-profits.

Marx evolved, therefore, a completely self-contained theory of economic development, a theory in which the causes of change are found within the economic system itself. It was necessary to his whole social philosophy that he should do so, because it is the movement of the economic system that he makes the dynamic force of all social change. Thus the causes of economic change must be endogenous. It is a bold effort to establish a comprehensive and self-contained economics, but its usefulness for prognosis is vitiated, as we have seen, by faulty

analysis and by the unscientific disregard of all considerations which might not lead to preconceived and desired conclusions.

J. A. SCHUMPETER

The "theoretic blight"[38] which began in Vienna and thence infested the greater part of systematic inquiry resulted in a sharp curtailment of the field of analysis. That portion of the field which we call economic change was left untilled except for the surface-scratching husbandry of the historical school. In our own day, a return to the interest of the classicists, started by a rebellion in Cambridge, has steadily gained adherents, and this movement, as widened by the writings of Professor Hansen, has come round to the study of long-run or secular change as a result of its earlier preoccupation with short-run fluctuations and disequilibrium. Earlier, however, and from the very centre of the infestation itself, Professor Schumpeter had moved from pure equilibrium analysis to a clear distinction between static and dynamic inquiry, a definition of the endogenous factors in the latter, and a comprehensive theory of development based on a given cause.

Postulating a "stationary equilibrium", Professor Schumpeter adopts the method of examining the causes that generate a departure from this equilibrium over time. He is not, therefore, in the tradition either of Smith or Karl Marx. Smith has no precise and vigorous norm (his "natural market" is far too vague and variable a concept), his description of capitalist economic society, rich and full as it is, is too much description and too little analysis. For all that, his approach always has the merit of attacking directly the development of the society in terms of wealth or welfare. Marx, too, believed change and development essentials of economic society and directed his system to the analysis of the principles of change in a capitalist society and their welfare and social effects, though, as we have seen, he did make use of a kind of static equilibrium concept.

[38] "Theoretic blight" is the phrase used by Professor E. R. Walker, *From Economic Theory to Policy* (Chicago, 1943), Ch. IV, to designate "The development of economics under the impulse of theoretic construction, for its own sake, along paths which become ever more remote from the real world" (p. 57).

It was, with him, however, an analytical tool towards welfare study.

With Professor Schumpeter we ascend to a more rarefied intellectual atmosphere. For him neither his static nor dynamic models are "realistic". They are all rigorously defined and maintained abstractions, methods of studying how a particular given cause of change, when introduced, affects the internal economic relationships. There may be welfare implications of such changes, but Professor Schumpeter is not—at least on this universe of discourse—interested in these. His primary concern is the extension of pure analysis to certain selected problems of development, and therein lies his contribution.

Professor Schumpeter's "equilibrium" is a modification, from which profits and interest are removed, of the Walrasian general equilibrium; values are created by demand, and costs represent foregone utilities, i.e., they are opportunity costs. The system is perfectly competitive and there are no indivisibilities or other frictions. The valuation of the factors is based on marginal productivity imputation, though, as we shall later see, some emphasis is directed towards production problems. Changes in the nature of the productive agents become important in a theory of change, so they must be allowed some independence in the stationary system. Capital, it is interesting to note, is resolved (Petty-like) into land and labour. There are no surpluses when all costs are opportunity costs, hence no profits, and interest also disappears because of perfect knowledge of all future markets and perfect adjustments immediately achieved.

To this equilibrium, "stationary flow", Professor Schumpeter introduces a single cause of change to see how economic society reacts and develops. He is not interested in how the cause comes to exist, and would not admit that it could come from within the model itself. This is perfectly proper, of course, because of the nature of the model: what would be improper would be to suppose that the cause would not arise within a given, real economic society, but must always be external to it. Sometimes, it seems to me, Professor Schumpeter verges dangerously close to this confusion between his model and real

society, and comes close to seeing causes of change as always external to the economic system.[39]

The single cause on which Professor Schumpeter concentrates and which he explicitly defines is technical advance. He defines an innovation as an economic, not a physical fact, the introduction into economic use of an invention, and his definition is sufficiently wide to include new methods of producing old goods, new goods, new materials and new sources of materials, new uses for old materials, new skills and techniques, including techniques of organization, administration, exchange, merchandising and financing.

The incentive to innovation is the possibility of profit, which cannot be earned in the stationary society. For Professor Schumpeter profit is tied by definition to innovation. It is the extra gain an entrepreneur can acquire by lowering his costs below his competitors. This is the function of entrepreneurship. But, in the model, costs can only be lowered below the competitive level by the introduction of some productive technique previously unused, in other words, by an innovation. Thus, for Professor Schumpeter, profits are a dynamic phenomenon only. They are possible only in a dynamic society, and this possibility is the incentive to development. This emphasis, we might note, implies that Professor Schumpeter, in spite of his recognition that an innovation may take the form of the introduction of a new good, in fact always thinks of it as the introduction of new—and less costly ways—of making old goods.[40] We feel that more attention might well go to the different kinds of innovations and the differences in their effects on economic change.

Professor Schumpeter believes that cost-reducing innovations are progressive in the sense that they enable society to produce more goods at less expense of effort. He relates the long-run process to the cycle through the way innovations are financed and introduced. He views the long-run process as consisting of an inflationary, prosperous movement when new innovations

[39] I am thinking here of the *Theory of Economic Development*, not of *Business Cycles*.

[40] This is not a serious neglect. An innovation in new goods also follows the profit incentive. See Ch. VIII.

are being financed by credit creation, followed by a movement back towards "equilibrium", but an equilibrium at a higher level of real welfare. His general attitude is optimistic, and he tends to under-state the seriousness of depression, and to remain comfortably detached from the miseries and sufferings of the masses which periodically accompany the "progress" of society.

The more recent writings of Professor Alvin Hansen, with a pessimistic tone resembling in some ways that of Karl Marx, require of us a critical approach to Professor Schumpeter's system. Professor Hansen believes that there are three given and external causes of economic change, which act more or less independently.[41] Professor Schumpeter does not explicitly deny these other causes. On the contrary, he insists on the complexity of the causes of change. But he does select a model for analysis in which only one cause is allowed to operate. We do not intend to be understood that this is not valuable analysis, and a perfectly proper use of limitation and selection. It is, however, distinctly limiting, and precludes the possibility of a comprehensive theory of economic change. Professor Schumpeter's general optimism would be dimmed, perhaps, by a study of other models, in which other causes present in the real world, were allowed to operate.

Another limitation of this method of a single model is that one is unable to discern with any particularity the relation between causes, the effects of innovations on population and the extension of the frontier. One is left with the feeling that if other causes are operative, they are so in a quite independent way. The inter-relationship of causes of change is omitted.

More important, in our opinion, are the limitations imposed by the selection of Walrasian equilibrium as the starting-point of the inquiry. Professor Schumpeter, alone among the writers we have been examining, is profoundly uninterested in policy and its relation to change. He has a keen interest in the financial system, the relation of its development and its institu-

[41] We have not included an account of Professor Hansen's theory in this chapter. It proved more convenient, and, I trust, will be more acceptable, to examine Professor Hansen's thesis at a later stage of this work.

tions to economic change, and he has condescended finally to write a popular book with characteristic charm of manner in which he discusses political institutions in relation to the economic system, but he does not in his systematic writing develop any theory of the relation of policy to the dynamic process of the economy. His selection of a general, fully competitive equilibrium model precludes the important study of the effects of change on the market, the effects of innovation on the degree of competition. His concept of profits neglects the possibility of monopolistic surpluses, their effects, their potential discouragement of innovations, and the effects on aggregate demand of such surpluses.

Again the Walrasian-like emphasis on general equilibrium leads to a neglect of the whole theory of the firm, and the theory of the process of the firm, a section of analysis recently worked through, and full of implications and suggestions for the theorist of economic change. Professor Schumpeter, exploring a new field, limited himself to a section of it, indicating a method of analysis capable of extension to other sectors, but leaving unattempted the problems of the inter-relationship of causes, models in which causes other than innovations operate, the relationship of firm analysis to general analysis, and, since the development of Keynesian theory, the relationship of what has been called "macro" or "aggregative" economics to "micro" or "real" economics.

CHAPTER V

THE GENERAL THEORY OF ECONOMIC CHANGE

From what has already been said it is now apparent that a general theory of economic change must relate the various major causes of change and explore their inter-relationships; it must offer a coherent account of how these changes work themselves out through the economic system and of their effects on welfare; it should examine the extent to which the given causes are within the order of economic causation and the manner and direction of their reaction to the process which they originate; and, finally, it should include the institutional channels through which the process operates, its effect on these institutions and the possibility and method of social direction of the development.

All the writers we have examined have admitted in one form or another the existence of a complex set of causes of economic change. Professor Schumpeter is the most explicit on this point, insisting that the economic system is operated on by the entirety of the social order, its events and systems of ideas and by the physical universe with its unpredictable happenings, such as the earthquake of Tokyo. From this totality one can only seize upon such causal events as seem of prime importance and susceptible of generalization. In that way one distorts reality; there is always the possibility that other and extraneous causes will act to modify or even reverse the probable direction of movement, but, at least, one gets an intelligible theory of some degree of probability. With this we agree. The real question is the selection of these causes of prime importance. The history of thought teaches us that different writers at different times have made different selections, and, in each case, at the time the selection was made it appeared reasonable and justified. Though no theory of economic change can be constructed, and remain a general theory, which takes account of all the causal

factors, an historical appreciation may teach us that at least a group of prime causes must be admitted, a group of which now one, now another is predominant in the economy. I am inclined to believe that that is the first lesson of our brief excursion into the history of economic thought, and that we must consequently include in our group of prime given causes at least innovation and population change. Geographic discovery and the extension of the frontier are also important, but I believe that by suitable definition they may be included under the term innovation.

Further, implicit in the work of Karl Marx and Professor Schumpeter and explicit in the writings of Adam Smith, David Ricardo, Lord Keynes and Professor Hansen, is the idea that human purpose, as expressed through social institutions and policy, forms an element, and a determining element, of the complex cause of economic change, and one which must be taken into account in any theory of change. Since, whatever may also be the degree of pure scientific curiosity in the motive for economic inquiry, it is of all the sciences most distinguished by the intention of its professors to inform social policy, the method of most writers is to show the operation of other prime causes, if left to themselves, and then to show how policy can guide the process so as to minimize socially undesirable results, and to maximize socially desirable ones. Thus, we have seen Smith argue that the progress he believes in can be most readily achieved in an economic system free from artificial restraints on competition and free commerce, and Professor Hansen believes that a suitable fiscal policy can prevent or postpone the incidence of the stagnation which, in the absence of such policies, he believes inevitable. Policy, in a word, is not included in the original causal pattern, but is led up to, and, the economic process having been set out on the assumption of a neutral policy, the writing becomes normative and the appropriate policy delineated. We believe this method is partly, but not entirely, justified. We believe that policy is conditioned by the socio-economic institutions, and that these institutions themselves affect and are affected by the process of change. For example, Professor Schumpeter's system of perfectly com-

petitive equilibrium is affected by innovation, and becomes less competitive. The choice of a perfectly competitive system also determines in part the effect on the system of innovation. Thus we must insist on the recognition of a particular kind of economic institutions (especially with respect to the degree of competition in the market) as a part of the causal process. We believe, of course, that once an area of decision has been defined, appropriate policy may determine the course of events, and thus reserve a place—and an important one—for normative writing. The policies deemed appropriate, however, must be considered in the light of their applicability under the institutional framework under which and on which the causal process is operating. I should like this to be very clear. Professor Hansen, and the others, are saying, "Here are certain causes—population change, innovation, and others—which operate on the economic system in a certain way, and here are their effects. If you don't like these effects, you'd better watch out, and do something about it." Usually, they follow this by indicating what ought, in their opinion, to be done. The point I am trying to make is somewhat different. I am trying to say, "Here are certain causes,—population change, innovations and so forth. They operate in different ways, according to the institutions of society. They also operate on these institutions and change them. Because they do that, people's decisions as to what ought to be done about it—decisions which are a part of the causal process—become affected. You are not as free as you think you are to select policy. We must examine the institutional effects of economic change in order to determine how free we are to act, and what policies we really can adopt effectively, to modify this causal process." I do not wish to exaggerate this difference, but I think it is a logically important distinction.

Once we admit this complex nature of the causes of economic change, we have to deal with the relations between these causes. Failure to do this is the reason for the strangely divergent views of the different authors we have examined. Smith believes that the introduction of machinery is at first beneficial to all classes, but in the long run leads to a declining rate of

profit and an improvement in the conditions of labour. Ricardo, however, believes that the introduction of machinery creates a "redundant" population and impairs the condition of the labouring poor. Karl Marx thinks it leads to no real improvement in the lot of the proletariat, but does result in a declining rate of profit and the eventual ruin of the capitalist class, while Professor Schumpeter believes machinery innovations are the source of profit, the salvation of capitalism and the cause of a gradual improvement in the standard of living. If we are not to write this off as another example of the perpetual disagreements among economists, of the scientific bankruptcy of the subject and its professors, we must seek for the explanation in the differing models set up, in the different assumptions made, and in the consequent different causal processes which are at work. Adam Smith obtains his results by concentrating on the real productivity of labour and by ignoring the demand effects of increased wages. Ricardo obtains his by his preoccupation with population change as a basic cause and by interpreting the effects of machinery through its effect on the labouring population and the margin of land cultivation. Professor Schumpeter interprets the results of the change on a purely competitive system, ignores the possible stagnating effects of certain types of innovation during a period of population decline and the evil possibilities of under-consumption. Karl Marx ignores the effects on the productivity of labour and on real wages, the very point Adam Smith stresses, and the possibilities of enhanced demand. Similarly, as we shall show in detail (Chapter VII) Ricardo obtains his conclusion that an increasing population is a cause of decline in welfare, while Professor Hansen shows it to be a stimulus to progress, by setting up quite different assumptions and by a quite different treatment of the relation of other causes.

What is needed to clear up this kind of confusion is a comprehensive theory of the relation of different causes and the different modes through which they operate, a relationship between effects on real productivity and on aggregate income and demand.

The causes themselves are not independent either of one another or of reaction from the processes they initiate. We

shall show, for example, that one sort of innovation affects the degree of competition in the market, and that a monopolistic economy tends to direct innovations into different channels from those into which they are directed in a competitive society. The kind of research that is promoted, the direction of research, invention and the adoption of invention, are all affected by economic conditions. Similarly, geographic expansion is a function of economic dynamics, and the development and the nature of exploitation of new markets and new sources of raw materials depend on the economic institutions and policies of the metropolitan power. Again, population change is far more related to economic circumstance than some population theorists, in reaction against crude Malthusian and Ricardian doctrine, have admitted. Shifts of population within countries, and, in a more limited way, between countries, still occur in response to changing economic circumstance. Moreover, it is possible, we believe, that the rate of population growth is not unaffected by changes in economic conditions. Population growth has also been affected by innovation, and changes in the structure of population, with respect to both age and sex, affect the nature of innovations.

A theory of economic change which does not inquire systematically into these associations must remain speculative. Different approaches are no doubt possible to this problem of analysis. It is to be hoped that they will be made, by statisticians and econometricians, by historical economists, and, in different ways, by analytical economists. In the present work, we shall experiment with inferential models, in which different processes may be observed and related, processes as they affect the economy as a whole, and as they affect the single firm.

The integrating factor in economic change we regard as institutional, using the word to comprise policy and the formation of policy, as well as existing and functioning social machinery. Out of the totality of the physical environment and the social structure which, as Professor Schumpeter observes, constitutes the full and complex cause of economic change, certain "prime" and "given" causes may be selected for investigation. These causes operate on the economy and are

reacted upon by the effects they cause. A point we have already observed, however, is that for any single cause several alternative effects may be predicted, with more or less equal validity from the analytical point of view, depending on the assumptions made as to the nature of the institutions which are the channel through which the process flows. This is simply a way of saying that in the real world the effects any "given cause" may produce, depend upon the institutions and policies of the society. Thus no single "given cause" of the kind we have been discussing can positively be said to be attended by a single set of effects. Indeed, these selected given causes are not complete causes at all. They are, as we have earlier argued, only the objective socio-physical element in social causation. The purposeful or meaningful element of cause must be added before we have a true causal situation and can infer a set of effects. A social event, containing all these elements, must be seen completely to be intelligible.

If I have properly understood and stated both this complex nature of economic causation and the manner in which neglect of the institutional element leads to contradictory conclusions following upon some one "given" cause, it follows that the proper method to see the economic event (the total cause-effect situation) wholly is to define the institutional assumptions and to observe the event under alternative institutional conditions. The determining effect of institutions may then be estimated, the modification of institutions under the impact of the given cause discerned, and the area, in which free decision is possible, defined. The possible structure of policy and the ends that may be attained may then be intelligently—though we must not say certainly—predicted.

PART III

MODELS OF THE GENERAL ECONOMY

CHAPTER VI

INTRODUCTION TO PART THREE

The general theory of economic change cannot exist as more than a sort of speculative skeleton, we have seen, until we know more specifically how particular or "given" causes interact with market motives and are transmitted through the complex institutional system of economic society. It is something of a problem to determine how particular changes can most effectively be studied.[1] The method which we have finally adopted is to adapt the usual technique of economic inquiry to our kind of problem. We shall select some of the more significant causes of change and, regarding them as "given causes",[2] shall attempt to study them in a series of temporal models, in which some approximation to real conditions may ultimately be achieved.

A note of warning must here be offered. It is perfectly clear that in the usual static models of economic theory, once the simplifying assumptions and the general axioms are given, the conditions of equilibrium are strictly determined. The task of the student is a purely logical one. It is equally clear that an economic historian studying a particular process is, if he takes reasonable care, assured of the order of events, though his postulate of cause must remain hypothetical. Now the models we propose to set up are not historical, yet the relations which we shall infer within the models are not strictly determined and cannot be rigorously demonstrated. In effect they are only strong probabilities based on a common-sense estimate of the way things would happen in such a temporal setting. The argument is thus on an empirical and, in a sense, speculative level. Many may object to this. The only defence we can offer for this method is that, in the first place, the simplicity

[1] See the introductory section of the Appendix to Ch. IX on the representation of the demand function in time.

[2] In the sense defined in Ch. IV.

of the models gives the inferences a fair degree of probability and, in the second place, that alternative methods proved, on experiment, unfruitful. Pure history is informative and can yield evidence, as we shall note later, that supports the general inferences of the models, but, as we have insisted, it cannot, of itself, yield generalizations. My own attempts to set up temporal models that would permit demonstration of a more rigorous sort produced a most arid formalism. If the formal demand and production functions in time could have been measured, it is possible that they might have served to tell us something concrete about the nature of economic change, but the problems of measurement are enormous and the work already attempted in this field shows that further progress will be expensive and slow and probably disappointing from the analytical point of view. (The Appendix to Chapter IX of this present part will attempt to summarize some of this work for the non-mathematical reader who might have difficulty in dealing with the originals, and at the same time indicate the theoretical significance and theoretical limitations of this method of approach.)

The method we have ourselves finally chosen has at least one real merit. Since it consists in setting up models of change on the basis of certain assumptions and of inferring what will then happen, it connects certain conclusions about the effects of change with the assumed conditions. It enables us, in consequence, to lay bare the implicit assumptions of various economists who have written on economic change. We are able to decide under what conditions their conclusions would be valid, and under what conditions they would not. It may possibly permit us to venture on a further step. We may be able to say that certain effects from a given cause would flow only under certain conditions (only given certain assumptions). Thus, if these effects are desired, the conditions necessary to them must be set up. It is along this line that our method might conceivably yield suggestions as to policy.

There is, further, the question, as there must always be in any inquiry involving a large number of variables, of selection. What temporal processes shall we select to observe and in what

categories of effects shall we be interested? To such questions our answers must in part be arbitrary. From the argument of this work to the present juncture, it is apparent that we are primarily interested in welfare. This directs our attention to the welfare effects of any process of economic change,[3] and our interest in the effects on equilibrium is incidental and is involved only to the extent that the matter of shifting equilibria implies progress or retrogression in welfare. Since the welfare progress of society, according to our argument, depends, at least in part, on its institutions, we shall have to observe the institutional effects of any process of change, and, eventually, the institutional reactions on the order of change itself. We shall, therefore, select models at first of comparative simplicity[4] in which the effects on welfare and equilibrium of a single process of change can be distinctly observed. In later, and more complicated models, we shall try to combine the processes of change in certain "natural associations" and infer the probable welfare and equilibrium effects and the ultimate institutional direction of the process of change itself.

Any process of change which we may select can best be observed as it works itself out through two forms or modes. These we shall call the "aggregative mode" and the "real mode".[5] By the "aggregative mode" we mean the form or mode through which any change works its effects on the economy via its effect on aggregate income. By the "real mode" we mean the form or mode through which the change affects the economy via the alteration in the margin of substitution

---

[3] This distinguishes the present inquiry from such studies as J. R. Hicks' *Value and Capital* (Oxford, 1939), in which economic dynamics is studied as a problem of adjustment of equilibrium.

[4] The simplicity, as always in economics, implies a level of abstraction remote from reality.

[5] I had originally attached different names to these "modes", calling them the "income mode" and the "substitution mode", but, as I feared the possible confusion with the "income effect" and the "substitution effect" distinguished by Professor Hicks (*Value and Capital*, pp. 27-33), I have now decided that the distinction I am drawing is of greater generality and less precision than his, and the use to which it here is put is substantially different from the use he makes of his distinction in *Value and Capital*. I have consequently selected the nomenclature used in the text. The word "mode" is used in its Spinozan sense, as meaning the formal channel through which a process operates.

of one good, or group of goods, or one factor, for another good, group of goods or factor, the structure of the market, the level of real income and welfare, and the real rates of reward. Thus an increase in population may increase aggregate demand and in so doing work certain changes in the economy. This would be the "income mode". The same population change would affect the labour-capital combination, under certain conditions increasing the proportion of labour to capital, and this would work further changes, not necessarily in the same direction, in the economy. This would be the "real mode". It is important, we shall show, to sort out these two modes[6] through which effects from causes of change work themselves out, and, as far as possible, estimate the probable predominance of the one or the other when they appear to operate in opposite directions. This can only be done, if at all, by careful selection of the processes to be analysed and definition of the level of abstraction.

Since we are interested primarily in welfare effects and hold the view that the institutional pattern is itself a determinant of the way a process of change proceeds and, in turn, is affected by the process, we cannot follow Professor Hicks in assuming perfect competition. On the contrary we must observe throughout our models the difference which the degree of trustification of an economy makes in the working out of a process of change and, in turn, the effect of the process on the degree of trustification. This implies, if the argument of Chapter III is recollected,

[6] It seems to me that a good deal of unnecessary controversy flows from the failure to see these two modes of change and to allow for both. Thus Professor Pigou, moving from "pure" theory to the problem of employment, is concerned with the "real mode". If the aggregate mode is neglected it follows that a reduction of wages, if adequate, will alter the margin of substitution in favour of labour, and employment will improve. Lord Keynes and others, on the other hand, are almost solely concerned with the "aggregate mode" and from their premises it follows that wage cuts will reduce aggregate demand, discourage consumption and investment and create further unemployment. Similarly the classical economists, who were interested in welfare rather than equilibrium, debated the welfare effects of technical advance and population growth. Ricardo, as a "real economist" (in Hicks' sense, meaning interested in the real quantities lying behind money valuations) observed the real or substitution mode, Lauderdale and to some extent, Malthus, had an eye on the mode of aggregates. This kind of debate is unresolvable until an effort is made to estimate the preponderance, under certain conditions, of the one form or the other.

that we can make no general assumptions about a temporary state of equilibrium at the beginning of any process. If the society were perfectly competitive we could assume a state of general equilibrium at the commencement of any process and go on to inquire how the process affected equilibrium. If the society were characterized by developed monopoly or monopolistic competition we could never assume that it was in general equilibrium.[7] All that we may do, if we wish, is to assume some kind of stability, at indeterminate positions, held in a monopolistic society by usage, balance of bargaining power, and successful, temporary closure of entry. When such a stability is disturbed there can be no way of saying at what point a new, temporary stability will be established. At best the general direction of change is all that can be indicated.

This creates a difficulty because we are most anxious to distinguish between the long-run or secular trend and the sort of cyclical change that originates within the dynamic economy as a result of monetary, investment and saving fluctuations. If we wish to observe the pure effects of some given cause of change (including, of course, the effects through the aggregative mode) we do not want the scene cluttered up by the complicated and complicating set of events which flow from monetary or exchange fluctuations, or fiscal policies, or from independent changes in entrepreneur's expectations, investment and the propensity to save. Yet we cannot, by the method of mathematical ignoration assign constant values to these variables, because all of them must eventually be allowed to vary as they become affected by the given cause of change. The only assumption which we are permitted, and do make, throughout the system of models with respect to these variables, is that no independent changes occur in them, that all changes which do occur are initiated by the cause of change under analysis.

We have already[8] prepared the way for our concept of the temporal process. Time for us means the concrete reality of the economic events in process. Thus the time in each model

[7] See above, page 44. This also is the view taken by Professor Hicks, see *Value and Capital*, p. 265.

[8] See Ch. III.

consists of the process of change flowing from the defined cause, and the time period, we may call it "a day" or "a week" for convenience, must be defined functionally, that is in terms of the events in progress. It will be possible, we shall find, to break down the eventual totality into a succession of processes in terms of which time periods may be defined. Our causal concept has also been defended.[9] The physical event, regarded as a given cause, must be meaningful and evocative of a motive-response within the economic rationale. Economic motives, in every case, form a part of the whole causal situation.

Other limiting assumptions are peculiar to the models, indeed define them from one another, and must be set out in each case. The general plan of the models is to enable us to study, first, population changes, then technical changes, and finally the two combined.

In each case we must analyse the effects of change through the aggregative and real modes, allow for the institutional pattern that will modify the process, and estimate the final welfare effects. We shall define the model in terms of the change to be studied and the level of abstraction, the unit of time and the nature of the time period. We shall then observe the process through the two modes of change, allowing for both competitive and non-competitive institutions. We shall make a final estimate of the probable predominant direction of change.

[9] See Ch. I.

CHAPTER VII

POPULATION CHANGES

MODEL I

In this model the economy at the beginning of the process is regarded as having been evenly progressive in the sense that capital accumulation was proceeding at a constant rate and population increase was at a constant rate, and the proportionate increase in capital was equal to the proportionate increase in labour supply, so that the elasticity of substitution of labour for capital was unity. Now let population come to increase at an increasing rate, and let there be no technical changes or innovations, meaning not only that there shall be no inventions of new machinery or production techniques, but also that there shall be no introduction of new goods, or new uses for old goods, and no discovery of new natural resources. There are no *independent* changes in aggregate money income, money issue, the rate of investment or the propensity to save. Time consists of the appearance on various markets of increased numbers of persons either as consumers or suppliers of services.

We have now to define our units of time. Clearly, for us, time is identical with the process which is going on and we must select from this process suitable units as the basis of its temporal measurement. It will prove convenient to choose two units, one of which will be longer (in calendar time) than the other, so that two or more of the latter units constitute one of the former. We shall call the long unit "the week" and the short unit "the day". The week shall be based on discernible stages of population growth. Thus Week 1 shall consist of the period during which the increased population growth affects the numbers of the population in the non-productive age groups. The new population appears from an increase in the birth rate relative to the death rate; thus in the first stage it takes the form of more children and (possibly) more old people. It is only as the

children reach maturity that any appreciable change occurs in the available labour force. Thus Week 1 has approximately the length of a generation. The appearance in appreciable numbers of the new growth in the productive age-groups constitutes Week 2 which continues until the rate of natural increase adjusts itself to the changed conditions, until, that is, the economic effects of the original cause become causes themselves reacting one way or another on the original cause. With Week 3, which is the period of population readjustment we cannot, in this model, be much concerned, for obviously the process which goes on in this week must run beyond the terms of this model, and very possibly will be in an opposite direction. We may then "enter" Week 3 only in order to see the final effects of our present process as they are manifested in population growth itself, but we shall not be able to "explore" therein.

The short-run unit, "the day", we define to consist of the stages of capital adjustment to population change. This is arbitrary, but convenient. Thus Day 1 of Week 1 is a period too short for any change in the quantity of fixed capital to occur as a result of the population change. Only changes in working capital are possible. In Day 2, fixed capital changes begin to occur. As the new labour force appears in Week 2, the same stages of capital adjustment will be observed. Day 1 will see the immediate adjustments of working capital, Day 2, the slower adjustments of fixed capital, and Day 3, an improbable date, the establishment of a new equilibrium. It will be seen that we depart from Lord Keynes' method of postulating a series of equilibria placed on an open cycle and inquiring into the conditions governing that system of metastable cyclically arranged equilibria. On the contrary we do not postulate any necessary return to equilibrium at all. We admit the possibility of certain metastable positions, but are properly concerned less with the conditions governing them than with the continuity of a process of adjustment, which, as we shall see, is always approximating what would have been a true equilibrium in the immediately preceding time period, but which, in the present time period is, except under special conditions, neither an equilibrium nor stable.

We do, however, assume the economy to have been stable, not necessarily in equilibrium, at the outset. No other, independent, causes of change exist. Actually if the stability is that which exists in a non-competitive economy, *incentives* to change must be present, but we assume these incentives to be temporarily held in check by the balance of power, custom and, in the lack of other checks, inertia.[1] It makes no difference if we assume this stability to be at full employment of labour or not, as long as it is stable and there is no cumulative process at work changing the quantity of employed. It will probably be simpler to suppose full employment of labour to exist, save for a small, "normal" labour reserve, of, say, five per cent of

[1] Inertia, as a source of stability in non-equilibrium situations, has received too little attention. There is great entrepreneurial insensitivity to small changes. This is especially marked when competition is not perfect. Monopolists are most reluctant to call attention to themselves by increasing prices, and require a stimulus of pronounced magnitude to do so. Small competitors of monopolistic concerns are also usually content to let sleeping dogs lie. As long as the situation is tolerable for them, they are inclined to leave it alone.

It is possible to give the concept of inertia an exact significance as in Fig. 1. The equilibrium position of a monopolist is defined by the intersection of the MC and MR_1 curves.

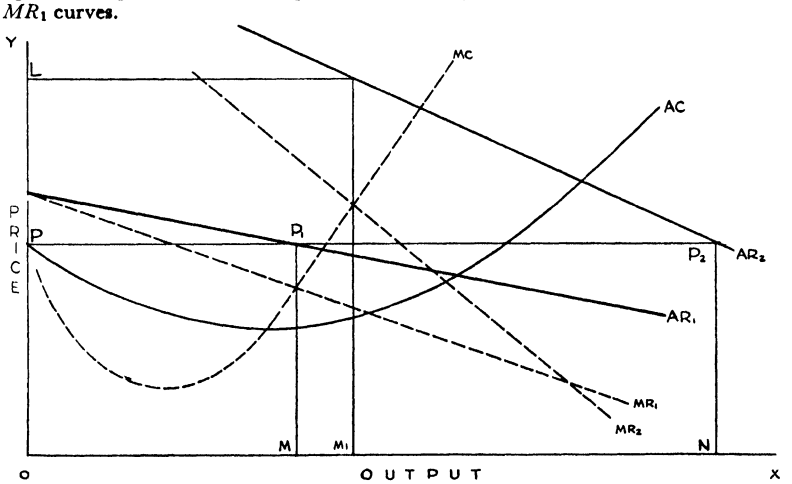


FIG. 1

Let us suppose him to be at this position. His average revenue now shifts positively to the position designated by AR_2. If he adjusts at once to the new situation he will move to the output OM_1 and the price OL. Such an adjustment would mean that his inertia was zero. If, however, he is completely inert he will attempt no price change whatever and his sales will automatically increase to ON. In the case here, as we have drawn it, this would palpably be an impossible situa-

the labour force. We cannot, of course, suppose there to be full utilization of resources in the non-competitive economy, but that assumption is implicit in the assumption of a perfectly competitive economy in equilibrium. We shall further assume the economy to be a mature one in which the point of diminishing returns to land has been passed and in which, in the lack of new technical advance, the point of constant returns to scale has been passed.[2]

We begin with the case of perfect competition in both commodity and factor markets. On the first day of the first week the increased population growth will produce increased numbers of people in the non-productive age groups. Thus aggregate demand will be increased. The effects from here on, throughout Day 1, can best be observed through the aggregative mode, because, for obvious reasons, as long as we are in Day 1, the effects through the real or substitution mode will be comparatively unimportant. The most probable inferences are that, with the increase in aggregate demand, prices will rise. Aggregate money income also will increase, as will real income, but, because of diminishing returns, the increase in the latter will be less than proportionate to the increase in population. Moreover the nature of the population increase—in the non-productive age groups—must accentuate the decline in average real individual income.

Before we can proceed with inferences as to the effects through the aggregative mode we must examine such real mode effects as are relevant at this stage. We began with "full employment" of labour and no "excess capacity". (We have perfect competition.) It is a grave misunderstanding to suppose, however, that this means that no increase in production can occur.

---

tion, but if Average Cost lay below Average Revenue it would be quite possible. Thus if $\frac{M_1N}{LP}$ is infinity, inertia is infinite. If $\frac{M_1N}{LP}$ is zero, inertia is zero. If we now call whatever increment in output accrues beyond the new equilibrium output, r, and whatever increase in price occurs beyond the old time equilibrium price, p_1, the ratio $\frac{r}{P_1}$ measures the degree of inertia.

[2] " . . . I am convinced that it is more convenient not to regard the case of constant costs one's standard case". Hicks, op. cit., p. 96. We shall show in Ch. IX the conditions under which returns to scale may be increasing or decreasing.

"Full employment" is an "aggregative" concept and means simply that no person is unemployed who would wish to be employed at going wage rates, with the ruling price level unchanged. If there is a change in prices or money wages, or both, the supply of labour can be increased by attracting persons to the labour market, such as housewives or children near school-leaving age, who are not normally attracted, or it can be increased by a lengthening of the working day. Excess capacity means, in economics, the difference which may exist between plant utilization to the point where the average revenue and average cost are equal, and any utilization short of this point, such as utilization to the point of output where marginal revenue and marginal cost are equal. Obviously, if average revenue or average cost is changed, or both, production may be increased or diminished. Thus we are quite justified in concluding that with the positive shift in aggregate demand, output will be increased by the more intensive application of labour and working capital to the fixed factors. Now, speaking generally, we are justified, on empirical grounds, in assuming that working capital (which is the only capital that can be increased in the short run) and labour are complementary, not substitutable, factors. Hence the margin of substitution will be altered in favour of these short-run variable factors as against land and fixed capital. Their physical joint marginal productivity will decline, but, because of the rise in prices, it does not follow that their marginal value productivity will decline. All that does follow is that their monetary rate of reward cannot rise in proportion to the increase in prices. The expansionist process will, however, lead to an increase in monetary rewards (there need be no demurrer about this because we are assuming a competitive and mobile economy) so that there will be a double incentive for entrepreneurs to increase fixed investment. The first incentive comes from the expansionist process generated through the aggregative mode; the second incentive is the increasing marginal cost and diminishing marginal physical productivity of the short-run variable factors which sets up an incentive to substitute long-term capital for

these factors. We may, therefore, conclude that Day 1 will witness the laying down of long-term capital formation.

The other real effect is within the aggregate demand function. The change in population and the increase in money prices must lead to the substitution in consumption of some classes of goods for others. On empirical grounds it is obvious that baby carriages and midwives' services, to select two examples, will be substituted in family budgets for motor cars and night club entertainment. Inferior goods will be substituted for superior goods. Necessities, with inelastic demand schedules, will rise in price with great flexibility so that more income will have to be devoted to these and the reduction of consumption of comforts in more elastic demand will be the greater. If the society were inflexible and immobile the spotty unemployment which would result might become a serious deterrent operating against the general expansionist trend, but under the conditions set out by our present assumptions it is probably correct to infer that this deterrent effect would not be preponderant and that the general expansionist trend observed as operating through the aggregate mode would be but little affected.

Let us return then to observe this trend more particularly. There are two related questions of some complexity which must now be answered. The first concerns the rate of saving. With an increasing aggregate money income the general expectation which modern economic theory would encourage would be an increase in the marginal propensity to save. More careful reflection, however, will show that this expectation is not justified. Lord Keynes himself does not claim that the marginal propensity to save necessarily increases with an increase in aggregate income. All that he says is that the marginal propensity to consume (unity minus the propensity to save) is positive but less than unity, and is likely to be a diminishing rate between the limits 1-0.[3] This likelihood, however, is based on the assumption that the economy is a mature one and that the increase in income is a cyclical one. Under the conditions we are assuming the probability is the other way.[4]

[3] J. M. Keynes, *The General Theory of Employment, Interest and Money* (London, 1936), p. 86.

Though money incomes are increasing, so are prices and the average size of families, both the latter more rapidly than money income. If we may use the income elasticity concept, it appears an almost certain inference that the proportion of new income spent will increase, i.e., the elasticity of spending to income will be more than unity. The rate of voluntary saving will decline. Hence the decisions of entrepreneurs to lay down new fixed capital equipment for use in Day 2 must impose or force real saving on the community by reducing the factors available for the production of consumers' goods and still further inflating

[4] The likelihood of a diminishing propensity to consume may be challenged on other grounds. In an, as yet, unpublished manuscript, Dr. H. G. Littler argues that Lord Keynes defines investment, consumption and income as primitive functions of time (*The General Theory*, pp. 54-63). Thus Lord Keynes is not using the usual concept of income as a rate in time, of, say, $100 a month, but is treating it as a cumulative aggregate over time. The difference for non-mathematical readers may be made clear graphically. Suppose an income to be earned for six months at the rate of $100 per month in the first, third and fifth months, and $200 per month in the second, fourth and sixth months. Then, in our everyday use of "income" it would be graphed like this: (*See pp. 120-1 for end fn. 4.*)

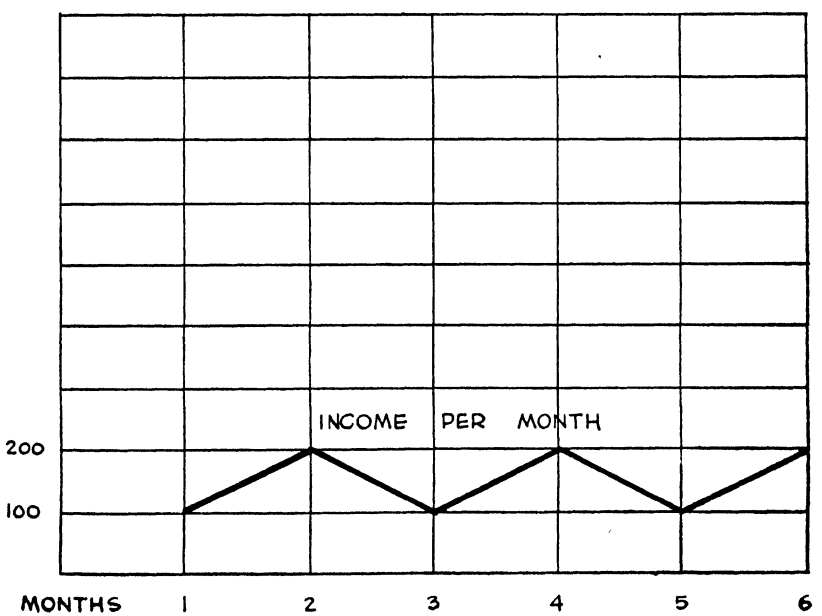


FIG. 2

the prices of those goods. Day 1 will consequently witness an accelerated increase in prices.

This leads directly to the second question to which reference was made above, viz., the question of entrepreneurial expectations. We are going to assume that under conditions of perfect competition the elasticity of expectations will be between zero

But the primitive function, the one Lord Keynes chooses to use, would look like this:

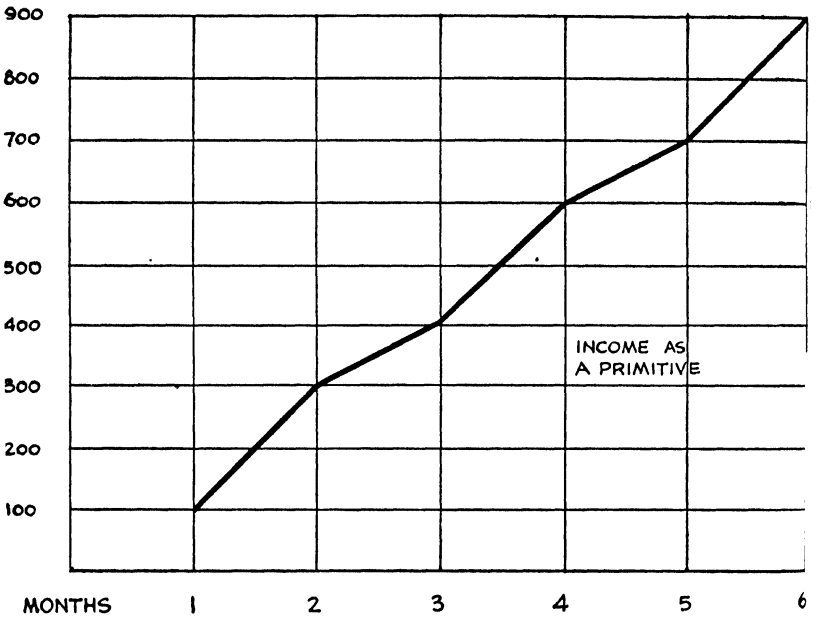


FIG. 3

Mathematically speaking, Lord Keynes has defined income Y as a primitive function of time, and consumption C as some other primitive function of time.

Thus $Y=f_1(t)$,

$C=f_2(t)$.

Hence what we ordinarily call income and expenditure are the rates at which Lord Keynes' "income" and "expenditure" increase with time, i.e., they are $\frac{dY}{dt}$ (income) and $\frac{dC}{dt}$ (expenditure on consumption).

Now it appears that Lord Keynes can mean one of two things by his "marginal propensity to consume". He can mean what he says, in which case marginal propensity to consume is $dC/dt \Big/ dY/dt$ or $\frac{dC}{dY}$. If this is so, the general psychological law is quite unwarranted. It would mean that in any short interval of time Δt, say a month, out of the sum you received that month, i.e., that month's increment to your cumulative total of receipts, you would always spend a sum less

and unity.[5] This means that as long as current prices are rising entrepreneurial decisions as to future outputs will be based on the assumption that the prices these outputs will command on the market when they are ready for sale will be increasing though not necessarily in the same proportion as are current prices. It is fairly clear that two factors influence expectations. One is the existence of organized "future" markets with quoted prices. Their existence makes expectations more definite and makes the concept of elasticity more precise. The other factor is the degree of monopoly, because cartels and monopolies make for inflexible prices and so, on the basis of experience of past price trends, lower the average elasticity of expectations. The absence of any degree of monopoly in our model justifies the assumption of a high elasticity of expectations.[6]

We enter Day 2, then, with entrepreneurs deciding on still greater outputs involving the intensive use of variable factors but with the appearance now of an increased supply of fixed

---

than the total amount received. This is nonsense, and, of course, not what Lord Keynes means to say at all. What he wants to get at is the way we change our *rate* of spending as we find our *rate* of income changed.

Let us, then, see if the second derivatives would give us Lord Keynes' "marginal propensity to consume". In that case the proper expression would be $\frac{d^2C}{dt^2} \Big/ \frac{d^2Y}{dt^2}$. At first sight this would appear to be what Lord Keynes really intends to mean by his marginal propensity to consume, for, in concrete terms, it means the rate at which you increase your total expenditures over time with a change in the rate at which your total receipts change in time. It is doubtful if one can conclude that this rate is in any general sense necessarily a diminishing one.

[5] "I define the elasticity of a particular person's expectations of the price of commodity X as the ratio of the proportional rise in expected future prices of X to the proportional rise in its current price" (J. R. Hicks, op. cit., p. 205).

[6] The data of the studies of the hog cycle would seem to indicate that under perfect competition in this form of agriculture, producers always anticipate the present price to be maintained for the two-year period, or, at least, always behave as though the present price were to be stable. This would mean that they had no elasticity of expectations at all, the change from the price in t_{-1} to t_0 being regarded as having no effect on the price at time t_2. The cobweb theorem is an ingenious analysis explaining an observed statistical phenomenon, but it is of doubtful validity. It supposes demand and supply functions to remain unchanged and entrepreneurial expectations to ignore all past experience. It is probable that there is a connection between the two alternating cycles and that this connection lies in expectations which shift with the previous year's experience. However, it may well be possible that the assumption of an elasticity approaching unity is too high, even in a competitive society.

factors. The general effects through the mode of aggregates may be expected to continue, though at a slower rate. This is because of effects discernible through the real mode. Though the new fixed capital may be regarded as being belatedly complementary to the increased use of the variable factors in Day 1, its present appearance must alter the margin of substitution in favour of fixed plant as against still more intensive use of labour and working capital. We have, in fact, increased our scale, but since there are no qualitative improvements, and since land and other factors in inelastic supply will decline in quality, we must expect less than proportionate returns to scale. Now the fact that the disproportionate intensity in the use of fixed factors of Day 1 is being corrected removes or lessens one of the incentives to still further increases in investment in fixed plant. With the increased supply of the fixed factor and the increased interest rates which must have accompanied the forced saving inflation in Day 1, the marginal productivity of fixed capital will decline relative to its marginal cost and hence advantages from further substitution will decline. The tendency will be towards some new equilibrium in the combination of the factors. This will mean less forced saving and the freeing of some factors engaged in Day 1 in making producers' goods for employment in consumers' goods industries. Prices will still be increasing, however, because the increase in population and aggregate demand is still continuing faster than the increase in production, but the reduction in the rate of investment in fixed plant and the transfer of resources to the production of consumers' goods must relieve some of the inflationary pressure on the price level. Hence, though prices continue to increase, they do so at a less rapid rate, and entrepreneurial expectations, with a high elasticity, are somewhat less optimistic as a result. Consequently less ambitious expansionist planning will exist and the general rate of expansion, including aggregate money income and the rate of investment, will decline. The economy, however, can scarcely cease to be expansionist as long as the increase in aggregate demand continues.

In the second week the new population appears in appreciable quantities on the labour market. This makes possible further

increases in production to meet the continually expanding demand, but still under conditions of diminishing returns. If new fixed capital could immediately be created and if returns to scale were constant, the existence of diminishing returns would not be important. It is worth noting here, for future reference, that it is because of the relative inelasticity of the supply of fixed factors and the assumption of less than constant returns to scale, that we get this persistent diminution in average household real income. Household real incomes may now increase, of course, as over Week 1, but they must be less than at the beginning of the population change and as Week 2 progresses must further decline.

The effects through the aggregate mode continue in Day 1 of Week 2 to be expansionary. Aggregate income, investment, employment, consumption all increase. There may be some easing of the inflationary pressure on prices because money wages will not, we shall see, increase in anything like the proportion of Week 1, if they increase at all. There may well be an increase now in the rate of voluntary saving (as a result of young wage-earners increasing the family income and undertaking long-term contracts for insurance or pensions) and the interest rate may be reduced.

The effects through the real mode are here closely allied to the aggregate effects. The appearance of new labour, seeking employment, will exercise a depressive influence on money wages. With aggregate demand continuing to expand, entrepreneurs will be anxious to expand output. The increased supply of labour, with no trades unions, will tend both to reduce its supply price and its potential marginal physical productivity, and the competitive pressure of more workers seeking employment offsets the pressure of rising money prices on money wages. In brief, labour must accept a declining real wage proportionate to the decline in real productivity if it is to find employment.[7]

[7] This conclusion, which conforms to the orthodox position on the relation between wages and employment, rests on the assumptions which define our present model. It is, perhaps, some vindication of our method of inquiry that it helps to make clear the underlying assumptions that are necessary for the conclusions achieved by various schools of economic thinkers.

In the Second Day new fixed capital will adjust itself to the new labour supply. The fact that labour is relatively cheaper and that aggregate demand continues to increase will mean that the new capital will be complementary rather than substitutable for labour so that there will be a "widening" of capital and a "thinning".[8] Thus the substitution mode effects will be continued until a new equilibrium is reached between the ratios of marginal factor cost to marginal value productivity for labour and capital.

The effects through the aggregative mode will be expansionary as long as the increase in aggregate income continues. The argument that there will be a failure of investment on which employment depends, if there is no deepening of capital, is invalid, because it fails to take account of continual increased investment in complementary capital which will continue as long as aggregate demand increases.

Thus we have an expanding economy with a falling real welfare level. Great pressure for technical innovation would, in fact, be generated, and it is unlikely that under real circumstances technical advance would not occur. It might be, however, that any technical advance that did occur would conform to Ricardian expectations and would only result in an eventual further decline of real wages. For the moment we may observe that failing the achievement of better than constant returns to scale, the eventual pressure of declining real wages will be reflected in population growth itself. If one takes a sufficiently large area, a sufficiently stable period, and a sufficiently short period, one can, indeed extrapolate population by means of logistics. In effect this means that future population growth is in part determined by past rates of growth and present population structure. But this admission does not mean that population is not sensitive to economic fluctuations and trends. No extrapolation of population for sub-divisions

[8] A widening of capital is an increase in the use of capital that is *not* more than proportionate to the increase in the use of labour. A deepening of capital occurs when more capital per labour unit is employed. Cf. R. G. Hawtrey, *Capital and Employment* (London, 1937) p. 35 et seq. A thinning of capital occurs when less capital per unit of labour is used. A fuller discussion of these concepts in relation to innovations and investment will be undertaken in the next chapter.

of the large area on a basis of logistics is valid because of migration from region to region within the area. Thus if the United States closes itself to migration, the population for the nation may be extrapolated with some confidence for a short period, provided no violent economic changes occur. But the population of Kansas could not. Further, marriage, birth and death rates alter with economic conditions. Crude death rates rose slightly during the last depression, marriages were postponed (the average age of marriage increased) and birth rates declined. During the wartime boom, on the other hand, in the countries not devastated by war or depleted of male population, the marriage age declined and the birth rate increased. Thus a declining level of real welfare will be reflected in population growth. Emigration, the postponement of marriage and the deliberate restriction of families will occur and will eventually change the basic structure of rates of the population so that no curve extrapolated by purely mathematical techniques will be valid. In the case of our present model at some time or other this change must occur. It may "start to begin" in Week 1 or Week 2, and will become evident in a considerable slackening of the rate of population increase in what we have called Week 3, which takes us, of course, out of our model onto a different level of assumptions.

Very obviously the present model is Ricardian. It gives us an expanding population, expanding economy, rising rents, but declining real welfare which ultimately leads to population restrictions. We see the critical assumptions to be that of less than constant returns to scale[9] (implying no technical advance or innovations which are "capital saving" and employment reducing), and of an institutional structure of freely competitive and mobile bargaining units. The importance of the institutional structure may be demonstrated by a study of the same model on the assumption of a trustified system.

We shall avoid tedium if we restrict the present inquiry to

[9] This is the important assumption, not the assumption of diminishing returns. Diminishing returns in the strict sense always exist, yet if returns to scale are increasing, output may be increased more than in proportion to the increase in the supply of the factors. See Ch. IX.

the point where the changed institutional assumption would modify the argument. What we are doing, in effect, is to introduce institutional rigidities to the model. The rigidities affect adjustment at the margin of substitution and they also affect the response of the aggregative factors of income, investment, employment and aggregate demand. Hence they make their appearance both through the real and aggregative modes.

Thus in Week 1, Day 1, the existence of monopoly and imperfect competition[10] will mean that there will be excess capacity and idle resources. The effects through the aggregate mode will mean a general positive shift in revenue curves and the more complete utilization of existing plant, so that there may well fail to develop anything like the expansion of investment that was expected when competition was perfect. We may therefore expect the aggregative effects to be less expansionary than under competition, and the laying down of new fixed capital less great. Further, the substitution mode effects will be modified. The shift in the structure of demand will find the factors less mobile in response, so that a general profit inflation may occur while considerable unemployment develops in industries whose products are sold to a highly elastic demand. The tendency to substitute labour and variable complementary capital for fixed capital will likewise be restricted, especially as trades unions will raise the marginal money cost of the labour factor with the increase in prices. Unorganized labour and rentiers will suffer from the increase in money prices. Organized

[10] Because of some confusion in the literature in the use of these terms, we shall define our use of them. We shall mean by perfect monopoly the condition when there is only one seller in the market. By perfect competition we mean the condition that exists when the revenue curves of all the firms are infinitely elastic. By imperfect competition we mean every condition that may lie between these limits. It is thus used in the most general possible sense, including as sub-categories the cases of oligopoly (a few sellers engaged in price competition—an impermanent and unstable situation), monopolistic competition (a few sellers competing to differentiate products with some sort of price agreement), and unequal competition, where a large price leader or a group in monopolistic competition set the price, and smaller firms accept it as a straight line revenue curve and behave within the framework so established as though under perfect competition. By a monopolistic or trustified economy we mean simply one in which imperfect competition in the most general sense prevails.

labour and entrepreneurs will benefit relatively, in that the distribution of the real national income will be modified in their favour.[11]

In the Second Day the amount of new fixed capital appearing will be less than under the conditions of perfect competition; the effects of this will depend on the degree of complementarity of this capital with respect to labour. On the whole we should not be justified in assuming that it would lead to a replacement of labour, unless technical improvements had been made. We might note here, in passing, that the pressure for capital innovation is great and that our assumption of no innovations is difficult to retain.

It is also hard to infer whether entrepreneurial expectations would be more or less elastic than under competition. The general assumption[12] is that, under conditions of imperfect competition, expectations are for steady prices and thus the elasticity of expectations is low. This would mean that Day 1 of Week 1 would not witness the same rate of continued new investment as it did when perfect competition existed.[13] On the other hand, it may well be that it is less price expectations, in the precise sense as defined by Professor Hicks, but rather general anticipations of demand that guide entrepreneurial behaviour under imperfect conditions. If so, the continued increase in population might result in intelligent anticipation of increased demand and so an increase in the laying down of new capital to provide for it.[14] As long as we assume no innovations to occur we may generally infer that this new investment will not be great. If innovations are possible, we should have to admit a probable deepening of capital to occur. This problem

---

[11] Cf. the discussion of the sharing of gains from a technical advance in Ch. VIII.

[12] Cf., J. R. Hicks, op. cit.

[13] Again we recall that no innovations occur. We shall show reason to believe that innovations are more likely to occur under imperfect conditions and, if they did, then an investment boom might occur at this stage, but the resulting fixed capital would be substitutable for, not complementary to, existing factors.

[14] I suppose one should consider the development of advertising techniques as an innovation. Certainly under the conditions here assumed, the kind of advertising directed towards commodity differentiation and discriminating monopoly would flourish, with consequent effects on the demand structure. See Ch. IX.

is one to which we shall return at a later stage. The increase in interest rate and in the wages of all organized workers will combine, at any rate, to check the rate of increase in investment.

Thus, when we move to Week 2, we see that the expansionary income effects are less great than they were under perfect competition, spotty unemployment is more serious, and the decline in real welfare is more unevenly distributed. It will be remembered that at this stage, even under perfect competition, the incentive to new investment was reduced. With imperfect competition, the reduction is the greater by reason of the less expansionary income effects. Moreover, there is no reason to suppose that the new working force now appearing will find employment. Unless wages are reduced it will not do so. It is probable that acute social conflict will be generated. Attacks on the trades unions will be made. The new unemployed, never disciplined in the union movement, may well provide the mass support for fascist or crypto-fascist political forces. However that may be, we must expect either a decline in wages, or unemployment, or both, with recessionary aggregative effects. Briefly, we may conclude that at this stage the expansionary movement will come to an end in the monopolistic economy, and to the suffering of those social groups who have so far borne the burdens of the decline in average real income, will be added the disaster of widespread unemployment and economic depression. The eventual decline in population will not necessarily lead to relief for it may only serve to "set" the stagnation and prevent recovery of investment and employment. Into that we shall later inquire.

The Ricardian conclusions, then, would not hold for an imperfectly competitive economy. Trustification accents the decline in welfare and modulates expansionary income effects, so that under the conditions assumed in this model, the economy becomes both declining and stagnating. We shall observe, however, that these conditions are unfair and inadequate to pass final judgment on the welfare effects of trustification. Before that can be done the question of innovations must be considered.

Model II

We turn now to consider the case where population growth is declining. It is doubtful if this model can be examined in detail with much profit, because, as we shall see, it is almost impossible to retain the assumption of no innovations. Further, since we are primarily interested in classifying the conditions necessary to certain theories of economic change, and since the thesis which we wish to examine in connection with declining population growth is that of Professor Hansen, it will only be profitable to consider declining population growth in connection with certain specific assumptions as to innovations and frontiers.

There are, however, one or two points to be observed in a pure population model when growth is declining. We shall, therefore, make the necessary assumptions that no other causes of change are operative and that any monetary changes are those induced by the change in population. The initial cause of change is a decline in the rate of population growth. If this decline continues unchecked it will become progressive as a result of the decline in the number of women of child-bearing age[15] and must eventually result in an absolute fall in total population.

Ricardo supposed that such a change must follow a decline in the standard of living and he infers that the decline in population will result in a recovery of living standards. The welfare effects, that is to say, would be progressive. Ricardo supposes this to be so because, with too little examination, he assumes that a perfectly reverse action will develop from that which follows an increase in population. That is to say, he supposes that the decline in labour supply will shift both the intensive and extensive margins of cultivation to the left, i.e., reduce both the intensity of use of the fixed factor and the use of low-quality land, so that average returns per worker will be enhanced. This argument neglects three probable complications, viz., the irreversibility of certain

[15] It is an obvious implicit assumption that the decline comes from a fall in birth rate. Some sudden increase in death rates from war or pestilence would not be sustained and continuous.

cost curves, the effects of declining population through the aggregate mode, and rigidities which stand in the way of economic adjustment to the changed population circumstances.

The assumption of diminishing returns, in the strict sense, simply implies that at any moment of time more intensive use of fixed factors will yield diminishing increments of product. This rule would be reversible. But a process over time, even if innovations are ruled out, involves the concept of returns to scale, and, as we earlier saw, it is really an assumption of diminishing returns to scale that underlies Ricardo's argument on increasing population. This concept is not reversible. Though, on the assumption of no cost-reducing innovations, one may assume that increases in scale will, for one reason or another, result in less than constant returns, once the increases have been made and the capital commitments undertaken, a reduction in output and eventually in scale will not follow in reverse the path of expansion. On the contrary the reduction of output may well result in higher average unit costs, and scale reduction may leave a heavy burden of unliquidated fixed and user cost to be distributed over the reduced output. Thus, any distribution of capital for labour with a consequent enhancement of labour's marginal productivity is unlikely. Indeed the aggregative effects are so disastrous that investment is sure to decline, and with it employment, and the decline in employment will be more rapid than the decline in population, unless labour is willing to accept a greatly reduced real wage. We can infer that the stagnation effects through the aggregate mode will predominate.

Moreover, there are always serious problems of adjustment to a declining population. It is true that adjustment problems and attendant rigidities exist in any economy experiencing population growth, but the increasing aggregate demand provides a strong economic incentive to make the adjustments. Entrepreneurial expectations are good, in spite of increasing costs, so that new investment to meet shifts and changes in the structure of demand is rapidly forthcoming. The opposite is true when aggregate demand is declining. New venture capital to meet particular increases in the demands for certain goods,

is not apt to be available, and in the declining industries every effort will be made to reduce wage costs and "hang on" with the existing capital formation. Bankruptcies will increase in number and still further discourage investment and new venture.

If the economy is trustified the struggle between organized labour and capital will become intense and may lead to attempts at revolution or counter-revolution. These social and political uncertainties and instabilities increase the caution of business men. It is highly unlikely, therefore, that under the circumstances here assumed, any welfare improvement would be achieved by a reduction in population.

CHAPTER VIII

INNOVATIONS

1. SOME DEFINITIONS AND GENERAL OBSERVATIONS

Technical advance, more or less continuous, but at a varying rate, is a characteristic of modern European society. The effects of technical progress on the economy have received the attention of economists of the modern period, but this attention, as the literature attests, has seldom been systematic. With a few exceptions, economists have rather regarded technical change as something external to the economic order, something to be excluded from the systematic models, and to be explained, as a modifying and destabilizing force, in the addenda and postscripts with which careful economists adorn and buttress their theoretic architecture.[1] If, however, one is intent on inquiring into the sources of material welfare and the cause of changes in the wealth of nations, the nature of technical progress and its economic effects become the very centre of study and such general principles as may be discovered become the foundations of a theoretic system.

Technical advance may assume many forms. It may consist of new processes of production, involving the use of new machines or new materials, or of new systems of organizing production, distribution, exchange or finance, or in improved techniques of plant management. It may take the form of the discovery of new goods, new materials from which new goods—or some old goods previously made from other materials—may be made, or of new sources of supply of old materials, or of the invention of new uses for old goods.[2]

[1] I am referring in the text to theoretic economic literature, not to economic history.

[2] Cf. J. A. Schumpeter, *The Theory of Economic Development*, tr. Opie. (Cambridge, Mass., 1934), p. 66. "This concept (i.e., development, defined on p. 55 as consisting of innovations) covers the following five cases: (1) The introduction of a new good . . . (2) The introduction of a new method of production . . . (3) The opening of a new market . . . (4) the conquest of a new source of supply of raw materials . . . (5) the carrying out of a new organization of any industry."

We call any particular advance in the technology of production an *innovation*, when it is adopted by enterprise. It is important to distinguish between an invention, which is a technical or scientific fact, and an innovation, which is an economic fact.[3] It is clear that an invention will not become an innovation unless entrepreneurs believe that it will enable them to increase their profits. Thus the state of entrepreneurial expectations and the institutional framework of the economy (whether making for a rapid adjustment of prices to costs, as under perfect competition, or for a slow adjustment allowing for the prolonged retention of windfall differentials between prices and reduced costs) must be considered as determinants of the rate of innovation. Hence, while it may be true that the rate of invention may be caused by "accidents" external to the economic causal order, the rate of innovation is, for the most part, within that order.[4]

All innovations in processes are cost-reducing. This proposition follows directly from the definition of an innovation. No invention of a new productive process would be adopted if it were not expected to reduce the unit cost of producing an article of the same quality as before, or to enable an article of superior quality to be produced for the same cost, which is the same thing.

[3] "As long as they are not carried into practice, inventions are economically irrelevant" (Schumpeter, op. cit., p. 88). This usage, which I have followed, differs from that of the historians of mechanical invention. Professor Usher restricts the term invention to mechanical invention, and speaks of innovation as the more general term denoting the general "synthetic, constructive and creative" mental processes. Thus "innovation is an integral part of the process of learning, an inescapable necessity for the individual as for the group as a whole". "Mechanical invention is only one phase of the innovating activities of the mind". See A. P. Usher, *A History of Mechanical Inventions* (New York, 1929, pp. 8, 10).

[4] My attention has been drawn to the phenomenon of stylistic changes in architecture which, by reason of their effects through the construction trades, may work profound changes on the economy as a whole. Thus the fashion for baroque building led to the construction throughout Catholic Germany in the sixteenth century of new monasteries. The period of chateau construction in France might serve as another example. In a sense, such building might be regarded as an innovation, though scarcely as we have defined the term. Though it is true that the motive for this type of new building is not an entrepreneurial motive, it does not follow that the development is non-economic in cause. It is really a demand change, as are most changes in styles, and, in the examples given, was a kind of Veblenesque "conspicuous spending" engendered by the rapid growth of European wealth in the period following the discovery of America and the opening of the Spanish American mines.

Innovations in goods cannot be said to be cost-reducing in the same sense, because, obviously, a new good not previously produced has no pre-innovation cost. New goods, however, are always, within the general demand structure, substitutes for old goods. They may be close substitutes for some old goods, and distant substitutes, through the reallocation of consumers' outlay, for all old goods.[5] They will not be introduced unless they can be made at a lower cost than existing close substitutes of the same quality, or if their costs are higher than those of close substitutes, unless their quality is expected to exceed that of close substitutes by more than the excess of costs, and unless aggregate income is expected to be sufficiently buoyant to absorb the new good at a price which will reward venture with profits.

Every innovation, then, results in substitution. Capital in a new form may be substituted for capital in an old form, as when the diesel engine replaces steam, or capital may be substituted for labour, as in most labour-saving devices, or one form of labour may be substituted for another. Most innovations in processes involve all these substitutions conjointly.

It is possible to classify innovations according to their substitution effects. Thus we may speak of an innovation as being "capital-saving" or "capital-using", and great ingenuity has been lavished in giving exact significance to these terms.[6] For our purposes we shall define a capital-saving innovation as one that results in an alteration of the margin of substitution in favour of labour, that is in a greater proportion of labour to capital per unit of final product. Thus capital-using innovations

[5] Cf. J. R. Hicks, *Value and Capital* (Oxford, 1939), Ch. II.

[6] Thus Dr. Lange ("A Note on Innovations", *Review of Economic Statistics*, Vol. XXXV, 1943) distinguishes innovations as being "factor-saving", "factor-using" and "factor-neutral", according to the effect on the employment of a factor of the change in the factor's anticipated marginal physical productivity. J. R. Hicks (*The Theory of Wages*, London, 1932) calls an innovation neutral when it raises the marginal productivities of labour and capital by equal proportionate amounts, labour-saving if it raises the marginal productivity of capital more proportionately than that of labour, and capital-saving if it raises the marginal productivity of labour more than that of capital, hence leading to a substitution of labour for capital. Mrs. Robinson ("The Classification of Inventions," *Review of Economic Studies*, Vol. V, pp. 139-42) accepts a suggestion of Mr. Harrod and classifies innovations in terms of the relative shares of the factors; thus, given a constant rate of interest, an innovation would be neutral that left the relative shares unchanged.

result in a deepening of investment (in Mr. Hawtrey's sense), a neutral innovation to a widening of investment, and a capital-saving innovation to a "thinning" of investment. It is a mistake to infer, however, that a capital-using innovation necessarily leads to increased total investment, or to diminished employment of labour. It is equally an error to believe that a capital-saving investment necessarily leads to a diminution in investment. If one assumes full employment of all factors when the innovation is introduced, it would then be true that a labour-saving innovation might create unemployment if the demand for the commodity affected were relatively inelastic. If, however, the economy is regarded as moving through the stages of the cycle, the aggregative mode effects may either offset or exaggerate the real mode substitution effects of the innovation. In our opinion the nature of innovation is as much effect as cause of cyclical movement (labour-saving innovations being bunched in the expanding and crisis stages, and capital-saving innovations in the depression stage), and the total effects are determined more by the aggregative factors than the degree of substitution in the combination of inputs.

We are also doubtful of attaching too much importance to what is essentially an equilibrium concept in the theory of long-run changes. A more important conceptual distinction for our purposes has to do with the "adaptability" of an innovation. Some innovations are peculiar to the conditions of a particular plant or particular industry. Some innovations of new goods are particular and have no further applications. These are innovations of limited adaptability. Other innovations, such as the introduction of a new material, or the development of the principle of the wheel or the valve, have wide adaptability and are taken up, in a series of modifications, by industry after industry. Some innovations not only have wide adaptability but necessitate still further innovations. We may call these, for want of a better name, "linked innovations",[7]

[7] They correspond to what Professor Usher has called "strategic inventions" (op. cit., p. 23). "This interdependence of individual inventions upon each other appears even at the beginning of technical progress, when all innovation is empirical" (p. 19).

because their acceptance is linked with a whole series of attendant and consequent innovations throughout a wide industrial sector. The link need not always be technical; it is sometimes economic in nature, as when joint products appear as a result of one innovation, and an innovation is introduced for the use of the by-product. Most important among such economic links are those which grow out of the existence of high overhead costs and excess capacity. Thus the development of a continental railway system led to differential freight rates and consequent economic stimulus to the development of new regions and new products, a series of innovations following the railway. Similarly with the use of electrical power at "off peak" rates. The link sometimes may be purely scientific or intellectual,[8] as when some development in science such as the Ptolemaic or Newtonian revolution is seen to have first one and then a series of applications to the industrial world. In most instances the links are scientific, technical and economic altogether. There may be a common scientific basis, followed by technical adaptability and the generation of economic pressure for successive adaptations. Most characteristic and most significant of such "linked" innovations are innovations in the sources of power and heat, especially in prime movers.[9] The invention of the steam engine was, in one form or another, adapted as the prime mover in textiles, shoes, steel, engineering, rail transport, shipping, hoisting, etc., etc. Its successful exploitation required innovations in metallurgy, in coal mining and use, in engineering, in precision instruments, in die-making and machine tool construction, in chemicals, in multiple machines, in coach building, in ship construction, in harbours and canals, in hostelry and catering trades, in new uses for metals, textiles, transport facilities, even in banking and exchange services, and so on through an almost inexhaustible list. Similar lists might be

[8] "The first stage in the history of these machines begins with the discovery and formulation of the underlying principles of pressure in gases" (Usher, op. cit., p. 25).

[9] Though we stress prime movers in the text, the importance of new sources of heat is not to be neglected, especially in connection with heat-using industries of the chemical sort. See, for the importance of the introduction of coal as a source of heat, J. U. Nef, *The Rise of the British Coal Industry* (London, 1932), Part II, Ch. III, Vol. I.

constructed of the innovations linked, by necessity, to the introduction of electrical power, the internal combustion engine, the diesel engine and, one suspects, before long, of atomic power. Linked innovations—or advances, to use a more general word to refer to the process in its entirety—usually bring in their train various mechanical innovations of wide adaptability. Professor Schumpeter refers to the "bunching" of innovations.[10] Innovations are "bunched", but in no haphazard manner. They are a constellation about a single sun. Particular innovations of limited adaptability may continue fairly continuously in a mechanical civilization, but the rate at which they are introduced, whether continuous or discontinuous, hardly matters.[11] They are not important in generating general effects on the economy. The linked advance is the important cause of economic change. We believe this linked advance to be associated most generally with an innovation in prime movers. That it must always be so cannot be said, but, in modern technical history, these innovations have been the significant and strategic ones.[12] Since there is no apparent reason why innovations in prime movers should occur at any regular intervals, and since, in fact, they have not done so, we must reject the appealing notion—for it surely cannot be called a theory—that the long-term trend of the economy has a regular cyclical periodicity.[13] There is in capitalist society a short cycle of seven to ten years' period which

[10] He is referring, of course, to the bunching of innovations at the upturn of the business cycle, a quite different phenomenon from that referred to in the text.

[11] That is a loose expression. The introduction of particular innovations "hardly matters" as a prime cause of economic change in the long run or secular sense, which is what is intended in the text. The rate of particular innovations will vary, however, with variations in the rate of investment and these variations do matter, of course, as a part of the aggregative aspect of the ten-year cycle.

[12] Cf. Usher, op. cit., Chs. XI and XIII.

[13] I am not denying the possibility of "swings" or reversals of trend. The "Kondratieff", in the sense of long-run or secular trends, subject to reversal, is admitted. It is the cyclical nature of the long-run trend with a regular periodicity that is here denied. The Kondratieff and Spiethof evidence does not cover a sufficient period to admit of the conclusion that there is a true "long cycle", of regular periodicity. The evidence on the building cycle (Cf. C. D. Long, "Long Cycles in the Building Industry", *Quarterly Journal of Economics*, Vol. 53, No. 3, pp. 371-404) does not establish a general economic trend, important as construction may be as a constituent of economic activity. In any case, Dr. Long's evi-

is associated with fairly regular, self-generated changes in the aggregative factors. That the long-run development goes in waves or cycles is an attractive fancy of the "naturalist" order of thought, for which, however, there is no reliable historical evidence.[14] The progress of the economy in terms of wealth depends on these major innovations of the "linked" variety. The movement from man power to horse, or ass or oxen power, to wind power, to water power, to steam power, to electric power, to oil power, to atomic power, evinces no cyclical regularity, no "natural beat", no wavelike motion, but a steadily accelerating rate of advance, as physical science has freed itself more and more of superstitious faith in preconceived ideas and required an ever increasing intellectual control over "nature".

We conclude that a linked advance is an important prime cause of economic change. It has effects to be studied under both the real and aggregative modes. Some of these effects are of a more or less permanent nature, that is, we believe the secular trend of the economy to be affected, and profoundly affected, by a linked advance. Other effects, directly and indirectly, are felt in comparatively short-run cyclical form, that is, in their aggregative aspects, linked advances are causally related to the ten-year cycle. Variations in the rate of introduction of particular innovations, on the other hand, we believe to be effects, not causes, of cyclical fluctuations—otherwise generated—in the rate of investment. In our present model we shall consider the long-run effects of a linked advance in the economy as a whole, the relation of such an advance to the ten-year cycle, and the distribution of the real gains of technical advance.

---

dence suggests not a secular trend but a 17-18 year cycle. Dr. Hansen points out (*Fiscal Policy and Business Cycles*, pp. 20-21) there are other fields of economic activity that reveal "a periodicity varying from that of the major cycle". He advances the "tentative explanation" that these special cycles are connected with the major cycle but evince a different periodicity because of time lags.

[14] Cf. A. H. Hansen, op. cit., "Attention should be called to the fact that it is scarcely appropriate, as has frequently been done, to speak of these prolonged periods as "upswing" or "downswing" periods. The long periods of hard times, at any rate, do not reveal an absolute decline in production, though they do reveal falling prices and interest rates. The trend of per capita output and real income continues to rise, though probably at a somewhat retarded rate" (p. 28).

2. The First Model

We assume that there are no changes in population and no independent changes on the side of money. In our first "approximation", we also assume perfect competition in all markets and a high degree of mobility of the factors. Because of the nature of the effects through the mode of aggregates, it matters a good deal whether we assume the economy to be in equilibrium at full employment or at some level of less than full employment. If we suppose the economy to be at less than full employment, we are faced with the further necessity of making some assumption about the direction in which the economy is moving along the cycle path, whether it is moving towards less employment, or towards more. Since such assumptions would introduce, at the very start of our analysis, seriously complicating "other causes" of a short-run or cyclical order, it is simpler at the present stage of inquiry, to make the classical assumption of full employment. At a later stage these cyclical aspects will be investigated.

When we say "let there now be an innovation of a major sort, such that it compels or necessitates the adoption of other innovations of wide adaptability"—what we have called a linked advance—, we have already implied a good deal about the nature, and effects, of the process. An invention is not an innovation until it has been adopted. Why was the original invention adopted? A major innovation, requiring large fixed capital outlay and some considerable delay before returns begin to come in, and necessitating the rapid obsolescence of old capital, will commend itself to entrepreneurs only under certain circumstances. If there is a rapidly expanding demand because of growing population or the opening of new markets, as in the eighteenth century, great pressure is generated for the invention of new, more rapid and cheaper processes. With a stable population, however, this pressure from the demand side is not apt to be great. Nor is there likely to be any intensive experimentation and research in a perfectly competitive economy, such as large-scale monopolistic firms frequently maintain. It might be that the innovation came during a war when

ordinary market motives did not prevail and obsolescence could be disregarded. Many innovations have come in this way. But this is not necessary to assume. In any competitive society, machine replacement becomes necessary in different plants at different times. One firm, or a group of firms, adopt an invention by way of replacement, because of the competitive advantage in costs and the expansion of output which it makes possible. In so doing they immediately render obsolete competitors' plants. It may, indeed, be true that it is hard in many instances to find the new capital disposal to replace comparatively new plant whose amortization charges have been based on an anticipated rate of depreciation and obsolescence now revolutionized. Such firms as cannot or do not make the effort, however, disappear from competition. Consider the fate of the once wealthy West Country woollen trade in England! It was the low capitalized West Riding industry that introduced steam-powered textile machinery, adapted from cotton, and the West Counties industry, with its heavy capitalization, now rendered obsolete in form, attempted to continue to compete without adopting the new methods. So the woollens of the world came to be made in Yorkshire.

The further innovations linked to the initial major innovation come by reason of its technical nature and are economically explicable in terms of the profit motive. To reduce fuel costs, condensers are needed, tighter fitting, more easily moving parts. Hence harder metals, more precise tooling of parts, better machine tools. Once initiated the advance is carried to its full technical limit by economic forces.

The real effects of a linked advance considered by themselves, are fairly obvious. New forms of capital are substituted for old, and, in individual employments, capital will be substituted for labour. One form of labour may be substituted for other forms and the society may be offered a choice between increased supplies of goods or increased leisure. The effects on material welfare as seen purely through the mode of real economics, are altogether favourable. With the same human resources the society can enjoy a higher level of real income. If the institutions of the economy are non-competitive, the welfare effects

may be minimal. This will depend on the degree of senility of monopolistic firms (see Chapter XI) and we must postpone discussion of the single firm's reaction to an innovation for the present. We may observe, however, that under oligopolist conditions, when firms sell to a kinked demand curve, the benefits of an innovation will not be passed on to the public in the form of increased output, unless the innovation reduces marginal cost so greatly as to move it below the sector of discontinuity of the marginal revenue curve.[15]

It is when we turn to the effects as seen through the mode of aggregates that confusion and difficulty are introduced. Long before economists had drawn particular attention to the question of aggregate effects, trades unions had, in practice, resisted innovations because of the feared effect on employment. The Lancashire weavers who burned the looms and the Welsh coal miners who for generations resisted mechanization of the mines did so because they saw that the new machines would cause widespread unemployment among their members. Economists, faced with this challenge, usually replied by admitting that such "technological unemployment" would exist, but argued that the demand for new capital goods and the increased income demand for other consumers' goods, would lead to a general expansion of employment and the rapid dissipation of any "spotty" technological unemployment in particular areas. They advocated various devices, such as labour exchanges, to increase the mobility of labour and to hasten the shift of labour from the areas of unemployment to the areas of increased employment. Always the economist had the last word in the confident assurance that everyone would be better off. This Panglossian philosophy, though difficult to rebut, failed to satisfy labour leaders, who, convinced there was a fallacy somewhere, continued to observe that innovations caused unemployment among their members. Professor Hansen has now lent the support of his authority to the heretical view.[16] Accord-

[15] Cf. O. Lange, "A Note on Innovations", *Review of Economic Statistics*, Vol. XXV, 1943.

[16] Cf. A. H. Hansen, *Fiscal Policy and Business Cycles* (New York, 1941). See also Testimony of Alvin H. Hansen, "Hearings before the Temporary National

ing to Professor Hansen, there are some innovations—capital-saving innovations—that reduce the volume of investment and hence of employment, and such innovations are becoming more and more numerous and so are contributing, along with the decline in population growth and the disappearance of the frontier, to a general secular stagnation.[17] This view is in contrast to that of Adam Smith[18] who seems to have taken it for granted that capital accumulation (investment) and technical progress were one and the same thing, and who believed that expanding employment (attended by increasing population) depended on inventiveness in the technical arts.

If we suppose the innovation to be introduced at the stage of full employment and to have been introduced because it offers increased profits, what can be inferred as to its aggregate effects? Clearly it may replace some labour in certain trades. Equally clearly it will require more labour in certain other trades, notably in the construction, engineering and machine-tool trades. The fact that the advance is "linked" means this process will be a cumulative one, and the fact that it is continually introduced throughout a wide industrial area implies that it promises increased profitability. It is because the rate of obsolescence is greatly accelerated that this process very definitely means increased investment and, on the usual assump-

---

Economic Committee", Part IX; Alvin H. Hansen and Guy Greer, "Towards Full Use of Our Resources", *Fortune* Magazine, November, 1942; A. H. Hansen, "Economic Progress and Declining Population Growth", *American Economic Review*, Vol. 29, No. 1, March, 1939.

[17] Professor Hansen does acknowledge the expansive power of strategic innovations. "This sort of transformation, involving vast expenditure of capital, can take place without extensive growth (i.e., of population), and under the progress of technology we shall doubtless experience again far-reaching revolutionary innovations" (*Fiscal Policy and Business Cycles*, p. 46). But he does not believe that such innovations are characteristic of the contemporary period, and holds that "now *extensive expansion* is largely over, and there remains only the possibility of intensive developments" (p. 45).

See also the argument of Ch. XVII.

[18] Professor Hansen is writing of a different economic period. When he speaks of the past he agrees in part with the views of Smith and thinks that the expansion Smith saw became even more pronounced in the nineteenth century. "It was a century of prodigious growth of population" (*Fiscal Policy and Business Cycles*, p. 361). Professor Hansen's emphasis on the frontier and on population, an emphasis first expressed in the article in the *American Economic Review*, March, 1939, is here apparent.

tion that the rate of employment is a function of the rate of investment, increased opportunities for employment. The improved marginal productivity of labour promises enhanced incomes, which may also be regarded as stimulating rather than retarding economic activity. A rich society, of course, may elect to take the real income improvement in the form of increased leisure. It is only as the process of adaption has spread about as widely as the technical limits permit that entrepreneurial expectations decline. Then, if there is not another new advance initiated, the cyclical decline from increased propensities to save and declining profit expectations may set in.

Thus, on the approximation of this first model, we see the Smithian rather than the Hansenian view sustained. A linked innovation does not generate the business cycle; it generates an advance in welfare and stimulates economic activity. It is only when its stimulating effects are worn thin that the ordinary cyclical causes are able to preponderate and bring about a decline in activity, a decline, be it noted, at a higher level of real welfare. It is, of course, the selection of the model and the type of innovation that gives us this result. An appeal to historical evidence supports the view. The introduction of steam power in England generated a forty-year advance with relatively minor cyclical declines, most of them associated with the vicissitudes of the Napoleonic wars or bungling in monetary policy. The innovation of the internal combustion engine in the United States initiated a similar period in that country, which only came to an end with the depression of the 1930's.

3. The Second Model

Under what circumstances might the aggregative effects of innovations, work in the opposite direction? First of all, we must assume that the innovations are not of the linked variety. It is apparent, from definition, and because of the effect on the rate of obsolescence, that an innovation in a prime mover linked to a general technical advance, must generate for some period an increased volume of investment. The failure of new major innovations to develop, innovations which would be linked to a further advance, might then lead to a period, follow-

ing the period of investment to replace obsolete installations, in which the rate of investment might actually decline to a level lower than would have obtained had replacement continued at a normal rate. Aggregative effects of stable or declining populations and the failure to develop new investment outlets in new areas and new forms of business undertaking will combine to produce the effect of economic stagnation. There was an important difference, noted by Professor Henderson,[19] between overseas investments in the nineteenth century when investment was complementary in character, that is it opened up new resources and sources of supply complementary to the mature economy of the metropolitan investing power, and overseas investment in the present century which has become more and more competitive. This is close to what Professor Hansen has in mind when he speaks of "extensive" and "intensive" investment. Under these conditions investment in new techniques will continue but there will be a cycle in their adoption and a general secular decline. It seems unreasonable to assume, though, that cyclical fluctuations in such innovations are the causes of cyclical fluctuations in the general level of activity. On the contrary, they would seem to be effects. We must remember that though inventions might appear in a whimsical and accidental fashion—there is absolutely no reason to suppose a regular cycle in inventiveness—they do not become innovations until they are accepted and introduced by a business enterprise. These economic motives and the state of business expectations govern the rate of such innovations as are not of the "linked" variety. Even those are adopted and adapted according to rational laws of economic advantage, but, whereas they have a definite secular effect, it is our contention that innovations of limited adaptability do not. The cyclical trend in such innovations, discernible during periods when no new linked advance is in progress, is surely the effect, not the cause, of economic fluctuations.

---

[19] Sir Hubert Henderson: various contributions to the discussion on Raw Materials and Colonies at the International Studies Conference, Paris, 1937, as reported in *Peaceful Change*, ed. Maurice Bourquin, International Institute of Intellectual Cooperation, League of Nations, Paris, 1938.

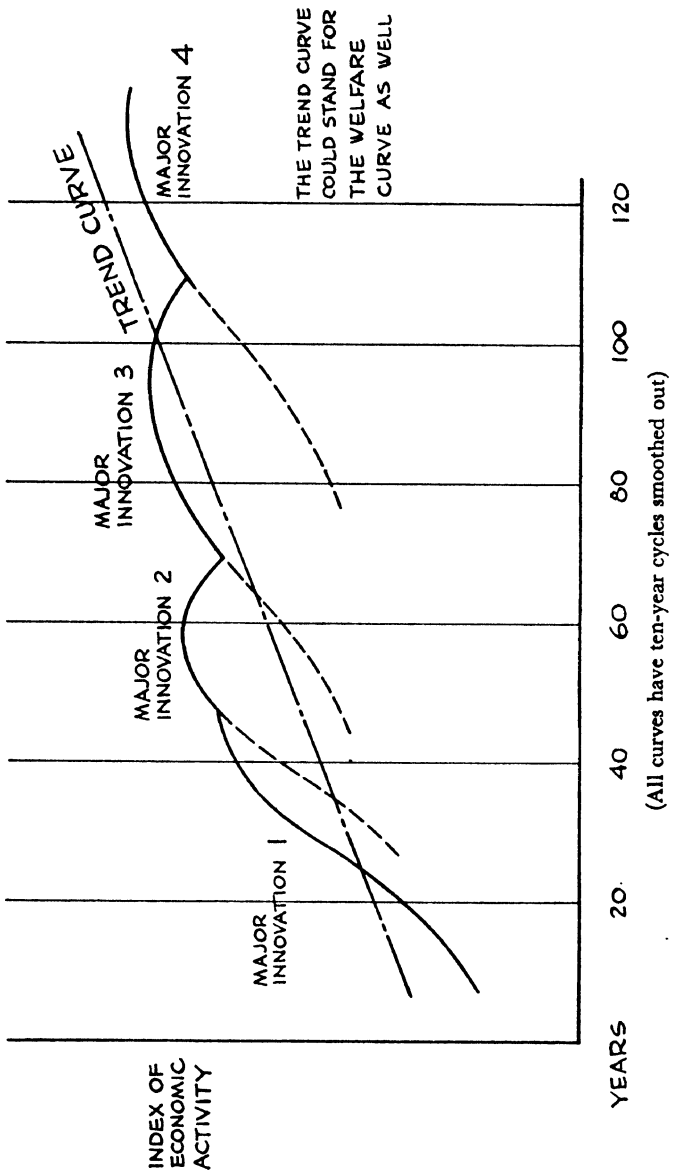
For suppose the economy to be approaching the top of the cycle with a high level of employment and good profit expectations. Entrepreneurs are still ready to expand investment and will welcome cost-reducing or output-expanding inventions. The rate of innovation will be closely linked with the rate of investment. Suppose, however, expectations to decline, with a declining marginal efficiency of capital, a high propensity to save and a securities' market bearishness making for increased liquidity preference. Then the rate of new investment will decline and entrepreneurs will be chary of taking up inventions, except such as offer great cost reductions at a small increase in fixed costs. As the general decline or depression continues and obsolescent and fully depreciated capital wears out, some general spurt in innovation may appear as the old capital is replaced by cost-reducing machinery of improved design. This is the sort of process, surely, Professor Hansen has in mind, when he speaks of "capital-saving" innovations. That they occur is not in dispute; but that they are causally related to a secular decline in economic activity—or stagnation—must appear more than doubtful. This phenomenon is the result of the comparatively short cycle in activity and investment. The type of innovation causally related to the long-run or secular trend is of the major linked variety, and its only possible aggregative effect is to act as a stimulus to investment, while its real effect is to create an improvement in welfare.

4. A Summary

Our argument may be summarized somewhat as follows.

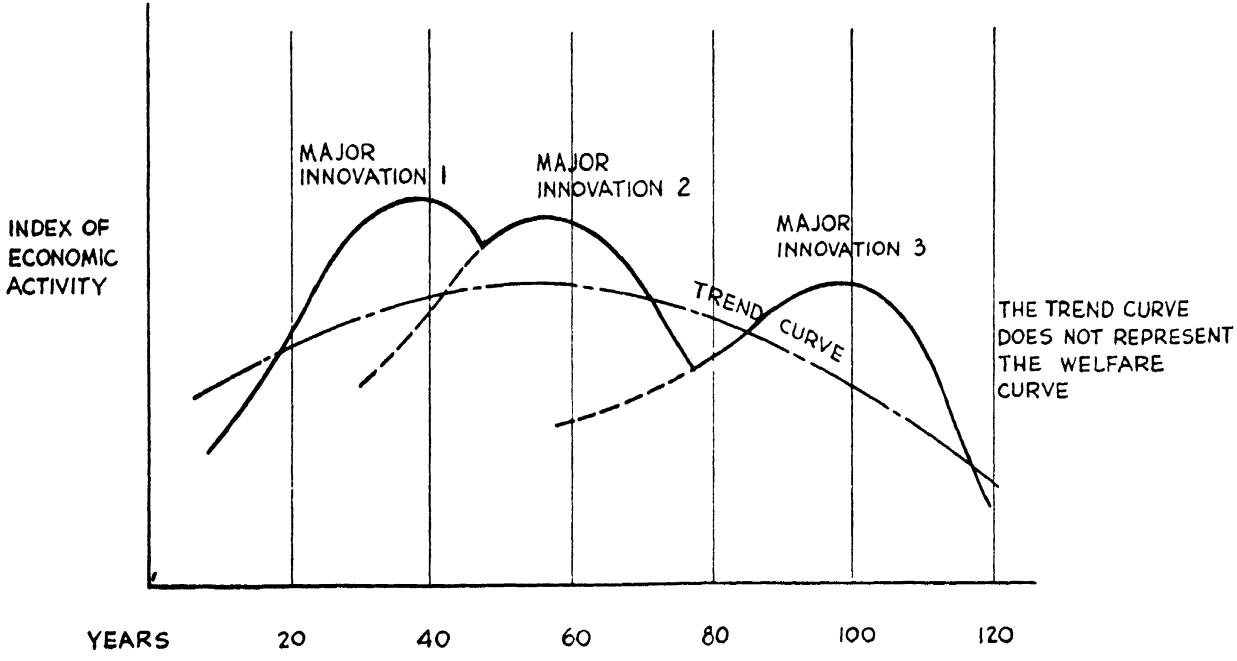
Linked advances occur from time to time, but are not cyclical in nature. Their effect on the general equilibrium of the system is to stimulate investment and employment, and their long-run real effects are an enhancement of welfare. If these major innovations, say in prime movers, follow one another sufficiently closely they may generate a long period of general economic advance, with only short recessions from the disequilibrating forces associated with the ten-year cycle. The movement could

be illustrated something like this:



(All curves have ten-year cycles smoothed out)

FIG. 4



(All curves have ten-year cycles smoothed out)

FIG. 5

If there are long gaps between the major innovations and the linked "bunches" associated with them, there could be periods of secular decline, with increased duration and severity of the depression periods of the ten-year cycle. The periods of secular decline will result from the failure to develop any innovation of wide adaptability requiring a large aggregate of new investment. The process would then be illustrated as in Fig. 5.

Innovations associated with the short cycle are not of the major or linked type and short-run fluctuations in the rate of such innovations are the effect not the cause of the ten-year cycle, which we take to be the result of purely aggregative forces disturbing the general equilibrium of income, saving, investment and employment.

There is, of course, implicit in this argument, a *conjunctural* relation between changes in the long-run trend and the ten-year cycle. A failure of a new innovation to appear in time to maintain the upward movement of the long trend may be in conjuncture with the downward swing of the ten-year cycle. In such cases the consequent recession would be long and severe. This conjuncture would look as in Fig. 6.

5. Institutional Change and Welfare Effects

We turn now to the institutional and welfare effects of a linked advance. In Part IV we shall show analytically how such an advance affects the position of the individual firm. In this section we shall content ourselves with a summary, without rigorous demonstration, of the now familiar argument that technical advance renders competition unstable and the market imperfect. In the first place, as we shall show in Part IV, the increase in the proportion of fixed to variable costs results in decreasing unit costs over time and, consequently, unstable equilibrium.

Each firm sees advantages in increasing output, and if all the firms throw increased outputs on the market so as to reduce their costs, they will so flood the market that they will begin to suffer losses. No curtailment of output, which means sacrificing the cost advantages of the advance, can be contemplated by a

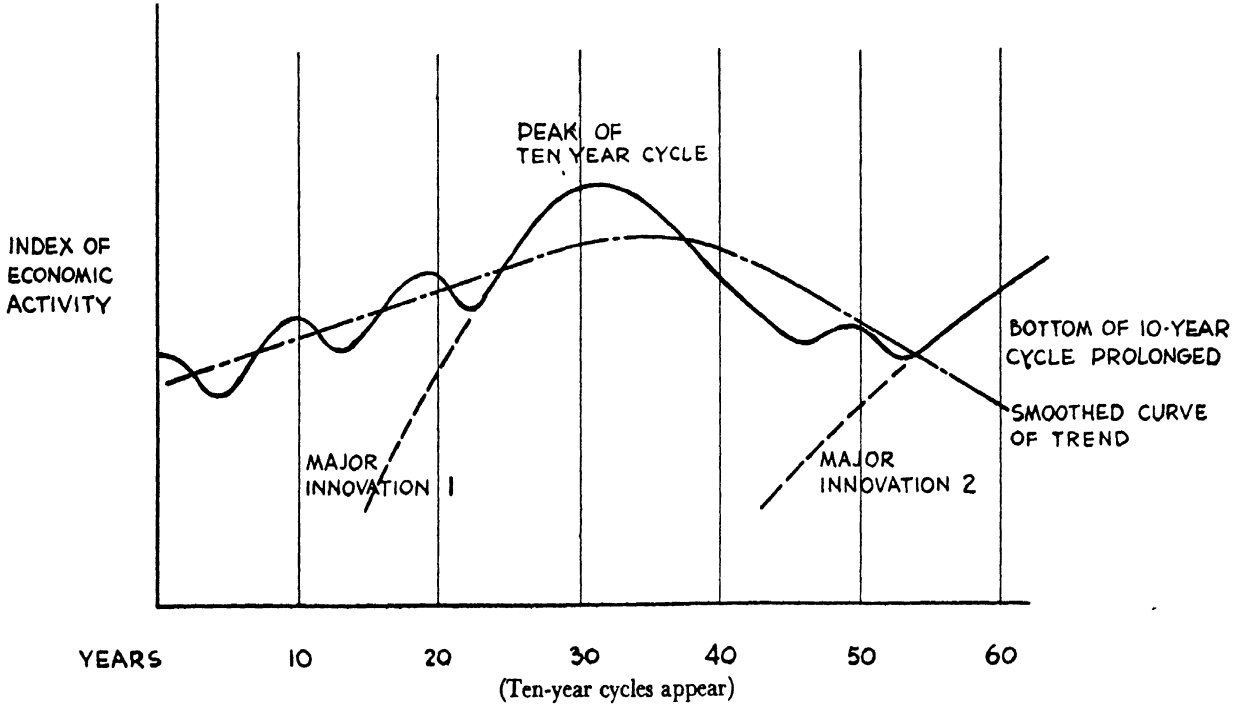

INDEX OF ECONOMIC ACTIVITY
PEAK OF TEN YEAR CYCLE
BOTTOM OF 10-YEAR CYCLE PROLONGED
SMOOTHED CURVE OF TREND
MAJOR INNOVATION 1
MAJOR INNOVATION 2
YEARS
10
20
30
40
50
60
(Ten-year cycles appear)

FIG. 6

single firm behaving *independently*; it can only press ahead, increase its output and, if possible, its share of the market, reduce its costs and try both technically, and by advertising, to surpass its competitors. Only when the technical advance comes to a definite halt, so that no further changes can be affected in time, can increases in output be regarded as "lateral", that is, as point increases along the *X*-axis with corresponding increases in unit costs, which might bring a return to competitive stability. As long as the technical advance continues the productive process makes for unstable competition. This implies (a) that weaker firms will be crushed out through bankruptcy, and (b) that a very strong incentive towards the creation of a pooling agreement, explicit or tacit, will be created. The tendency towards trustification which we observe in the later stages of the industrial-technical revolution is consequently no historic accident, nor a purely financial phenomenon; it is a direct result of the nature of continuous technical change.

Technical advances, and the internal economies they embody, are frequently only possible to the firm able to command large accumulations of capital. The second way, therefore, in which a technical advance corrupts competition is through the incentive it creates towards financial integration, the concentration of ownership, within the industry. In the pulp and paper industry in Canada the period of great technical advance—the period from 1920 to 1928—witnessed a reduction in the number of small establishments, growth in the size of the typical establishment, growth of the "optimum size" of establishment,[20]

---

[20] This is a concept we have not previously introduced and should be defined. It does not, of course, correspond to the optimum production of any given firm, a notion which can be given statistical reality only by an accountant with access to the costing and production sheets of the firm. But in most industries it appears that any index of efficiency shows a definite positive correlation with size of establishment up to a certain point. Beyond that point further increases in the size of the establishment do not appear to be accompanied by any further economies. This general proposition is supported by the findings of Professor E. A. G. Robinson (*Structure of Competitive Industry*, London, 1931), Professor Sargant Florence ("Economic Research and Industrial Policy", in *Economic Journal*, Vol. XLVII, No. 188, Dec., 1937), and Mr. H. C. Hillman ("Size of Formation in the Boot and Shoe Industry", in *Economic Journal*, Vol. XLIX, No. 194, June, 1939), and my *Economic Effects of the War on the Maritime Provinces of Canada* (Halifax, 1943) Ch. III.

and the integration of the industry by the development of the great combines.[21] The following table indicates the trend in a section of the Canadian pulp and paper industry.

PULP AND PAPER INDUSTRY

*Quebec and the Maritime Provinces**

	No. of establishments	(size of establishment measured in thousand-ton capacity) *Median size of establishment*		*Optimum Size of establishment*
		Quebec	*Maritime Provinces*	*(Quebec)*
1920	57	19.5	10	125†
1939	57	62.5	37.5	200

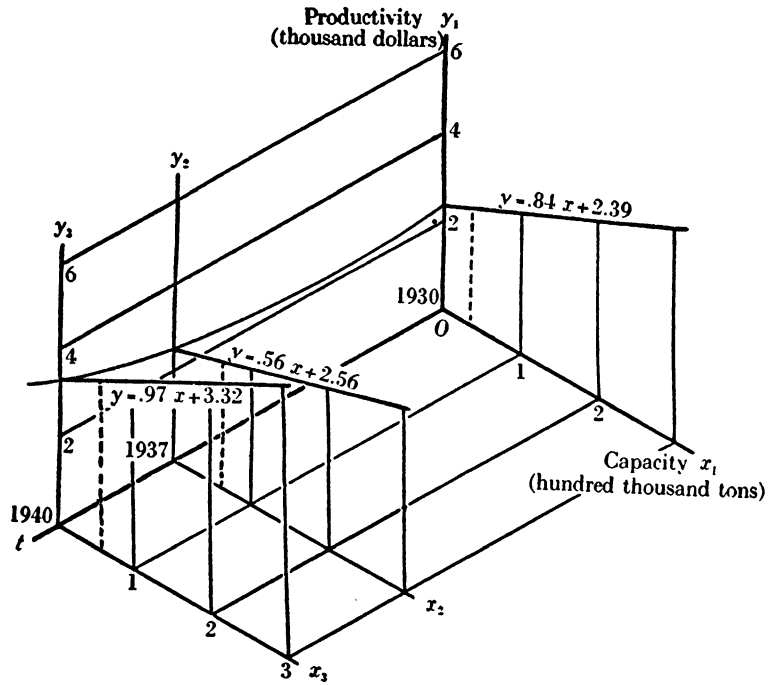
*The data from which this table is compiled are obtained from my *Economic Effects of the War on the Maritime Provinces of Canada.* For permission to use this material I have to thank the Institute of Public Affairs, Dalhousie University, under whose direction the work was done, and the Dominion Bureau of Statistics, whose co-operation has made the work possible.

†This figure estimated, as calculations are not complete.

In Figure 7 the pulp and paper industry in Quebec is pictured. In this case, where the actual measurement of firms' costs is impossible, the index of efficiency, which only roughly indicates the relative optimum cost points of the different firms, is the net value productivity of labour. Size of establishment is measured by rated tonnage capacity of the plants. It can be seen here how, over time, the efficiency of all establishments increased with the technical advance, the smaller establishments tended to disappear, the size of the median establishment grew greater and the differential advantage of the larger

[21] Admittedly it was also a period when there was a great expansion of demand for newsprint. American mills could not supply the growing demand and were also facing a depletion of their forest reserves. American and Canadian companies built new mills with larger, faster machines that could tap the more cheaply accessible Canadian forests. The fact of the positive shift in demand does not impose modification of our argument. With decreasing costs competition would be unstable—as it definitely was, *vide* the series of rebellions against the price leadership of Canadian International—regardless of increased demand, and every incentive would be given to mergers and amalgamations to maintain price. Smaller establishments and firms were swallowed up by the growing trusts. One should not, perhaps, ignore the interest in this process of the stock promoter.

establishments increased. During this period there was, as is well known,[22] a concentration of ownership within the industry,



Courtesy "Canadian Journal of Economics and Political Science" and The University of Toronto Press

FIG. 7

Pulp and Paper Industry in Quebec

Net productivity of labour (y) against size of establishment (x) in the years 1930, 1937 and 1940. Shift in the position of the median establishment is shown by the dotted perpendiculars dropped to the x-axes.

and the general expansion in industrial capacity, was accompanied by a tendency towards imperfect competition in the market. Under the stimulus of a technical advance firms tend to grow larger, total output is increased, but in the industry smaller firms tend to disappear either from bankruptcy or, more likely, by merger,[23] the optimum production per plant grows larger,

[22] Cf. J. A. Guthrie, *The Newsprint Paper Industry: An Economic Analysis* (Cambridge, Mass., 1941).

[23] Cf., Investigation of Concentration of Economic Power by the Temporary National Economic Committee of the 76th Congress (U.S.A.), Monograph 13, *The Relative Efficiency of Large, Small and Medium-Sized Businesses.*

and the industrial tendency is towards non-perfect competition and ownership integration.

Our examination, and general theoretic considerations, would lead us to expect that the gains of a technical advance would be socially unevenly distributed. We should expect the gains to be passed on in part to the consumer in the form of greater output, in part to skilled labour (especially if highly organized, because it is in inelastic supply and, if unionized, in a strong bargaining position), to the owners of certain other factors in inelastic supply not socially important, and, for the rest, to be retained by the owners of the firms that were able to take advantage of the advance. We should expect few of the gains to be passed on to the great mass of labour (except in the capacity of consumers) in elastic supply, and none to the owners and managers of establishments unable either to take advantage of the technical advance or to sustain and endure the period of unstable competition it might initiate. The argument that factors in inelastic supply would benefit from increased remuneration and factors in elastic supply would not is too well known to require review here.[24] That consumers as a whole would gain in their capacity as consumers is indicated by the fact that increased quantities of the commodity, or commodities, affected by the technical advance would be thrown on the market, and on the assumption that there has been no change in demand, this would mean a reduced selling price. A positive shift in demand would increase the quantities made available to consumers, but might reduce the price benefit to consumers. A negative shift would tend to offset the benefit passed on to consumers. If the shift in demand is unconnected with the technical advance and would in any case have occurred, the consuming public will in the event be better off, despite the effects of the shift in demand, than if the technical advance had not occurred. If the industry had formerly been perfectly competitive and, as a result of the technical advance, becomes imperfectly competitive or monopolistic, the gains, to the consumer may not materialize and will, in any case, be less than

[24] Cf. for example, Paul Douglas, *The Theory of Wages* (New York, 1934).

they might have been because of the incentive of the imperfect market to artificial restriction of output.[25] Thus, because the process of technical advance is, under our institutions, bound up with the tendency to trustification, an institutional organization which gives an incentive to restrict output, the full benefits of technical progress are never passed on to the community as a whole.[26]

The profits of the firms able to adopt the new process will be increased from one or both of two causes. Altogether apart from the general effects of the technical advance on the industry as a whole, the profit area of the firm, as we shall show—Part IV—is increased as the advance continues. Net profits are a function of the time process and, for the firms able to advance, increase in time. Moreover, the relative bargaining position of the firms able to take advantage of the advance may be improved if weaker competitors are forced out, or the dominant firms may improve their position on the market by a pooling arrangement. Anything which increases the degree of monopolistic bargaining improves the ability of the firms who survive to earn profits. We should consequently expect the benefits of a technical advance to accrue to the stronger firms in the form of increased profits, while weaker firms would gain few, if any, benefits, if indeed they were able to survive.

Dr. Spurgeon Bell,[27] approaching this problem with statistical techniques and from the point of view of aggregates, obtains findings for the United States which in the main support the theoretic expectations. He finds that, in the major industries of the United States, excluding agriculture, the technical advances of the twenty years between 1918 and 1937 produced benefits which, in the aggregate, were passed on to the consumer in the form of more goods and services at cheaper prices. He finds that, whereas wage rates, particularly in certain crafts and skilled

---

[25] Compare our earlier comment on the retention of the gains of innovation by firms in monopolistic competition with kinked average revenue curves.

[26] Cf. the statistical evidence and analytical argument on this point in M. Kalecki, "The Distribution of the National Income" in *Essays in the Theory of Economic Fluctuations* (London, 1939).

[27] Spurgeon Bell, *Productivity, Wages, and National Income* (Washington, Brookings Institution, 1940).

types of employment, went up, the total wage bill did not show any increases of significant magnitude. Similarly he finds that total dividend payments did not increase by any considerable amount. Thus, in terms of broad social classes, neither labour nor capital succeeded in holding any gains, but individual workers and individual recipients of dividends did gain in their capacity as consumers. However, if one breaks down Dr. Bell's classes, which are conglomerate and undifferentiated, it would appear that if certain grades of workers did gain in increased wages, and if no gains were registered by the workers as a whole, then certain grades—those employed in depressed industries, and in many normal industries the mass of unskilled and semi-skilled workers—must have lost, there being no significant change in the volume of employment. Similarly we know that many small capitalists disappeared during the period, consequently the profit-making capacity of those who remained must have increased. Thus it would appear that, under our present institutions, the process of technical advance has the social effect of improving the position of the smaller groups of labour and capital who already enjoy relatively strong positions and of worsening the position of the unskilled workers and the small, independent capitalist. The gist of the theory of technical advance seems to have been contained in the pessimistic aphorism of Holy Writ—"unto every one which hath shall be given; and from him that hath not, even that which he hath shall be taken away from him". The technical process would seem, then, to intensify class division and the concentration of wealth and power, and it would seem to give to the great trusts and their owners and to the unions of skilled workers incentives to retain the present social and property institutions, and to create two classes with real discontent, the mass of unskilled workers and the small-scale, dispossessed owner capitalist. The political conservatism, therefore, not only of such organizations as the manufacturers' associations, but also of old-line unionism such as the American Federation of Labour, and the political unrest and social discontent of the masses or workers and of the petty bourgeoisie are both explicable in terms of the long-run technical process. The bitter social strife of our times has a real source,

not in the static fact of the class society, but in the historical process which is intensifying the conflict of interest in that society, dividing the group interest of those small groups who hold the weight of political power from the interest of society as a whole, and preventing the formulation of national policy by disinterested statesmen capable of thinking in terms of common welfare.

6. Relation to Short-Run Cycles

Not only are the benefits unevenly dispersed, but the technical process involves costs of a social nature, which may bear unevenly on different social groups. The long-run or secular trend is towards increased production at lower entrepreneurs' costs. But the secular trend is an uneven course involving, as we have admitted, cyclical fluctuations. The depression periods of the cycle, which are definitely related to the technical process, bring great hardships which we might think of as the social costs of the advance, and these costs fall with unequal incidence on different sections of the community. I wish now to sketch the relation between the theory of the technical advance and cycle theory and then to indicate briefly the incidence of the social costs of the process.

Equilibrium in the capital market at any moment of time is effected by the instantaneous rate of interest which equates the demand for funds with the available supply. The demand for loans will depend on the current anticipations of the net marginal value productivity of new capital and the supply of loanable funds at the current liquidity preference schedule. It is, of course, true that the rate of interest must adjust the total demand for money with the quantity of money available at any one time because all the money in existence must be owned by someone. But the rate of interest which satisfies the desires of the people, taken together, to hold money must, by the same token, satisfy the demand for loans. If, at the prevailing rate of interest, entrepreneurs are willing to offer more for the quantity of capital disposal available to them for investment they will bid up the price of loans, forcing a new adjustment

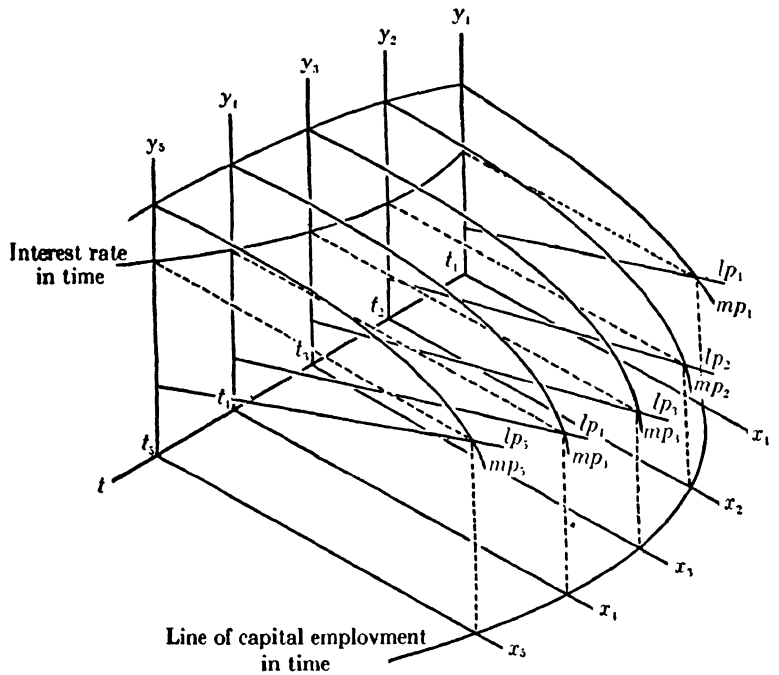
of the rate of interest. But both the marginal productivity schedule and the liquidity preference schedule are static; they are potential at a moment of time.[28] The equilibrium instantaneous rate of interest is thus, in our language, a potential price. We are here interested in the price in process under the stimulus of a technical advance.

Many things will affect the liquidity preference schedule over time and many things will disturb equilibrium of the capital market. The desire to hold money rather than securities will be primarily affected by the psychology of the stock market, and if people as a whole believe stock prices will fall they will want to hold more money and fewer securities. If the prevalent expectation is that prices will rise they will prefer to hold securities. The prevalent attitude towards the market is a product of so many forces, many of them subjective, defying analysis, unpredictable, that they seem beyond current techniques of economic analysis. It will be necessary, for the purpose of studying the effect of technical advance on the supply of and demand for loanable funds, to suppose no *independent changes* in market psychology to occur.

A technical advance will affect the supply of loanable funds through its aggregative effect on the national income, and the demand for loanable funds through its effect on the anticipation of marginal value productivity in process. Anything which increases the total of the national income, or which alters the structure of the national income in favour of the larger incomes will, *ceteris paribus*, increase the volume of savings and the willingness and ability of people to lend funds to industry and commerce. Anything which decreases the national income will have the opposite effect. Marginal productivity in process is to be regarded as the rate of increase in total produce which may be imputed to a single agent as its employment is increased over a period of time. The effect of a technical advance, we have seen, will be to increase the employment of capital over

[28] Dr. H. G. Littler has argued ("*The Equation of Money Saving*", as yet unpublished) that these concepts are static concepts in Keynesian analysis, but, properly considered as monetary phenomena, they are essentially temporal. If I understand him aright he believes that the cycle can only be understood in terms of the temporal maladjustment of these and other monetary rates.

time under conditions which give a rising productivity with increases in the employment of the factor as one moves outward in time. In Figure 8 the employment of capital is increased along a line that runs outward with the T-axis and, of course, to the right along the X-axis. Productivity in process at first



Courtesy "Canadian Journal of Economics and Political Science" and The University of Toronto Press

FIG. 8

increases because of the economies of the technical advance, but after a certain employment of capital, if the technical advance does not continue at an augmented pace, a point will be reached where the decline in potential productivity offsets the increase from the technical advance, and realized productivity in process then declines. Consequently the long-term trend in the demand for capital is apt to be cyclical in nature, increasing with each fresh spurt in the pace of the technical advance, then decreasing under the effect of diminishing marginal productivity in process, thus constituting a period

during which the level of investment can only be sustained by lowered interest rates.

During the period of linked technical advance, when entrepreneurs are increasing investment, employment will be increasing and the national income growing. If the technical advance is too limited to affect employment favourably it will at any rate, as we have seen, affect the national income in favour of greater concentration. In either case there will result a greater volume of saving to support the increased investment and, in terms of probability, liquidity preference will diminish, so that loans will be offered at reduced interest rates.[29] This means that equilibrium is shifting towards an ever-increasing use of capital at falling rates. In Figure 8 we have pictured the process over a period of time, with five time-points selected to show the movement of the equilibrium position. The impact of the technical advance on the curved surface of marginal net value productivity is indicated by the rising surface as one proceeds outwards along the T-axis from the origin. From time 1 to time 3 (t_1 to t_3) the lift in the surface is considerable under the stimulus of a spurt in the pace of the advance. From time 3 to time 5 (t_3 to t_5) the lift is negligible as the pace of the advance slows down. In the first time interval the liquidity preference schedule will not continue to fall. The continuation of the advance during the second temporal interval maintains the trend, but between time 3 and time 4, the rate which entrepreneurs are willing to pay to sustain the rate of investment falls more rapidly than the fall in the liquidity preference schedule. There is thus a curtailment at t_4 of capital employment. This has the double effect of curtailing the expansion of savings in the aggregate and of affecting the psychology of the market in favour of increased desire for liquidity. Thus in the fourth time interval the liquidity preference schedule rises abruptly, at a time when entrepreneurs require a falling rate to sustain the level of investment and capital employment. The new equilibrium adjustment is for a reduced capital investment. In Figure 8 we observe that the equilibrium employment of

[29] This is at the beginning of the boom. Rising interest rates, generally associated with expansion, appear later when the boom is established.

capital increases from time 1 to time 3 while the interest rate falls. From time 3 to time 4 the increase in the employment of capital is checked and the rate of interest shows a tendency to rise. From time 4 to time 5 the interest rate rises and the employment of capital is reduced. This reduction would be even more marked if we extended the period to some future time 6, and would continue until the stimulus of a fresh spurt in the rate of technical advance.

Thus the cycle is not unconnected with the long-run trend of technical advance. Our technical process is one which, in the long run, confers certain benefits upon the society as a whole, but it confers these benefits unequally. In its process it involves employment at various levels through shifts in the equilibrium of investment, shifts which follow a cyclical pattern. These shifts, as we know, involve certain social costs. We need not here evaluate the incidence of the cost of depression, but such an evaluation would not improbably show that the real burden falls most heavily on two classes, namely, the classes of labour in elastic supply, from reduced employment, and the small-scale capitalists, from bankruptcy. One of these classes, at any rate, exists on such a meagre standard of living at the best of times that any restriction in that standard brings real suffering, and both classes are those that benefit the least from the long-run effects of the technical process. The implications of these conclusions for social policy are fairly evident. Social justice is a concept extraordinarily difficult to define, but it is hard to imagine any definition that would be accepted in a democratic society which did not insist that the benefits of science and invention be given to the community as a whole and that any of the pains of material progress that appropriate action fails to eliminate should not be permitted to fall, without relief, on those least able to support them.

CHAPTER IX

POPULATION AND INNOVATIONS

1. SECULAR STAGNATION

Professor Alvin Hansen's thesis of secular stagnation involves three causes for the long-term trend towards decline in economic activity and the increased intensity and prolongation of depressions. These three causes are the decline in the rate of population growth, the tendency of new innovations to require "intensive investment", and the disappearance of the moving frontier. The thesis does not depend on a logical or necessary relation among these causes, they are regarded each as sufficient in itself, and as all operating in a conjunctural association in the same direction.[1] Thus, for example, to attack the capital-saving concept is not seriously to dispute the general thesis. Indeed, it is doubtful how much stock Professor Hansen himself puts in this notion. Chiefly his thesis depends on the aggregative effects of a decline in population growth and the disappearance of the moving frontier.[2] Now the "moving frontier" in our

[1] "The constituent elements of economic progress are: (a) inventions, (b) the discovery and development of new territory and new resources, and (c) population. Each of these in turn severally and in combination, has opened investment outlets and caused a rapid growth of capital formation" (A. H. Hansen, *Fiscal Policy and Business Cycles*, p. 352).

[2] "One overwhelmingly important fact characterized the century which preceded the first World War. It was a unique epoch. It was a century of rapid expansion into new territory and it was a century of prodigious growth of population. . . . This one central fact of growth and expansion dominated the whole of economic life. *It minimized the risk of new ventures.*" [Italics mine] "If optimism carried railroad building too far at the moment, . . . expansion and growth soon made good the error. Business now could look far into the future with gigantic plans, with anticipatory capital outlays, with investment plans which had no relation to the present, and which were based upon the expectation of growth and expansion" (*Fiscal Policy and Business Cycles*, p. 347).

Professor Hansen believes that of the three causes of economic progress ("inventions, the discovery and development of new territory and new resources, and the growth of population") innovations can only lead to a saturation of new capital equipment with a consequent cessation of investment unless growth and territorial expansion "make good the error". Hence the secular stagnation of the present period.

language becomes simply an innovation. It consists, in its economic essence, in discovery and economic introduction of new factors of production and new materials. Its effects are, taken by themselves, the effects of any major innovation of wide adaptability.[3] That the frontier has ceased to move is not a fact of American geography: it is an economic fact. Innovations of this particular sort have, for one reason or another, been less attractive. In the purely geographic sense of unexplored and unsettled land the North American economy still has frontiers (potential new resources and materials) in many parts of the continental United States, as the Tennessee Valley Authority has shown, in Alaska, in the Canadian North and in South and Central America. The European economy has such "frontiers" in parts of southern and central Europe and, more obviously, in Africa. Institutional structure and economic reasons have prevented these geographic frontiers from "moving" in Professor Hansen's economic sense. They are like inventions that do not become innovations.[4] Under the stimulus of war, innovations developed in the way of moving into the Canadian

[3] Professor Schumpeter also includes "the conquest of a new source of supply of raw materials or half-manufactured goods" among this "new combinations" or innovations. See his *Theory of Economic Development*, p. 60.

[4] "The purely geographic phenomena of history are essentially spatial, and for this reason they are not only deterministic in their bearings, but essentially unhistorical. The technological problems with which economic history is concerned present a sharp antithesis to these geographic factors. Changes in technique involve series of individual innovations that are finally embodied in practical accomplishments. These series or sequences of relatively independent inventions are among the most intense manifestations of the dymanic processes of history" (A. P. Usher, *A History of Mechanical Inventions*, New York, 1929, p. 4). In our view the "spatial, unhistorical" facts of geography stand in this antithesis to economic innovation, but the economic utilization of new resources, including the peopling and cultivation of new lands, is an historical and dynamic process of exactly the same character as other innovations. Professor Usher implies a recognition of this when he includes (p.15) explorers with scientists and inventors among the "true heroes of economic history". It is only proper to state, however, that Professor Usher, in his introductory chapter suggests a distinction between geographic expansion and mechanical invention from which our argument departs. We must admit that geographic fact is a limiting and, in fact, a determining factor in economic, and social, development. It is part of the physical element in economic causation. The purposeful, human response, however, which is also an element in economic causation, is economic and dynamic and, at least in its economic aspects, is essentially an innovation, the working of new combinations of the productive factors and the introduction of new resources, new materials and new goods.

North for new materials and factors. The "moving" or "static frontier" is not a given cause of economic change, external to the economic order, it is a rate of innovation, dependent on economic forces. This point Ricardo saw long since.

Thus we believe that the moving frontier, in the economic sense, is simply one form of innovation. Innovations are economic facts, with economic causes, and even invention, depending, as it does, on the economic direction of research, is in part, at any rate, a result of economic causes. The rate of population growth is also governed by economic events, though here we have to observe that there is a time-lag in the adjustment of the rate of growth to economic changes, and that this time-lag is occasioned by the effect on future births of the trend of the past, through its effect on the age structure of the population. Further we have to allow for the qualifying influence of deliberate contraception and the various human motives which lie behind this practice. Our point, however, emerges sufficiently clearly. The frontier, the rate of innovation, the rate of population growth, are not exogenous causes, external to the economic order, but are internal to that order. They are a part of it, mutually interdependent, and a theory of economic change must take account of this mutual interdependence. They cannot be treated as exogenous causes, but as within that complicated pattern of physical stimulus-motive-response-institutional effect which is the chain of economic causation.

The case of increasing population and expansion of a growing economy with frequent major innovations is probably neither complicated nor obscure. The aggregative effects of expanding population are expansionary and set up a strong incentive for innovations. Thus innovations take the form of developing new territories and resources with the surplus population, as well as the form of mechanical processes. Except for the comparatively minor vicissitudes of the ten-year cycle the secular trend is increased activity and increased real welfare.

Thus population increase in the eighteenth century was made possible by the improved standard of living which in turn came from better communications, the expansion of the west European frontier beyond the seas, the discovery of new goods, the more

economic allocation of resources and the development of science. The increased population, the movement of population in response to economic motives to overseas colonies and the discovery of new markets led, in turn, to pressure for new processes and a most rapid increase in the rate of innovations, involving the introduction of steam power, a major innovation generating a linked advance. Population responded to this technological advance by rapid growth and increased mobility, moving rapidly to the areas where the richest new resources were being discovered, and so pushing back the frontiers and further stimulating the rate of innovations.

Contraception, an innovation in consumers' goods, cut across this population movement, reducing the birth rate faster than the death rate continued to fall, and so reducing the rate of natural increase. That rate then, for the first time in human history, became directly subject to purposive control and thus became related to the nature and use of social institutions. The causal dependence on economic conditions, however, remained, though in a modified form. Population continued to move, wherever it was free to do so, in response to the relative attractiveness of various labour markets, and the average age of marriage (reflected after a time-lag by birth rate) correlated in the short run with economic fluctuations, the average age declining in good times and increasing in times of depression.

The development of the Canadian economy in the period 1901-1914 offers an attractive case study of an economy expanding under the threefold impact of innovations with capital formation, population increase and expansion of the frontier. During this period railway construction and extension was continued, the grain elevators were built and the beginnings of hydro-electric power generation laid down. Though Canada, at that time, failed to develop a motor car industry of her own, the impact of the innovation of the internal combustion engine was felt in the Canadian economy in various ways—demands for motor roads, provision of motor fuel, automobile repair, servicing and sales promotion. Population increased from 5,371,315 in 1901 to 7,206,643 in 1911 and to an estimated 7,632,000 in 1913. Immigration ran as high as 300,000 in some

years, and though much of this was offset by heavy emigration to the United States, the net migration gains more than doubled the natural increase. The prairies were opened up, settled and brought under the plough. Average productivity per worker in manufacturing industries increased from $1,418.30 in 1900 to $2,263.10 in 1910, and this increase should be set against an increase in capital employed per worker from $1,317.70 in 1900 to $2,445.20 in 1910.[5]

The expansion of the Canadian economy under these stimuli is attested by all the indices of activity. Yet the welfare improvement is problematic. The highly increased productivity of the Canadian worker should have raised the real welfare level, but the calculations of the Royal Commission on the Standard of Living[6] indicate that improved money wages were quite offset by increased prices. In fact a generally inflationary situation existed with full, or "over", employment. Expansion, in itself, did not necessarily mean improvement in welfare. Mr. Bell[7] explains this by arguing (a) that the new capital formation was based on capital imports and that much of the increased physical production was in the form of capital goods and much also was exported to service the external debt; (b) that there was a considerable displacement of non-wage farm labour by wage paid labour in industry and in the commercial cash-crop agriculture of the west, so that money wages deflated by a cost-of-living index, was an inadequate measure of the standard of living, and (c) that there was an increased share of the national dividend paid to capitalists, merchants and professional classes, so that a real welfare gain was registered, but unevenly distributed. He concludes that the expansion in economic activity was accompanied by some real gain in welfare, though some of the welfare effects were transferred to England in the form of cheapened foods and raw materials.

---

[5] I am indebted to Mr. E. A. Bell for these figures. Mr. Bell's estimates are based on Canadian Census returns.

[6] *Report of the Board of Inquiry into the Rise of the Cost of Living* (King's Printer, Ottawa, 1915).

[7] E. A. Bell, in an as yet (March, 1947) unfinished thesis, McGill University, "The Expansion of the Canadian Economy, 1900-1914".

How are we, however, analytically to relate to one another and to the economic process as a whole the fluctuations in population growth and in the rate of innovations when their effects move in opposite directions? It will be noted that this is a problem of a different order from that which we tackled in the two just previous chapters where population growth and innovations were regarded as *given* causes and thus were capable, with somewhat violent distortion of reality in the level of abstraction, of separate treatment. We feel that the most effective approach here is through the study of the general demand and production functions. This, it will be seen, cuts across, but need not confuse or impair, the observation of effects through the aggregate (or general equilibrium) and the real (or welfare) modes. It will involve us, however, in certain digressions which we propose to keep within narrow limits, referring the reader in the case of demand, to the appendix to this chapter for further elaboration.

2. The Demand Function[8]

The demand function expresses the functional relationship between quantities of goods demanded and certain other variables. In the Marshallian formulation of the demand function the relationship is between the quantities of any good which will be taken from the market by buyers and the prices at which the good is offered, population, tastes, techniques, the marginal utility of money and income being regarded as constant.[9] Marshall is not explicit, to my seeming, as to the nature of his assumption about other prices. It is apparent that his demand function is a particular, or partial, not a general, one, so that he does not allow for independent changes in the prices of other goods, and their possible effect, through available

[8] The material here is a brief summary of some conclusions of recent works in the theory of demand. It is highly selective in that only points relevant to the development of our own proper argument are considered. The Appendix to this chapter is intended to treat this whole subject more fully for the benefit of students who may wish a short resumé in non-mathematical language of some of the more important points emerging from recent studies and for the benefit, too, of readers who may wish to challenge the present brief summary.

[9] Since this material all rests on the Appendix to this chapter and is there documented, I have refrained from encumbering the present text with footnotes.

income, on the quantity taken, at any given price, of the commodity under consideration. Since, however, it is impossible to assume, without contradiction, at one and the same time, that real income remains constant and that the quantity consumed and price of one good vary and that all other prices remain constant (unless the demand of that one good has an elasticity of unity throughout) it must be inferred that Marshall allows compensatory changes in the prices of other goods.

The Walrasian and Paretan formulations of the demand function achieve complete generality and hence picture the interdependence of all prices, but the gain is accomplished, it seems to me, by a considerable sacrifice of definiteness and usefulness. The effects of changes in other prices, except in special cases of which Marshall takes account, are of little importance and can safely be ignored and the result is the development of a theory of demand of great flexibility and precision, capable of immediate application with gratifying results to the analysis of markets, firms' behaviour, tax incidence and international trading. The sacrifice of simplicity involved in the achievement of generality is more serious and is, perhaps, the explanation of the barrenness of the Lausanne school of any great development of applied theory. Professor Hicks, however, has wrung from the Paretan formulation a method of distinguishing between "income" and "substitution" effects. In a word, Professor Hicks presents a demand function which relates the quantity purchased of any commodity with changes in the price of that commodity (the Marshall function) and with changes in the prices of other commodities (the generalization of the function) and of distinguishing analytically the two effects.

The late Professor Henry Schultz in his monumental *Theory and Measurement of Demand* also succeeded in setting out a formally correct generalized demand function which had, in addition, the merit of being capable of fitting to time series, with a certain amount of statistical manipulation, and thus possessed the virtue of mensurability. Now the results of Professor Schultz's measurements are of the highest importance to the theory of demand considered in its temporal aspect.

It must be remembered that Marshall's function was presented under assumptions that, in effect, prescribed a stationary state. Walras's and Pareto's were static and rested on the notion of instantaneous adjustments. The Hicksian function, too, as first developed, is non-temporal, static really, though capable, Professor Hicks holds, of partial dynamic extension. Yet the essence of a generalized function is that it must be temporal, as adjustments to other prices require time and cannot be considered, as the Marshallian function can, as representing a series of potential quantities functionally dependent, irrespective of time, on a series of potential prices. Over time the demand function must be considered as existing in a three-dimensional manifold, the temporal axis of which measures the trend in time of consumers' choices. The actual curve of prices paid and quantities purchased over time follows the curve of no instantaneous demand schedule, but rather the curve of intersection of a cylinder across the demand surface.

Thus in Figure 9 the d_0, d_1, d_2 curves represent Marshallian (or could represent summed Hicksian) functions for one commodity at times 0, 1, 2. On the double assumption of continuity and orderly consecutive increases in consumption as time passes, the shaded curvelinear surface represents the demand function in time and the curve D_0D_2 is the true demand curve in time, i.e., a curve suitable for fitting to a time series.

This is, however, just illustrative. Professor Schultz found, of course, no such continuity; nor was there any such direct one-one relationship between increases in sales and the passing of time. It was, nevertheless, possible to isolate by ingenious statistical techniques three distinct relationships between quantities purchased and other variables. In the first place it was possible to show the functional dependence of quantities of a commodity purchased and the market prices of the commodity. This relationship was of the expected sort, quantities purchased increased with lower prices. Then statistical evidence is adduced to support both the Marshallian and Hicksian functions. We say both the Marshallian and Hicksian functions because price changes in related goods were not isolated, hence the resultant relationship includes both the "price" and "income"

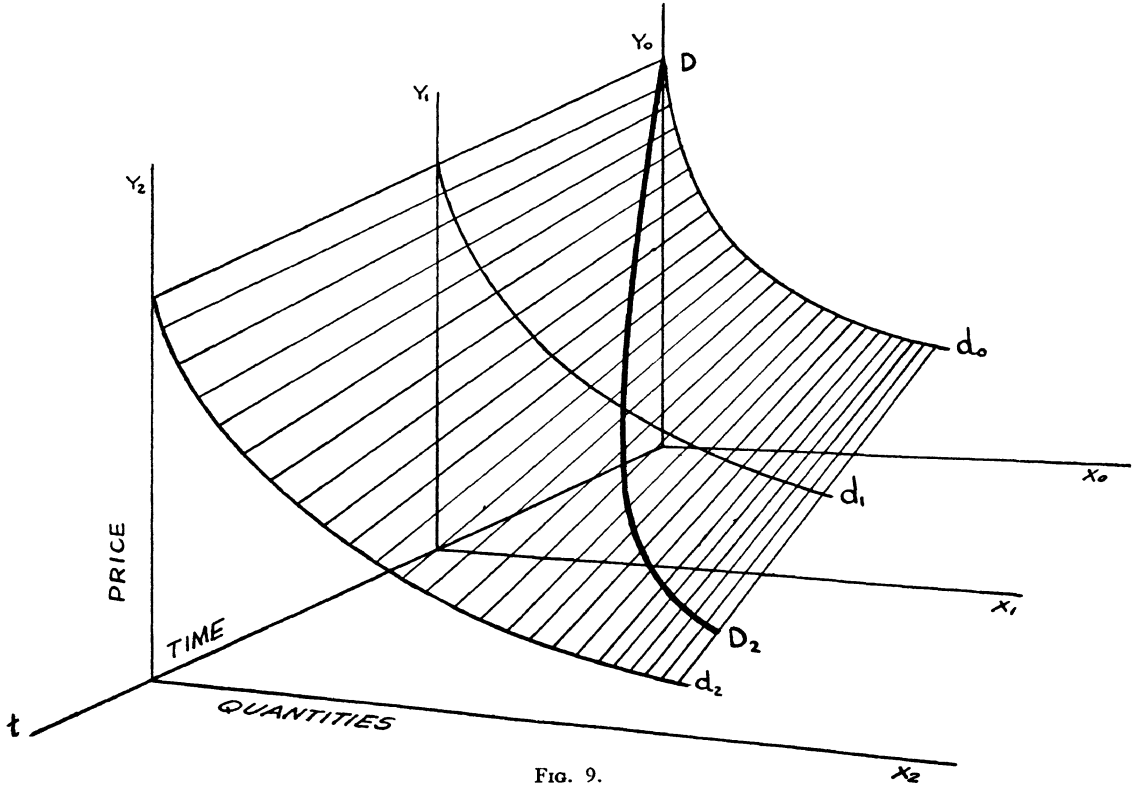


FIG. 9.

effects, to use Professor Hicks' language. Since, however, the curve has a negative slope, it follows that the slope would be negative if only the Marshall "price" effect had been considered.

In the second place Professor Schultz isolates another type of "income" effect, that is the effect on the demand for a single good of changes in the aggregate money income. This is what Professor Hansen calls the income elasticity of demand. This relationship is cyclical in character, the quantities purchased increasing with a diminution in the value of money. In the staple commodities studied by Professor Schultz, the variations from this effect were comparatively small in magnitude. It is probable that in the case of other commodities, say radios, the amplitude of the cyclical movement would be more important in magnitude.

Professor Hansen and Professor Samuelson contend, however, that consumption habits are very sticky and have small propensity to change on their own account. Their degree of response to changes in the national income remains a question of some doubt and controversy.

Perhaps the most significant relationship that Professor Schultz isolated was what he called "trend". He discovered a general trend in per capita consumption which was independent of price or cyclical fluctuation. People after 1922 just did not consume as much sugar. The same was true of certain cereals. This is not the result of the decline in the rate of population growth, for it is worked out on a per capita basis. It is to be attributed to some of the other forces "impounded" in Marshall's "ceteris paribus", to a change in tastes and consumption habits, to dietary reform, to the introduction of new feeds for cattle, and so on. Now this "trend" is very important. We cannot always know its causes, in fact, by its very nature, it defies any causal generalization. A change in population structure or quantity may well affect this trend, as may innovations of new goods, yet it may go on independently, guided by education, advertising and other institutional factors.

We return now to our main argument to ask how population change and innovations react upon and are related through the demand function. Population changes affect the demand

function of any commodity through their aggregate mode effects on total income. Such effects may be important in the case of radios and semi-luxury articles. They are not important in the cases of sugar and barley or rye. They may also affect the trend of the demand function by resulting in changes in tastes or needs, as, for example, to take the extreme illustration we have previously used, baby carriages become more popular when there is an increase in the number of people in the marriageable age group. Innovations affect the demand function by their effect on income, reducing the prices of various goods, and by their effect on trend, as when new goods are introduced or new uses for old goods discovered.

Now we saw (Chapter VII) that when the rate of population growth declined the aggregate mode effects were stagnating by reason of the decline in aggregate demand, and, though the effects through the "real" mode promised an improvement in welfare, the stagnation effects in a "trustified" economy were likely to preponderate, so that the sum effect on welfare would probably be deleterious. This general conclusion, which agrees with that of Professor Hansen, rested, however, on the assumptions that there were no innovations and no purposive direction of the process through social institutions. It was also assumed that we were dealing with an economy in which population had not pressed too intensively on the fixed factors of production. There are clearly areas in the world, such as the British West Indies, where an absolute reduction in population would increase per capita real income by so much as to offset, and more than offset, in the sum effect on welfare, the stagnation effects through the aggregate mode.

Let us now drop these assumptions. It becomes at once apparent that the decline in the rate of population growth will have the aggregate effect of impairing expectations, the rate of investment and the rate of new innovation. Innovations will not be introduced where they involve heavy fixed costs if expectations are diminished. It is also true that the long-run consumption trend of commodities is in part independent of population. The downward trend of certain commodities may be accentuated by a decline in population, the upward trend

of others will continue. Innovations both in processes and in new uses and certain types of new goods are, therefore, still possible and a good deal of pressure is created for them because they offer, and they alone offer, chances of windfall gains. It is probable, as we have shown, that the more trustified the economy is, the more likely such innovations would be.

We come now to the crux of the whole matter. A population decline *may* reduce aggregate demand, *must* reduce the demand, in terms of trend, of some goods. But it *need* not reduce aggregate demand, and the demand, in terms of trend, of many goods may continue to increase. As Dr. Moulton has pointed out,[10] Sweden has experienced a decline in the rate of population growth to the point where population is about constant, yet throughout the period of this decline her aggregate income and aggregate consumption continued to increase. This was done by maintaining the rate of innovations, often under state direction,[11] in Swedish industry, by fiscal policies directed against cyclical declines in consumption and private investment (thus offsetting many of the aggregate mode effects) and by a redistribution of income of a sort to increase per capita consumption of many basic articles. The trend of demand functions can be maintained upwards for a vast variety of articles by education and by placing in the hands of the masses adequate purchasing power. The real welfare effects of a stable population and improved technique then become possible for the entire community to share.

The nature of the guidance that may be given through institutional channels to such an economic process is beyond our present purpose. (See Part VI.) Our immediate point is

[10] "The history of Sweden affords a good illustration of the lack of correspondence between the rate of population increase and the rate of economic expansion. During the 30-year period from 1900-1930, the increase in population in Sweden was 20% while the increase in production was over 300%. Foreign investments were negligible, expansion being almost entirely internal." H. G. Moulton, *The New Philosophy of Public Debt* (Washington, 1943), p. 23.

[11] For example, the Swedish law against stream pollution forced Swedish producers of chemical pulps to work out means of utilizing waste liquors. Waste liquors from the sulphate process are ordinarily recovered in Canada, but waste sulphite liquor is not. The recovery process of waste sulphite liquor in Sweden has consequently given the Swedish sulphite producers an advantage over their Canadian and American competitors.

that there is no inevitable stagnation attached to a decline in population growth. There are other factors affecting the demand function that can be operated on to offset the aggregate effects of a population decline. There are many cases where a moderate population decline could be permitted with obvious real gain to the welfare of the population, and the decline could occur, under proper guidance, without stagnation effects.[12] In all cases, even if a decline were not necessarily desirable for its welfare effects through the real mode, the decline does not carry any inevitable stagnation. The relation of population decline to other factors is not fixed, but is institutionally controllable. Moreover, if people deliberately prefer more domestic freedom and leisure to having children, that in itself is an economic good and must be included in the total effect on welfare.

The effects of freedom from cyclical depression and innovations reducing prices of other goods and in new goods may well in turn affect the demand for babies. It has been our contention throughout that the introduction of contraceptive devices has simply cut across the usual pattern of population growth but that, once contraceptive practices and knowledge have become general, the usual pattern, at a greatly reduced average birth rate, will again emerge clearly, with higher than average birth rates following the reduction in the average year of marriage in good times and the reverse in bad times. (The time-lag between marriage and births of children may bring the high birth rate to coincide with cyclical recession, of course.) Thus a steady secular trend towards improved welfare may have the effect of reversing a declining rate of population growth or of stabilizing population at the socially desired level. The whole process, we repeat, is far from an inevitable chain of

[12] We understand, of course, that this conclusion is in no way a departure from Professor Hansen's argument. His writings constitute, and are intended to constitute, a powerful plea for just such institutional and policy guidance. Since he is able to show (Ch. XI, *Fiscal Policy and Business Cycles*) that changes in habits of consumption are not generative of cyclical fluctuations, but are a function of income, he argues that the proper method for offsetting the secular decline of population growth and investment opportunities is extensive public investment of a non-profit sort. To this question of policy we shall return in Part VI of the present work.

cause and effect, but is a typical case of social causation where all the factors of the process react on one another and where the human purpose and the social institution through which it operates can play a decisive guiding role, determining the ultimate direction of the process.

3. The Production Function

A consideration of the production function will also throw some light on our problem. The production function expresses the functional relationship between real product and the employment of the factors of production, or, to use language more in fashion nowadays, between real outputs and real inputs. Thus, if P is Product and A, B, C, . . . the factors of production, the most generalized form of the function may be written $P=F(A, B, C, \ldots)$

On the assumption that if all the factors are increased in the same proportion, the same proportionate increase must occur in P, i.e., that if scale increases occur, returns to scale are constant, we may substitute the constant λ for the function, with its usual connotation of a linear homogeneous function of the first degree. Then, by Euler's theorem, partial differentiation gives us: $P=\frac{\partial P}{\partial A}.A+\frac{\partial P}{\partial B}.B+\frac{\partial P}{\partial C}.C \ldots$[13]

Since the partial differentials $\frac{\partial A}{\partial P}$, $\frac{\partial P}{\partial B}$, etc., are the marginal products of the factors A, B, etc., on the assumption of perfect competition in the factor markets, it follows from this equation that the distributive shares of the factors exhaust the product, leaving no residual share or shares.[14]

This argument rests on certain assumptions which must be brought out.

---

[13] For this simplification of Wicksteed's rather cumbersome notations, see A. W. Flux, Review of Wicksteed's "Co-ordination of the Laws of Distribution", *Economic Journal*, Vol. IV, No. 2, 1894.

[14] What exactly is implicit here can be shown in the following manner:

Let P be product and $q_1 \ldots q_n$ the factors of production, so that $\frac{\partial p}{\partial q_1}.q_1$ is the marginal productivity of factor q_1. The marginal product of the factor may be

1. It assumes perfect competition in the factors' markets. Otherwise distributive shares do not equal marginal products and windfall surpluses of a residual sort remain.

2. It assumes that all factors are alike mobile and divisible, that is they can be increased in a continuous manner by a series of small increments. Otherwise the function would be discontinuous.

3. It is a static law, assuming instantaneous adjustments. Otherwise it would be non-linear. We shall have to inquire, in this connection, if the static assumption is logically or pragmatically justified.

Now it is our contention that these assumptions are unjustified, not because they are "unreal"—every simplifying assumption is unreal in that it abstracts from reality—but because pragmatically judged they distort rather than clarify the aspects of reality selected for analysis.

Moreover, we shall show that the assumptions are in at least one respect mutually exclusive, so that they fail to meet either the pragmatic or the logical test.

It is desired to prove that there are no residual surpluses. The assumption of perfect competition in factorial markets,

---

written $dp_{q_1} = f_{q_1}\, dq_1$, when f_{q_1} is $\frac{\partial p}{\partial q_1}$

Then $dp = f_{q_1}\, dq_1 + \ldots + f_{q_n}\, dq_n$

Now let C be the "elasticity" of production (see W. E. Johnson, "The Pure Theory of Utility Curves", *Economic Journal*, Vol. XXIII, 1913, p. 507) or the proportional increment in output to the proportional increase in input of services, so that

$$C = \frac{dp}{p} : \frac{dq_{1 \ldots n}}{q_{1 \ldots n}}.$$

Now if this is always equal to some constant k,
then $dp = f_{q_1}\, kq_1 + \ldots + f_{q_n}\, kq_n$,
or substituting Cpk for dp,

$$kC = f_{q_1}\, q_1 + \ldots + f_{q_n}\, q_n,$$

which is the equation for the exhaustion of the product, thrown into a form which shows that the distributive shares exhaust the product only when the sum of the inputs times their marginal productivities are equal to the elasticity of production times output. This means that the product is exhausted only when the increments in output are exactly proportional to increments in inputs. This condition, an elasticity of production equal to unity, must be fulfilled, otherwise the product will not be exhausted. For a fuller treatment, from which this note is adapted, see Sune Carlson, *A Study on the Pure Theory of Production*, Stockholm Economic Studies (Stockholm and London, 1939), No. 9, pp. 12-18.

at one stroke, gets rid of the differential between the marginal product and the reward paid by a monopsonistic employer. Perfect competition is a proper analytical concept when used to define a limiting market condition. It must not be used as a controversial device for obtaining the general conclusion, so full of policy implications, that the distributive shares, as determined by marginal productivity, exhaust the product. However, we should not press this too far, because it is true that if trades unions exist, it is possible for labour, even when dealing with a monopsonist, to obtain the full marginal product. That is not to say that other factors, dealing with a monopsonist, are not exploited.

There are, to move to the second assumption, grave difficulties of a methodological sort in the classical concept of capital as consisting of real instrumental goods. It is hard to determine in what sense all may be called capital, difficult to establish a satisfactory unit of measurement for varying quantities of employment. There is the dualism, nowhere more marked than in Marshall's theory of quasi-rents, between "free" capital, which has not been invested, and "real" capital in the form of producers' goods. These difficulties led to the conception of capital as a free, mobile fund of productive force, homogeneous (in the qualitative, not the mathematical, sense) and infinitely divisible, the use of which increased the roundaboutness of the productive powers, thus giving capital the dimension of time. For in whatever form capital "manifested" itself, whether in land, factories, machinery, raw material or power installations, these forms were mere temporary incarnations of an indestructible and immortal power. Such a view of capital as homogeneous and capable of continuous incremental increase in employment is only tenable in the very long run. It is thus quite incompatible with the short-run or static assumption on which the linearity of the production function is based.[15] In the short run, this almost mystical concept of capital blurs the real differences between land and other factors relatively fixed and indivisible in supply and the factors variable and divisible in supply. In fact, increases in

[15] Cf. B. S. Keirstead and D. H. Coore, "A Dynamic Theory of Rents", *Canadian Journal of Economics and Political Science*, May, 1946.

output are never accomplished by a continuous scale increase in the use of homogeneous factors. Certain fixed factors will be used more and more intensively, then there will be a discontinuous and sharp break when a "scale" increase occurs, that is when additions to land and fixed plant are made. Again this may be followed by increases in the employment of the variable factors, but it may often result in the substitution of some fixed factor for some of the variable factors. Thus the production function is neither continuous nor homogeneous (in the mathematical sense).

That it is linear is only tenable on the opposite assumption to that on the strength of which its continuity is claimed. For to be linear the functions must rest on the very short-run assumption that no innovations, economies of scale or qualitative improvements—the very essence of the temporal process—may occur. But, in fact, this discontinuous, non-linear, increase in the employment of the factors occurs because the process is essentially a temporal one, the main characteristic of which is that the factors, far from being homogeneous over time, are subject to qualitative improvement. It is the possibility of realizing the windfalls from qualitative superiority that leads entrepreneurs to increase employment and to alter the combination of the factors. As we shall see in Part IV, the characteristic of the firm in expansion is a discontinuous alteration of the margin of substitution under the stimulus of possible gains from innovations.

We conclude that the production function is discontinuous and non-linear, that there are surpluses or rents of a residual sort which, though emergent from the social productive process as a whole, accumulate to the benefit of entrepreneurs and the owners of factors of superior quality, relatively fixed in supply. We see that these surpluses are dynamic, that is that they result because of the temporal nature of the process. They are not definitely determinate.[15a]

[15a] Mr. Jean Waelbroeck, a McGill student, has prepared, since this was written, a Note for publication, in which he shows that under certain assumptions these rents may be shown to be determinate. He works out an ingenious technique, using the concept of expectations, to show how dynamic rents are determined and how they affect entrepreneurial behaviour.

Now, to return from this second digression, a decline in the rate of population growth affects on the one hand the supply of one of the factors, namely labour, and, on the other hand, through the aggregative mode, the possibility of windfall gains and hence the derived demand for innovations and the employment of the fixed factors. Innovations, themselves, are the basis of qualitative changes in the factors and the source of differential gains. It follows that there will be stagnation effects from a decline in population growth if the following conditions are satisfied:

1. The decline in population growth must bring about a general diminution in aggregate demand. This, we have seen, depends on institutions and policy.

2. The economy must depend on the existence of large surpluses of a residual sort for incentive to technical advance. A highly competitive economy could experience continuous advance from the competitive pressure on each firm to maintain its market position: surpluses could not be retained. A trustified economy depends on the retention of large surpluses for technical advance. This again is clearly an institutional matter. The trend, during the nineteenth century expansionist period, towards larger productive units and integration of ownership, a trend causally connected with innovation yielding economies of scale, has given us an economy of a trustified sort, for which the failure of population to go on expanding is a serious matter. Innovations requiring increases in scale and yielding increased outputs at lower unit costs are less welcome, as surpluses cannot be counted upon to develop if the demand over time slopes more sharply downward than the cost curve. An economy that freed itself from this dependence on differential surpluses could continue to experience technical advance, taking the gains in the form of increased leisure, if it desired, as much as in the form of increased goods for consumption.

3. Innovations must be cost reducing with decreases in scale. This is the sort of "capital saving" innovation that Professor Hansen has in mind. Innovations of a "capital-consuming" sort would act as a stimulus to aggregate investment. Clearly

"factor-saving" innovations, which make for a diminution rather than an increase in aggregate investment, will be stagnating in their aggregate effects, especially if they are labour-saving as well as capital-saving, and it is probable that they will occur if the economy, institutionally speaking, is unwilling or unable to undertake large-scale investment in plant of a real income-increasing sort without the promise of large gains. As Professor Hansen's fiscal proposals show, he is aware of the institutional issue involved and of the need for freeing the economy from its dependence on large differential gains for technical advance. In a word, the production function has always been regarded (and in the past rightly) as a positive function, i.e., output varies in the same direction as employment of the factors, though, as we have argued, the rates of change may differ. Professor Hansen is saying that a diminution in scale, a reduction in the employment of the factors may be accomplished without any reduction in output. Given a certain type of innovation, this may well be true, but it is serious only if the public taken as a whole prefer more leisure to more goods and if the increased leisure is unevenly distributed, so that some work as hard as ever, while others are totally unemployed and dependent on charity or public relief. These are clearly institutional matters, subject to the control of the social purpose.

4. Synthesis

The experience of modern capitalist society has been one of rapidly increasing population growth accompanied by an almost continuous technical advance, followed by a decline in the rate of population growth with, however, no corresponding decline in the rate of technical progress. The increased severity of the depression phase of the ten-year cycle, resulting from conjuncture with post-war dislocations and the stagnating effects of the declining rate of population growth, has led some thinkers to the general conclusion that western capitalism is a stagnating, and even "decadent" society.

Our position is that this is not so much an incorrect as a too simple view of the historical process. In the first place,

we distrust the method of examining certain economic forces in a kind of abstraction, whereby one says, with Marx or Professor Hansen, the decline of investment opportunities will lead to increasing stagnation until catastrophe envelops society, unless society mends its ways and adopts the remedy of socialism or a "functional" fiscal policy. We have to see that what we decide to do with and through our institutions is a part of the causal process and that, for example, there need be no decline in investment opportunities. What is wanted of economics is not so much a single remedy, a panacea, as an analysis which defines the various choices we can properly elect with their probable consequences.

Our thesis throws primary emphasis on technical advance. Our study of population change shows that the possibility of increased population in the eighteenth century depended on innovations which raised the productivity of labour. Had this not been so the pressure of population, as in the Ricardian model, would have imposed a limit on growth. Again, in later periods, innovations affected tastes, consumption habits, the demand for leisure and the regulation of the size of family, and hence the rate of natural increase. At all stages of the process, the rate of population growth may have stimulating or retarding effects on economic activity and the level of welfare, but, as we have seen in Chapter VII, the nature of these effects will depend on the rate of technical progress and the characteristic institutions of the society.

Technical advance, however, is causally related to these institutions. In the early period the new machines broke the old trusts of the mercantile economy and led to a period of active competition. The progress in techniques led, however, to unstable competition, growth in the size of firms and ownership concentration. This corruption of the market, or increase in the degree of monopoly, is of the utmost importance, because, as our argument has shown, it has a direct negative effect on welfare in that it prevents a general sharing of the gains of technical progress, it tends to offset the stimulation from accompanying population increase, it is the prime cause of large surpluses or residuals in the distribution of the social dividend and

hence leads to accumulations of capital which cannot profitably find investment because of limited purchasing power. When population, responding to this process and coming under conscious control, failed to increase at an accelerating pace, the stagnating effects through the mode of aggregates became accentuated. Moreover the trustified society is much less able to adjust rapidly and easily to changes in the general demand function, changes which come as innovations (including education) and changes in the age structure of population which affect consumption habits. This rigidity results not only in misallocation of resources, but in permanent "spotty" depressed areas, areas which have depended on one or two "staples", the world demand for which has undergone marked negative change. As we shall later show (Part V) the problem of depressed areas is only the regional aspect of this general process of change.

We are not yet in a position to pronounce on the possible effects on innovation of this process which they themselves initiate. We shall, in the next Part, discuss their effect on the individual firm and reach certain conclusions about the tendency of monopolistic firms to slow down the rate of innovation.

The position adopted in these pages differs, therefore, from Professor Hansen's, in emphasis. Our emphasis on the institutional effects of innovations leads to a conclusion not unlike his thesis of secular stagnation. On the policy level, however, the difference in emphasis in the analysis betrays itself. We believe, as will be shown, that a functional fiscal policy has a place in the purposive control of the economy, but that basic causes of instability and lack of welfare must also be attacked by other methods.

Our thesis, though like Professor Schumpeter's in the emphasis on innovations, differs also from his. By marrying indissolubly his analysis to a perfectly competitive model, Professor Schumpeter prevents himself from observing the effects of innovations on the institution of the market and confines his attention to the monetary institutions. He thus fails, in our opinion, to see the serious menace contained in the technical process to the institutions of western society, and is able to reach his

rather cheerful conclusions with urbane regrets of the inevitable cycle involved in progress.

I am struck by the similarity of much that I have written to the view presented by Mr. Maurice Dobb in a book which has reached me since this work was near completion.[16] Mr. Dobb's approach is historical and analytical and his analytical concepts are derived from Marx. While, in Mr. Dobb's hands, the Marxist thesis is freed of some of the limitations we noted in Part II, I feel that he still fails to reveal clearly just how innovations have led to corruption of the competitive market, and I am also oppressed by the Marxist implication that the process is an inevitable one that cannot be directed into desirable channels within the general framework of our present social institutions. However, this implication, which I sensed in much of the work, is positively repudiated in the final paragraphs. There Mr. Dobb writes, "The continuance into peace time, as well as into the transitional period between war and peace, of a form of State Capitalism, democratically controlled and operated in the interests of labour towards the maintenance of full employment and the curbing of monopolistic business organization, seems the most probable outcome over at least a major part of Europe".[17] Mr. Dobb views such an outcome with approval, and I have no quarrel with that.

[16] M. Dobb, *Studies in the Development of Capitalism* (London, 1946).

[17] M. Dobb, op. cit., p. 386.

APPENDIX TO CHAPTER IX

THE THEORY OF DEMAND

In the main text we referred to certain developments in demand theory and their significance for the theory of economic change. It seemed too extreme a digression, even for this peripatetic volume, to review the literature on which we were drawing at that stage. Yet for a certain class of student reader, especially those who have difficulty with mathematical argument, such a review may have value and may aid in the comprehension of our main line of argument. Hence this Appendix which aims at a simplified and essentially non-mathematical presentation of the outlines of the theory of demand and of the measurement of temporal changes in demand.

1. THE GENERAL THEORY OF DEMAND

Currently used by economists for various analytical purposes are three formulations of the demand function. We distinguish these as the Marshallian, the Paretan-Hicksian,[1] and Keynesian. The Marshallian function is that familiar to most undergraduate students. It rests philosophically on the theory of utility and is alleged to assume the measurability of utility in cardinal numbers. It relates the change in sales of any commodity directly with the change in price—incomes, the purchasing power of money, taste, fashion and population being regarded as unchanged. The Paretan-Hicksian function rests on the

[1] We have called this form of the general demand function "Paretan-Hicksian" for purposes of identification, using the names of the writers most widely associated with it. Actually, as Professor Hicks tells us in his *Value and Capital* (Oxford, 1939), Ch. I, it has a long history of development, with contributions by Walras, Edgeworth, E. Slutzky, R. G. D. Allen, Luigi Amoroso, E. Phelps-Brown, Henry Schultz and others. Professor Knight refers to the same demand theory in his "Realism and Relevance in the Theory of Demand", *Journal of Political Economy*, Vol. LII, No. 4, Dec., 1944, as the "Slutzky Theory", because of the important part played in its development by the Slutzky article "Sulla teoria del bilancio del consumatore" (*Gior. degli econ.*, LI, No. 1, 1915). See, also, J. R. Hicks' acknowledgments, op. cit., p. 19.

objective comparative evaluation of one commodity in terms of others and assumes only ordinal measurement of commodities—"I prefer two of this to one of that"—and relates the change in sales of any commodity with the change in its price relative to other prices and with changes in real income. The Keynesian function is of a different variety entirely. It expresses the functional relationship of aggregate demand with changes in aggregate income. It differs from both the other two in that it is concerned with the aggregate demand for all consumers' goods taken together, rather than with the demand for a single good, and it relates the dependent variable to income rather than price as the independent variable. It is true, of course, that Professor Hicks provides for the "income effect", but he clearly means real income, whereas Lord Keynes is talking—or ought to be[2]—about changes in aggregate money income. In brief, the Keynesian function is an analytical concept of "macro-economics", whereas the Marshallian and Paretan-Hicksian formulations belong to "micro-economic" analysis. We have earlier dealt with some of the aggregative aspects of demand changes and in the present context are not further concerned with them. The problems of demand discussed in the text of Chapter IX lie in the field of micro-economics, so it is to a discussion of the Marshallian and Paretan-Hicksian functions that we now limit ourselves.[3]

The important difference between Marshall and Professor

[2] "Effective demand is the demand that is causally effective in an economic system employing money. It is the real demand of a community modified by its capacity to pay" (H. G. Littler, "A Pure Theory of Money", *Canadian Journal of Economics and Political Science*, Vol. X, No. 4, Nov. 1944, p. 433).

[3] Let us not suggest by the incautious language of the précis writer that micro-economics and macro-economics have no problems in common, or that there is no bridge between the two. Prof. Hicks in *Value and Capital* erects such a bridge, moving by stages from the study of equilibrium in a micro-economic world to the discussion of just such problems as engaged the attention of Lord Keynes. "The particular version of the model (of the trade cycle) which is set out in the concluding chapters of *Value and Capital* is only one possible version; it is obviously of less practical utility than those versions which make more use of Savings and Investment, concepts for which I felt a certain aesthetic repugnance. Yet there is something to be gained from being able to set out a theory in several different ways . . . " This quotation is from Professor Hicks' contribution to the symposium on "Inter-relations of Shifts in Demand" by D. H. Robertson, J. R. Hicks and Oscar Lange, *The Review of Economic Studies*, Vol. XII, No. 31, p. 74.

Hicks is the choice of parameters. Strictly speaking, a parameter is an arbitrary constant, any value of which will define the character of any member of a system. The system may be a system of curves, functions, or other mathematical relations. For example, we might set temperature at 50° Centigrade and plot the functional relationship of pressure and the volume of a gas. The resulting function could be graphed, volume against pressure, and the temperature parameter of the graph would be 50° Centigrade. If we let temperature be 60°, then 70°, then 80° and so on, we should derive a system of curves, all showing the functional relation of volume to pressure, and that relationship would be characteristic throughout the system in the sense that volume would decline with pressure and the curves of the system would all be falling curves. The arithmetic values for volume at any pressure would, however, differ for the different members of the system, that is the average height of the curves in the system would differ with the change in the value assigned the parameter. If we suppose that some chemical investigator into the properties of gases had entirely ignored changes in temperature and had selected for his experiments some other parameter, say the position of the heavenly bodies, he might well have found some odd variations in the results of his experiments with pressure, occasioned by the unnoticed changes in the temperature of his laboratory. The advance of science is marked by growing appropriateness in the selection of parameters.

In the theory of demand the functional relationship to be observed is that of quantities purchased plotted against prices per unit. This may be written quite simply

$$D_1 = f(p_1)$$

when D_1 is the quantity purchased of commodity 1 and p_1 is its price. The parameters selected for the study of this relationship have an important effect on our results. For example, if we neglected to assign a constant value to the purchasing power of money as one parameter, we should find we had a surprising relationship during a period of rapid deflation. The price of our commodity would fall with the general fall

in prices, but, if our commodity had a particularly high income elasticity, the quantities purchased would also fall. This would strike us as very curious indeed, and like Alice we might well inquire, "Which way? Which way?" The selection of the appropriate parameters is as important for the economist as for the chemist, but the difficulty is greater because the choice is so much more complex.[4]

It will be seen that the parameters of the demand function constitute in their entirety the contents of economic time. It is the change in these constants that marks the passage from one momentary position to another. The actual demand for any commodity seen over time is a system of pure demand functions with shifting parameters. Our argument will consequently consist of the following steps:

(a) The definition and selection of the parameters,
(b) The statistical distinction of changes in demand deriving from changes in certain parameters,
(c) The indication of the relevance of this to the theory of change.

Marshall sets out the parameters of his demand function in the following words:

"The demand schedule," he says, "represents the changes in the price at which a commodity can be sold consequent on changes in the amount offered for sale, other things being equal. But in fact other things seldom are equal over periods of time sufficiently long for the collection of full and trustworthy statistics. There are always occurring disturbing causes whose effects are commingled with, and cannot easily be separated from, the effects of that particular cause which we desire to isolate. This difficulty is aggravated by the fact that in eco-

[4] "The number of parameters we may introduce into our demand functions are legion; they may include anything on heaven or earth which has any relevance for human conduct. And there is no reason why two commodities which are sympathetic (i.e., affected in the same direction) with respect to one parameter, should also be sympathetic with respect to another . . . Mustard may be said to be companionable (i.e., sympathetic with respect to most parameters) with beef, because nearly any parameter which leads to an increased demand for one will also lead to an increased demand for the other. (But clearly not every parameter; the preaching of vegetarianism would diminish the demand for beef, but might at the same time maintain the demand for mustard for use as a salad dressing)" (J. R. Hicks, "Inter-Relations of Shifts in Demand", pp. 72, 73).

nomics the full effects of a cause seldom come at once, but often spread themselves out after it has ceased to exist.

"To begin with, the purchasing power of money is continually changing, and rendering necessary a correction of the results obtained on our assumption that money retains a uniform value . . .

"Next come the changes in the general prosperity and in the total purchasing power at the disposal of the community at large . . .

"Next come the changes due to the gradual growth of population and wealth . . .

"Next, allowance must be made for changes in fashion, and taste and habit, for the opening out of new uses of a commodity, for the discovery or improvement or cheapening of other things that can be applied to the same uses with it."[5]

This seems perfectly clear. Marshall is intending to discover how the quantity demanded of any one good varies with its price, and he is going, in his own phrase, "to impound in *ceteris paribus*" all other possible changes that might affect or influence the relationship he wishes to investigate. However, in two respects, Marshall is less definite than was his wont. "Impounding in *ceteris paribus*" is a dangerous business and lacks the clarity of proper mathematical ignoration. The method of mathematical ignoration is to assign a definite constant value to each parameter and to hold that assigned value constant over a series of observations. The results are unequivocal; no hidden or disguised changes in a parameter can occur. The method of impounding, if it may be so called, is loose and full of traps for the unwary. One does not know what values are assigned the parameters so that a change in one or several may go undetected.[6] There is no certainty that the (indefinite) values assigned them are consistent with the assumption that they may, without mutual antagonism,

---

[5] Marshall, *Principles of Economics* (2nd ed. London, 1891), pp. 166-68.

[6] "Thus the demand curve for wheat when the price of rye is kept constant at seventy-five cents a bushel may be considerably different from what it is when the price of rye is kept constant at a dollar and fifty cents a bushel." Henry Schultz, *The Theory and Measurement of Demand* (Chicago, 1938), p. 9.

all be held constant together. The record of text-book treatment of the Fisher quantity of money equation is testimony to these dangers.

The second ambiguity in Marshall's position is his attitude towards changes in the prices of other goods. It is clear, I think, that Marshall does not intend to allow for independent changes in the prices of other goods to interfere with his problem. By assuming no change in real income or in the purchasing power of money he rules out any such general price change. On the other hand, he cannot rule out all changes in other prices. It is quite improper, as Professor Knight has shown, to assume at one and the same time, *both* that income is constant and that there are no changes in other prices.[7] If the price of one particular commodity alters and there is an alteration in its sales there must either be compensatory changes in the prices (and sales) of other goods *or* in real income. (There is theoretically the unlikely exception in the case of a commodity the demand for which had an elasticity of unity throughout.) It must be concluded therefore that Marshall meant to allow for compensatory changes in the prices of other goods. The Marshallian function might then be written

$$D_1 = f(p_1, K, R)$$

When R is real income and K is all the other constants assigned under "*ceteris paribus*".

Now $R = \sum_{k=2}^{n} p_k D_k + p_1 D_1$.

If p_1 changes and D_1 changes as a consequence, the product $p_1 D_1$ will be altered by some fraction "h". (Unless the demand for commodity 1 has an elasticity of unity throughout.) Then we must have.

$$R = \sum_{k=2}^{n} p_{k1} D_{k1} + (1+h) p_1 D_1 ,$$

where p_{k1}, D_{k1} represent the changed magnitudes of p_k and D_k, necessary to maintain the equation.

[7] F. H. Knight, "Realism and Relevance in the Theory of Demand", *Journal of Political Economy*, Vol. LII, No. 4, 1944.

Since $\sum_{k=2}^{n} p_k D_k = \sum_{k=2}^{n} p_{k1} D_{k1} + h(p_1 D_1)$,

$$\sum_{k=2}^{n} p_k D_k = \sum_{k=2}^{n} (p_k D_k + p_k \triangle D_k + D_k \triangle p_k + \triangle p_k \triangle D_k) + h(p_1 D_1),$$

writing $p_k + \triangle p_k$ for p_{k1}, etc.

Then

$$-h(p_1 D_1) = \sum_{k=2}^{n} p_k \triangle D_k + \sum_{k=2}^{n} D_k \triangle p_k + \sum_{k=2}^{n} \triangle p_k \triangle D_k$$

This is the most general statement of the condition that must be filled if the equation is to be retained and $p_1 D_1$ to be allowed to vary.

The Marshallian formulation of the demand function has been, as we all know, most fruitful in all cases where what Professor Hicks calls "income effect" is negligible and where one was not concerned with the complications arising from long-run changes in the parameter. "The simple law of demand—the downward slope of the demand curve—turns out to be almost infallible in its working."[8] The weaknesses of the Marshallian function are discovered in the cases of "inferior goods", where the income effect is important,[9] and in the long run where shifts in the temporal parameters are important.[10]

---

[8] J. R. Hicks, *Value and Capital*, p. 35.

[9] Ibid., p. 28.

[10] We ignore in the text the claim made by writers of the Pareto-Hicks school to have freed demand theory from dependence on the theory of utility. This claim rests on the technique of the indifference curve by means of which commodities are directly valued in terms of one another. The diminishing marginal rate of substitution replaces the concept of diminishing marginal utility. We are not inclined to take this claim very seriously. The Jevonsian theory of utility never did tie economics to a Benthamite utilitarianism. It simply implied that people did have some notion of the utility of a good to themselves and could measure this sufficiently accurately to base their market behaviour upon it. This is no more extravagant an assumption than that which underlies the indifference curve. Indeed the convexity of the indifference curve to the origin is itself dependent on the diminishing marginal utility of goods with an increase in the quantity possessed. Even if one accepts Professor Hicks' elegant—too elegant—proof of convexity (*Value and Capital*, p. 23), it applies to marginal utility, for, if there must be a dim-

The contribution of Professor Hicks and the others of what Professor Knight calls the "Slutzky school" was to free demand theory of its dependence on utility and to modify the parameters of the function so as to allow for the income effect. The former contribution is of disputed value (See footnote 10), the latter, while of little practicable use in the majority of "standard" cases, has the theoretic virtue of greater generality, enables the Giffen paradox to be explained, and yields a demand function more susceptible of temporal and statistical treatment. While retaining the assumption of a constant purchasing power of money, Professor Hicks drops the assumption of constant real income. Thus a change in price, with an accompanying change in purchases, will result in altered real income. If the commodity in question is a sufficiently important item of expenditure this income effect will influence the purchases and prices of other goods. Changes in real income, and via such changes, the inter-related changes in prices and purchases of all goods, are thus reflected in the demand function. In some cases this may be important. The sales of margarine will decline with increases in income, even if its price is falling. In any event its sales decline rather than increase with increases in real income. Professor Hicks works out ingeniously a method for distinguishing theoretically the "price effect" and the "income effect". For purposes of explaining how quantities demanded change with income and price over a period of time, however, some method of statistical measurement and isola-

---

inishing marginal rate of substitution, the marginal utility of any good may then be said to decline with increases in the quantity possessed. We are on this point inclined to agree with Professor Knight when he says that the "Slutzky school do not free economics from a theory of utility, but merely substitute for the item of utility as the property which causes it (a good) to be consumed and chosen for consumption in preference to other goods the idea of economic value as solely a comparison between the particular good and other goods." (*Realism and Relevance in Demand*, pp. 293-94). This substitution is a questionable step. The indifference curve implicitly assumes the principle of diminishing marginal utility. If this principle did not hold, successively smaller quantities of one commodity would not be offered to obtain any given quantity of another. There can be no serious doubt of the principle of diminishing marginal utility and the "Slutzky school" in effect imply the principle in their argument. "Even if it were possible to explain the facts of demand without it (diminishing marginal utility), the procedure would be interesting only as an intellectual 'stunt' " (Knight, op. cit., p. 298).

tion must be discovered.[11] This problem is dealt with, and a tentative solution offered, in the work of the late Professor Henry Schultz, *The Theory and Measurement of Demand.* When one sets out to measure changes in quantities demanded, one has to begin with a time series of quantities purchased, and another time series of prices. The crude relation of prices to quantities will not, however, yield a demand function. In the temporal process, income is changing and so are the other parameters. "Economic Theory", said the late Professor Allyn A. Young, "has never professed to deal with the temporal succession or the spatial distribution of unique combinations of circumstances, while statistics has to deal, in the first instance, with nothing else".[12] Thus if we are (a) to demonstrate statistically the shape of the demand function and (b) to explain the nature of temporal change in the demand function by quantitative means, we must be able to isolate and measure the changes in the parameters. It is to this question that we now turn.

2. The Measurement of Changes in Demand[13]

The dangers of a direct attempt to use the Marshallian theory, with the corollary that demand and supply curves can shift in any direction, for purposes of temporal observation, appears in the use made of this theory by Professor Leontief.[14] Professor Leontief fitted straight-line curves, one positively inclined "supply" curve and one negatively inclined "demand" curve, to the scatter of the logarithms of the observations of market transactions over a period of time. He assumed (a) that each market transaction represented the intersection of instantaneous demand and supply curves which change their

[11] Cf. Professor Pigou's footnote on Professor Moore's "Measurements of the Elasticity of Demand for Pig-Iron", A. C. Pigou, *The Economics of Welfare* (London, 1920), p. 925.

[12] Quoted from Allyn Young, "English Political Economy", *Economics*, Vol. VIII, 1928, p. 10, by Schultz, op. cit., p. 61.

[13] I am indebted to Mr. S. V. Suggett of the Economic Research Section of the Bank of Canada, formerly a graduate student at McGill University, for assistance in the preparation of this section.

[14] Wassily Leontief, "Ein Versuch zur statistischen Analyse von Angebot und Nachfrage", Weltwirtschaftliches Archiv, XXX, Heft I (July, 1939). I have depended entirely on Professor Schultz's account of Professor Leontief's method, because I am unable to read German. Somewhat naturally, therefore, I have

position from time to time, (b) that for each curve the elasticity is reasonably constant with alterations in position, (c) that the shiftings of the demand and supply curves are independent of one another, and do not affect the shape (elasticity) of the curves.[15] Professor Schultz, in considering Professor Leontief's results, was able to show that they failed to account for the facts. Professor Schultz attributed this to the use made of the Marshallian function and the corollary that the two curves are independent of one another and of changes in other prices. Thus the Leontief data does not sort out the different kinds of changes that may occur over time and attribute them to their probable causes. A one per cent increase in the price of sugar never did call out a 15% increase in output, as Professor Schultz finds a testing of the Leontief method to show, and the output increase must be imputed to some other cause of change. It is argued that if we wittingly introduced large errors of deviations into the data of sales and prices, these errors would be explained, under Professor Leontief's method as coordinates and points of intersection of hypothetical demand and supply curves. They would all be attributed to pure demand response to price or supply change. Only when the various causes of change were properly isolated, could the deviations be detected and their probable error discerned, tested, and the error itself eliminated after re-examining the data.

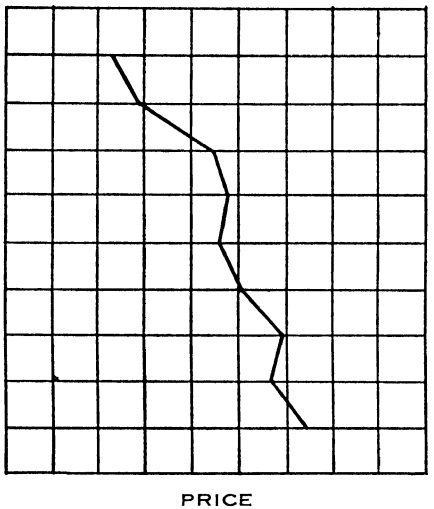
If, on the other hand, we begin with the assumption that the demand curve is relatively fixed over time, while conditions of supply change, we can treat our observations (i.e., time series

---

accepted the former's criticisms. I should hasten to add that all this work in the measurement of demand is in the experimental stage and each investigator contributes both by his achievements and his errors to the development of a solution. Of Professor Leontief's work, Professor Schultz says, "Students of the subject will, however, always be grateful to Leontief for his bold and painstaking attempt to deduce the true statical, Cournot-Marshall demand and supply curve from statistics. His efforts will not have been wasted if they serve to convince economists and statisticians of the futility of trying to obtain these curves without first examining . . . whether and to what extent the Cournot-Marshall law of supply and demand has meaning in terms of operations" (Schultz, op. cit., p.95).

[15] Here as occasionally elsewhere I have condensed from Professor Schultz and have retained some of his phrasing. I have not used inverted commas, however, because I have altered the phrasing in the process of summarizing. In this case, see Schultz, op. cit., p. 83.

of prices and quantities) as representing the intersections of a series of different supply curves with the same demand curve.[16] An imaginary time series would then look like this:



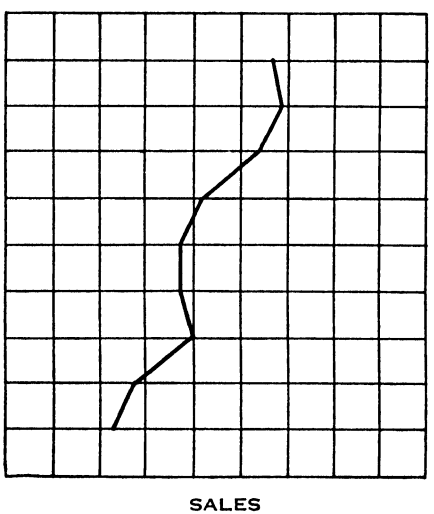


FIG. 1. APPENDIX TO CH. IX.

[16] See E. J. Working, "What do Statistical Demand and Supply Curves Show?", *Quarterly Journal of Economics*, Vol. XLI, 1927, pp. 212-35, cited by Schultz, op. cit., p. 23.

A scatter showing the correlation could then be prepared like this:

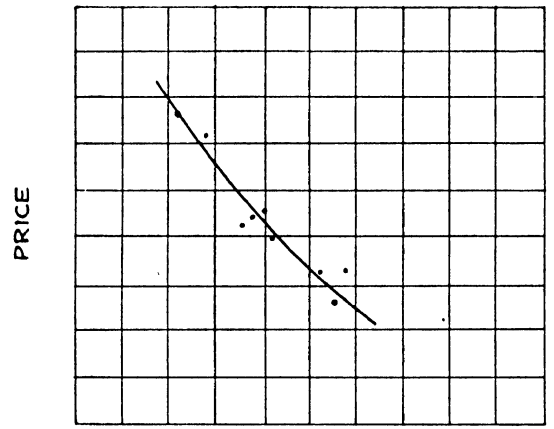


FIG. 2. APPENDIX TO CH. IX.

Smoothing the scatter with a hand-fitted curve yields a "demand" curve, on the assumption that demand is stable over time, and the various points of sale mark the intersections of the series of supply curves.

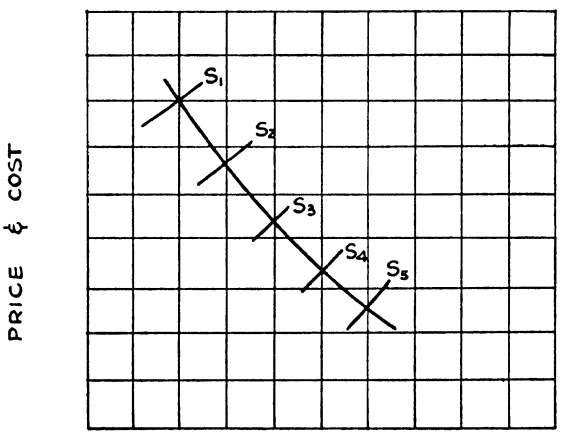


FIG. 3. APPENDIX TO CH. IX.

This gives a helpful start for the problem of measuring the demand curve and for the statistical verification of the general demand function. It does not lead, however, to anything bearing on the mensuration of shifts in demand, and it is, of course, suitable only for application to such commodities as may reasonably be assumed to have relatively fixed demands.

Professor Schultz attempts to remedy these defects. Following and developing the methods of Mr. H. L. Moore,[17] he assumes that there is a certain routine or generality in the consumption pattern of human beings, that the statistical data of consumption and price are a fair reflection of this pattern and that, consequently, the statistically unknown theoretical demand function "can be approximated by various empirical curves."[18] Agricultural produce, he believes, can most safely be used for statistical treatment because there can be little doubt of the appropriateness of the assumptions when applied to them. Professor Schultz selects, for reasons we have made clear in Section 1 of this Appendix, the formulation

$$q_1 = f\,(p_1 \ldots p_n,\ R,\ t)$$

as the most correct statement of the theoretical demand function. In this equation, q_1 is the quantity purchased of commodity 1, $p_1 \ldots p_n$ are the prices of n commodities, R is income, and t is time, as composed of the various long-term trends in consumption habits, i.e., the changes in the parameters of the instantaneous curves.[19] In spite of the theoretic preference Professor Schultz expresses for the demand function in this form, the formulation he uses for measurement comes much

[17] H. L. Moore, *Economic Cycles: Their Law and Cause* (New York, 1914), Ch. IV, and other works, as cited by Schultz, op. cit., p. 67. Of Professor Moore's contributions, Professor Schultz says (p. 63) "The statistical study of demand is a new field in economics and may be said to be the creation of only one man, Professor Henry L. Moore."

[18] Schultz, op. cit., p. 65.

[19] It is thus clear that Professor Schultz in spite of a rather formal seeming treatment of time, does in fact define it as we have done in terms of those changes that occur over time.

closer to the Marshallian function. His attempts to allow for the income effect fail to be completely satisfactory.[20]

When we now come to correlate prices with quantities purchased over time we have to distinguish between gross and net regression curves. This means, in concrete terms, that we must, by some means, distinguish between the changes in consumption that occur over time because of trend in consumption habits (taste, fashion, age structure of population, etc.) from the gross changes in consumption. The net changes in consumption (the change that is left after correction has been made for "trend") can be imputed to changes in price. Then, if some measure can be devised to approximate consumption response to changes in R, a further correction can be made and the temporal deviations can be imputed in part to income shifts and in part to "trend".

The first step is the isolation of trend. We start with a gross regression which expresses a high negative correlation between price and consumption.[21] Professor Schultz applies this to sugar consumption between 1875 and 1895. He assumes that a straight line fitted to the scatter is the true demand and then calculates the difference between this and the observations. He finds these differences or deviations can be plotted against time with a strong correlation. Thus he finds in the 1870's that per capita consumption of sugar lay consistently below the "true" demand curve at all observed prices and after 1883 lay above the true curve. This enables him to conclude that the general trend was a steady increase in per capita consumption of sugar regardless of price, i.e., at the same price per pound more sugar per capita was consumed in the 1890's than in the 1870's. Thus Professor Schultz reaches the demand function for sugar in this form: $x = a + b\,y + c\,t$

[20] See p. 190. In applying the various alternative formulations of the demand function to empirical data, Professor Schultz found in favour of the most simplified forms of the Marshallian function in dynamic form, viz., $x = f\,(y,t)$ *and* $x = f\,(y,R,t)$. Only in the case of "related goods" did he elect to use $x_1 = f(y_1 \ldots y_n,t)$, and there admits that the increase in the number of variables impairs the accuracy and reliability of the results. See op. cit., pp. 136-39.

[21] The balance of this section is based entirely on Professor Schultz's *The Theory and Measurement of Demand.* I am indebted to the publishers, the University of Chicago Press, for permission to quote the demand equations.

where x is consumption per capita, computed from both variables, y price per pound, t trend and a, b, and c the constants of the regression equation.

Fitting the curve to the observed data by minimizing the sum of the residuals (y and t being assumed to have no deviation), he arrived at the following demand equation for sugar:

$$x = 70.62 - 2.259y + 0.8371t$$

(x = pounds per capita consumed.
y = cents (corrected for changes in value of money to 1913 base)
t = July 1st, 1885)

This means that over the period under observation, for every increase of 1 cent per pound in price, there was a decline of 2.259 pounds per capita in consumption, other things remaining equal. That is the "pure" demand function. But other things did not remain equal. There was a normal increase in consumption of .8371 pound per capita, per annum.

Professor Schultz calculated the demand function for sugar of the period 1896-1914 and then for 1915 to 1929. His results are shown below, along with the estimated elasticities of demand and the regression of x to y. (For non-mathematical readers, the latter may be taken as serving as a check on the results. It measures the perfection of the negative response of sales to the increase in price. If the regression was equal to -1, that would mean that there was no deviation in the observed data, corrected for trend, from a 2.259 lbs. decline in sales for every 1 cent increase in price. It thus acts as a measure of the degree of reliability of the equations.)

		$\mathcal{N}$(Elasticity)	$r(x,y)$
1875-1895	$x = 70.62 - 2.259y + .8371t$	$-.36$	$-.915$
1896-1914	$x = 92.90 - 3.341y + .9197t$	$-.26$	$-.920$
1914-1929	$x = 134.51 - 7.8031y + .0900t$	$-.34$	$-.901$

We might note here that the regressions are good, but there is an odd variation in the elasticity values. The sharp decline in the importance of t in the final period is also notable, as it suggests that the trend towards higher per capita consumption was about over. It may be that, with dietary reform, for

example, the present period would show a negative value for t. Moreover we do not know whether some of the variation in t might be owing to changes in income. This possibly would explain the odd break in elasticity. Ordinarily we should have expected a general slight decline in elasticity as sugar became a more and more accepted necessity of ordinary diet.

Professor Schultz corrected the data for income variations. He was able to establish deviations of per capita consumption with income changes as measured by the index of industrial production and trade (till 1919) and (1915-1929) by direct income data. With w as the index of income, he established the following equations.

$$\begin{aligned}
&1875\text{-}1895 \quad x = 82.02 - 2.4758y - .0950w + .7900t, \quad n = -.397\\
&1896\text{-}1914 \quad x = 120.50 - 4.6810y - .2033w + 1.0558t, \quad n = -.361\\
&1915\text{-}1929 \quad x = 165.29 - 7.0589y - .3314w + .0969t, \quad n = -.306
\end{aligned}$$

The regression values for x against the three variables were satisfactory.

The conclusions that follow from this analysis have some odd features. Though the general shape of the "true" demand curve, i.e., the theoretic curve, is definitely established as a downward sloping curve, the dependence of per capita consumption on price is seen to be much less affected by trend in the later than the earlier periods, while more affected by income fluctuations. In the first period a 1 cent increase in price led to a decline in per capita consumption of about 2.5 pounds, while from 1915-1929 an increase of 1 cent in price led to a decline of 7 pounds in consumption. (The proportionate decline in consumption to the proportionate increase in price, the elasticity of demand, was, however slightly less in the later period than in the former.)

An extraordinary and unexpected conclusion is that the income elasticity of demand—in the case of sugar—was negative in sign and comparatively insignificant in magnitude. One would have expected that increases in income would have led to an increase in per capita consumption. The contrary was the case in all three periods. If sugar were an inferior good, this might be explained by Professor Hicks's analysis, but sugar

is not an inferior good. Either the result indicates mathematical error, which Professor Schultz admits is possible, or the aggregate income hides changes in other prices, which might cause a reduction in sugar consumption. Herein is the admitted weakness of using a Marshall type function. We do not know what changes in other prices are hidden in w and/or t which may affect x in a manner contrary to what we should otherwise expect.

Professor Schultz tested his equation against the data 1922-1936 with the following result.

$$x = 108.83 - 6.0294y + .1644w - .4217t.$$

This shows the change in trend anticipated by the t value of the data in 1915-1929 and gives the expected positive correlation of changes in income and consumption.

Professor Schultz applied the same method to the analysis of the demands of various other commodities. His results are summarized below.

The Demand for Corn

1879-1895	$x = 49.2472 - .4398y - .0611t$	$n = -.717$
1896-1914	$x = 48.0692 - .3468y + .2929t$	$n = -.595$
1915-1929	$x = 37.1718 - .2348y - .5377t$	$n = -.484$
(1917-21 excluded).		

No significant change appeared when the equations were corrected for income, so the variable w is not introduced.

The Demand for Cotton

1875-1895	$x = 15.3392 - .8514y + .0989w + .2588t$	$n = -.509$
1896-1913	$x = 14.8649 - .5740y + .1650w + .3478t$	$n = -.249$
1914-1929	$x = 6.4026 - .2727y + .2366w - .2100t$	$n = -.124$

In the earlier periods there was trend towards increasing consumption, reversed in the last period when cellulose fibres were introduced. Cotton demand became much more inelastic to price change, but more sensitive to income fluctuation.

The Demand for Hay

A	$x =$ Consumption per animal	
1875-1892	$x = 4.6193 - .1968y + .0044t$	$n = -.785$
1899-1914	$x = 4.6612 - .1550y + .0070t$	$n = -.622$
1915-1929	$x = 5.2899 - .1993y + .0845t$	$n = -.463$

The usual growing inelasticity appears. Trend is completely unimportant.

B x = total consumption, z = numbers of animals.

1875-1892 $x = 30.1923 - 2.9523y + 2.6006z$ $n = -.740$
1899-1914 $x = 35.4110 - 3.7907y + 3.1425z$ $n = -.657$
1915-1929 $x = 174.0251 - 5.6379y - 1.3153z$ $n = -.522$

Animal population appears, therefore, as the chief temporal cause of fluctuations in the consumption of hay. This links the demand for hay with the demand for meats and other animal products. In some periods this influence is as great on consumption as price shifts.

The Demand for Potatoes

1875-1895 $x = 4.9418 - .0336y + .2227t$ $n = -.682$
1896-1914 $x = 5.5757 - .0312y + .0487t$ $n = -.542$
1915-1929 $x = 4.4403 - .0155y + .0136t$ $n = -.317$

Trend towards increasing per capita consumption in this case resembles the trend in sugar consumption. Probably the observed trend is now being reversed with dietary changes.

3. Conclusions

I wish now to draw briefly certain general conclusions. They must be regarded with some reservation. There is always a tendency towards ingenuousness on the part of non-mathematical economists, such as most of us are, towards the work of the econometrician. The habit of exactitude, precision and certainty, in which mathematics invests an argument, leads us both to accept the conclusions uncritically, and to be ready to suppose the conclusions to hold with great generality in the real world, which they seldom do. In the present case it is clear that the commodities whose demand has been studied are few in number, far from typical in kind, and that even in this limited area the conclusions rest on some highly simplifying assumptions and, in certain cases (as in the income effect on the demand for sugar), the mathematical argument itself is subject to query. Professor Schultz's work is experimental and trail-blazing. It does not propose to be a concrete high-road to absolute truth. It is to be hoped that others will con-

tinue to develop and refine his methods and extend their application to a wider variety of commodities. In the meantime the theoretic economist may well engage in a little tentative stock-taking.

A—The first conclusion may be said to be definite rather than tentative. The statistical verification of the downward slope of the theoretic demand curve, while interesting and not without some value—it enables a teacher, for example, to illustrate to a class of scoffers how deductive analysis often reaches true conclusions much more simply than laborious induction—is largely a work of supererogation. Ordinary theoretic demand analysis had demonstrated the shape of the demand curve to the satisfaction of all but the "vulgar" who hold "that the safe methods on political subjects are those of Baconian induction", a view which Mill adds "will one day be quoted as among the most unequivocal marks of a low state of the speculative faculties".[22] Far more important and, because based on selected and slight empirical evidence, less certain, is what Professor Schultz's argument can tell us about elasticities and temporal shifts in demand and elasticity.

B—The empirical measurement of elasticities, though statistically less satisfactory, is probably of more significance. The isolation of the income effect enables Professor Schultz and other investigators to find negative elasticities more or less corresponding to the Marshallian concept. This is in distinction to Professor Moore's measurement of the elasticity of the demand for pig-iron, which was positive, reflecting, as Professor Pigou pointed out,[23] positive temporal shifts in the demand curve rather than the elasticity of an instantaneous curve. Professor Schultz's calculations show low elasticities, i.e., of less than unity, for the agricultural produce whose demand he measured. This also agrees with the expectations of theoretic economics, but in this case the theoretic conclusions were based on an estimate of the probabilities of human behaviour and

---

[22] All three phrases in inverted commas are from J. S. Mill's famous denunciation of empiricism in his *System of Logic* (London, 1884), Ch. X.

[23] A. C. Pigou, *The Economics of Welfare* (London, 1920), App. II.

were not rigorously demonstrated. Hence the statistical argument contributes to the extent and certainty of knowledge.

An interesting phenomenon is the tendency of elasticities to diminish with time. This is true for the demand of all commodities studied. In the case of cotton elasticity of demand became as low as $-.124$, barely one-eighth of unity, in the period 1919-1929. This declining elasticity towards extraordinarily low magnitudes seems to be a temporal process and a general characteristic in the modern period of the demands of such products as are here studied. Is this because of growing rigidity in consumption behaviour, or because increasing incomes have resulted in consumers' reacting more to income changes than to price changes with respect to the purchase of standard commodities such as these, or because these commodities have tended to form a less and less important part of consumers' expenditures so that changes in their prices are negligible in relation to consumers' outlay and the demands in consequence have become increasingly inelastic? Or is it some combination of these causes? The increasing inelasticity of the demand for cotton is particularly surprising in face of the growth of competing goods based on wood pulp and other cellulose fibres.

The practical significance of this growing inelasticity of demand for standard agricultural produce is that it means that a "bumper crop" will yield a lower total income to farmers than a small crop. It thus comes to have great importance for those responsible for planning agricultural production and marketing policies. From our point of view it strongly suggests a long-run secular trend in demand behaviour not caused by population changes, and related only indirectly to innovations. Clearly innovations of new goods and the increase, from innovations, in real income, may well be connected with the decline in the elasticity of the demand for agricultural staples.

C—This reflection leads directly to our third conclusion, which is the great importance of trend generally, so far as the commodities studied are concerned. So much stress in recent times has been laid on cyclical fluctuations that it is interesting to discover, in the case of sugar, that the secular trend in de-

mand, *corrected for population change*, had, in the period 1875-1895, eight times as strong an effect on consumption as the change in income. That means that the passage of one year of time changed the per capita consumption of sugar by eight times as much as did the alteration of one point in the index of industrial production. Indeed, the figures indicate that the passage of three years was sufficient to offset the increase of one unit in price. This means, to take a rough arithmetic example, that if sugar rose from three to six cents per pound during nine years (within this period) the per capita consumption, at the end of the nine years, would have been approximately the same as at the beginning. In the subsequent period (1896-1914) per capita consumption would have increased by one pound for every year that passed and would have fallen off by .45 pounds with every increase of one cent in price. Thus if price in ten years rose from six to ten cents the price effect on per capita consumption would have been a decline of 1.8 pounds, which would have been more than offset by the trend effect of an increase of ten pounds. In spite of a 40% increase in price, per capita consumption would have increased by 8.2 pounds. This indicates the very great practical importance of the temporal shift in demand. In the third period (1915-1929) the temporal change in the demand for sugar was unimportant, but this is almost certainly because the trend reversed itself, changing its sign within the period. The evidence of this is that the t value for 1915-1929 is $+.0969$, while for the period 1922-1936 it was found to be $-.4217$. Thus the low positive value 1915-1929 disguises a higher but declining positive value, succeeded, some time in the 1920's, by an increasing negative value. Trend values for the other commodities are less striking, but they are in all significant, except in the case of hay, where the important temporal change is shown to be the variation in animal population. This refers the temporal change of human consumption to the demand for animals and animal produce. We do not attempt to guess the various causes of these temporal changes and of the reversals in their direction, though inquiries into their probable causes might well be commended as fruitful fields of inquiry to research

students in economics and statistics. Our present purpose is to stress the importance of the temporal shift in demand, and to indicate that such changes, great with significance to the general process of economic change, are not lightly to be dismissed in order to frame wide speculative hypotheses about population and innovations. It is indeed most probable that, independently of population, modern society is consuming more and more of certain commodities and less and less of certain others and this process requires a continuous alteration of the price structure in order to maintain some kind of equilibrium. Since the shift is normally against agricultural staples, and since it includes a growing inelasticity in their demands, the process implies a reduction in agricultural incomes and suggests the disequilibrating effects of a rigid price structure.[24]

D—A further, highly tentative, conclusion is the relative unimportance of income changes in affecting demand for products such as these we have been considering. This runs counter to the present emphasis of theoretic economics, and must be regarded with some suspicion. It may be partly because of the nature of the staple commodities selected for study. It may come partly from the admittedly rather unsatisfactory mathematics of the treatment of the income effect. Further, the periods are chosen so as to smooth out cyclical income effects and may therefore fail to do justice to the full effect of violent income changes in the course of the ten-year cycle. Finally it should be noted, in spite of the preceding observations, that income effects become more marked in the later than the earlier periods, suggesting that the cyclical fluctuations became more important in their effects on demand in the later than they were in the earlier periods. In spite of all these qualifications, however, we should do well, in the light of these findings, to reconsider, so far as agricultural produce is concerned, the

[24] If the price structure is flexible, however, the agricultural population must sustain a continuous adverse income effect. This, in turn, will have short-run "aggregative effects" of an adverse sort. It is clear that a policy problem of great importance is posed. At the present stage of our work the discussion of this problem would be premature. We might point out, however, that it is again illustrative of the dangers of oversimplifying the problems of economic change by concentrating on certain very general aggregative aspects.

relative emphasis we place on cyclical income fluctuations and the long-run secular trend.

E—We have said little in the course of Section 2 about the question of inter-related shifts in demand. Much attention in the recent literature has been paid to this question. In this connection the work Professor Schultz has done has a most important theoretic bearing. He has been able to show (in statistical argument which for reasons of brevity we did not attempt to summarize in Section 2) that the general Slutzky-Hicks theory of competing and complementary goods is satisfied by the statistical evidence, but that the indirect income effects even in such important items of expenditure as meats is insignificant. "It is probable that the income effect is also small for most articles of wide consumption on which only a small proportion of the income is spent".[25] This lends weight to Professor Knight's scepticism as to the relevance and realism of the newer developments of demand theory.[26] Professor Schultz, however, emphasizes the need for further empirical investigation in this field.

F—We note finally, and very briefly, what to some may appear the most important conclusion of all to come from this type of inquiry, viz., the practicability of extrapolation of curves of demand so as to forecast probable future consumption at contemplated or anticipated prices. Clearly if one knows elasticity and trend and if income effects may be disregarded, one should be able—or so the practical man will argue—to say what future consumption will be at any given price. There is something in this, but not much to give comfort to the business man of the private enterprise economy. If we were a world planner, or the government planner of a closed socialist state so that all supplies for a self-contained market were simultaneously determined together, this kind of extrapolation might have some value. It is not very clear, however, what value it could have for a single enterpriser who would have no knowledge of the production plans of his competitors. Even in the case of the socialist planner the extrapolation has dangers.

[25] Schultz, op. cit., p. 646.
[26] See Knight, op. cit.

It is based on the assumption that the causes of the *trend* are fixed, so that indefinite extrapolation is safe. This, as the evidence shows, is not true. Causes of the temporal shifts alter, and the temporal trend is capable of change in magnitude and direction. Until these causes of change are analysed and understood and their probable effects estimated, policy based on prediction, which in turn is based on extrapolation, must be highly dangerous.

PART IV

MODELS OF THE FIRM

CHAPTER X

THE TRADITIONAL THEORY OF THE EXPANSION OF THE FIRM

In this Part we turn to the single firm to examine how it behaves under conditions of temporal change. The traditional theory of the firm has always contained a certain difficulty in market analysis because it supposes instantaneous adjustments to equilibrium positions, whereas the explanation of the firm's process of adjustment, in terms of the process of altering employment and output until equilibrium is reached, always involves time. Thus Marshall tells us that if firms taken together put more on the market than can be sold at the equilibrium price, i.e., more than can be sold at a price to cover costs, they will, in the next period, reduce output. If they reduce their output below the equilibrium amount, abnormal profits will then lead to an expansion of the output in the subsequent period. It becomes necessary, therefore, to set up the assumption of a stationary state, from which all other temporal changes are abstracted, and to study the process of the firm's expansion and contraction in an empty time. The difficulty of saying precisely what is left of the time process has led to discussion and controversy over the nature of the firm's expansion path and over such questions as the reversibility of supply curves,[1] the existence of multiple points of equilibria and their relative stability,[2] and the stability or metastability of such equilibrium as is achieved.[3] These difficulties, I believe,

[1] See S. Slichter, *Modern Economic Society* (New York, 1928), pp. 294-98; K. E. Boulding, *Economic Analysis* (New York and London, 1941), p. 441.

[2] See A. Marshall, *Principles of Economics* (2nd ed., London, 1891), pp. 491-92; J. Robinson, *Economics of Imperfect Competition* (London, 1933), pp. 57-58, and others.

[3] See L. Robbins, "On a Certain Ambiguity in the Conception of Stationary Equilibrium", *Economic Journal*, Vol. XL, 1930, p. 194; D. H. Robertson, "Increasing Returns and the Representative Firm", *Economic Journal*, Vol. XL, 1930, p. 86; N. Kaldor, "The Equilibrium of the Firm", *Economic Journal*, Vol. XLIV, 1934, p. 60; A. P. Lerner, "Statics and Dynamics in Socialist Economics", *Economic Journal*, Vol. XLVII, June, 1937; J. Viner, "Cost Curves and Supply Curves", *Zeitschrift fur Nationlökonomie*, III, 1931-32, p. 23; and others, as cited infra.

arise from differing assumptions about the nature of time, the choice of which is, of course, arbitrary. In the first chapter of this Part, we shall review the traditional theory of the expansion of the firm, and in the second chapter we shall attempt to select a satisfactory assumption as to the nature of time, such as will permit a resolution of some of the problems of inter-equilibrium adjustment and afford further knowledge of the nature of change in modern economic society.

1. The "Expansion Path"

(I have thought it wise, because of the comparative neglect of this field in most text-books, to provide an elementary review of it here, which I hope will be helpful to some students. All readers familiar with the Frisch-Carlson theory of the expansion path will wish to skip direct to Section 2, which they may do without loss of continuity.)

We begin with an account of the theory of the expansion path of the firm, that is the theory which describes how a firm increases its employment of the factors in order to obtain greater output. There are two problems involved, one a problem of minimization, the minimization of total cost for any given output, the other a problem of maximization, viz., that of maximizing total profit by appropriate adjustments of minimum cost outputs to meet a given demand.[4] The latter problem and its traditional treatment we shall assume to be familiar to the reader and, in this chapter, shall treat it only incidentally as its relationship to the minimum cost problem requires. To the solution of the latter we now turn.

Let us consider the simplified case of a firm employing only two factors. The principles discovered within such a model can readily be modified to cover a model in which more than two factors are employed. These two factors (we shall call them "input services" or "inputs") may be technically related to one another in one of three ways. They may be absolutely

[4] For a precise treatment of the economic problems of minimization and maximization in mathematical form, see R. G. D. Allen, *Mathematical Analysis for Economists* (London, 1938), pp. 351-84.

complementary or non-substitutable, they may be perfectly substitutable, or they may be partly or imperfectly substitutable.

Where the two factors are non-substitutable there is only one combination of the factors possible to produce any given output. In other words, there is a technically fixed ratio in which they must be employed. In this case no economic problem is involved, because the entrepreneur has no economic choice in the combination of the factors. His minimum cost combination must be the only such combination technically possible. Such a situation may arise in chemical manufacture. We might choose, however, a humbler illustration from the kitchen. In the case of baking a certain cake two eggs must always be combined with exactly one cup of shortening. Whatever the size of the cake these two inputs must bear the same relation, so that, if we want twice as big a cake, four eggs must be combined with two cups of shortening, and so on. In the case of perfect substitutability the converse relationship exists. In this case, a given output can be produced by varying the two inputs in inverse proportion. In making our cake, for example, cream of tartar or baking powder may be used. We may use a tablespoonful of baking powder or a tablespoonful of cream of tartar or half and half. Every reduction in the input of baking powder can be exactly compensated by a proportionate increase in the input of cream of tartar. If we wish to double the size of the cake we double the total use of leavening agents, but we have perfect substitutability of the one for the other.[5]

It is clear that neither of these cases is usually to be found in the industrial world. There, the prevalent case is of imperfect substitutability. In this instance to obtain any given output some quantity of each factor must be employed, but a certain substitution of one for the other, by varying the proportions is possible within certain limits. Thus in baking our cake, we must use at least one egg-white and one cup of milk, and we

[5] I am warned by my wife that, though this may appeal to me as a good illustration of the economics of perfect substitutability, my readers should be put on guard against attempting this substitution in the kitchen. It appears the culinary facts are not quite as I have stated them.

must use in addition further egg-white and milk. We might use one whole additional cup of milk and no further egg-white, or we might use only three-quarters of a cup of milk additional and add a further egg-white, or a half cup of milk and three egg-whites, or no additional milk at all and eight egg-whites. It will be seen here that every reduction in the application of milk required an increasing increment in the application of egg-whites. This may not be sound in a culinary way, but, as we shall show, it is sound economics in the industrial world.

Let us, however, elect another example. We wish, say, to produce 1,000 bushels of wheat. To produce it we must have some land and some labour, because it is quite impossible to produce wheat using only land and no labour or only labour and no land. Let us suppose that one man-year could produce the 1,000 bushels by the extensive cultivation of 100 acres of land. The same total output could be produced by two men-years applied to a lesser area of land. We have to ask whether two men would require a full 50 acres to produce the same total quantity of wheat. If they did we should have a case of perfect substitutability, a proportionate increase in one factor exactly compensating for a proportionate diminution in the other. This, however, would not happen. The factors are not perfectly substitutable. When one man tried to cultivate 100 acres, land was under-employed, so that when two men cultivate 50 acres an increased total product would result. Something less than 50 acres would be required to give the same total product. This follows from a simple application in reverse of the law of diminishing returns. The application of a large number of units of land to a single unit of labour had resulted in diminished returns to labour. Increasing the employment of labour in proportion to land consequently yields increasing returns. As we continue the substitution we begin, after a certain point, to get diminishing returns to land, so that we must apply increasingly larger "doses" of labour to replace the units of land up to the limit where no quantity of labour would maintain the total product if a further reduction in acreage occurred. We might imagine the substitution to run something like this:

Labour	*Land*	*Total Product*	*Marginal Rate of Substitution*
1	100	1,000 Bushels	
2	40	1,000 Bushels	60/1
3	20	1,000 Bushels	20/1
4	15	1,000 Bushels	5/1
5	13	1,000 Bushels	2/1
6	12	1,000 Bushels	1/1

We would graph this as follows:

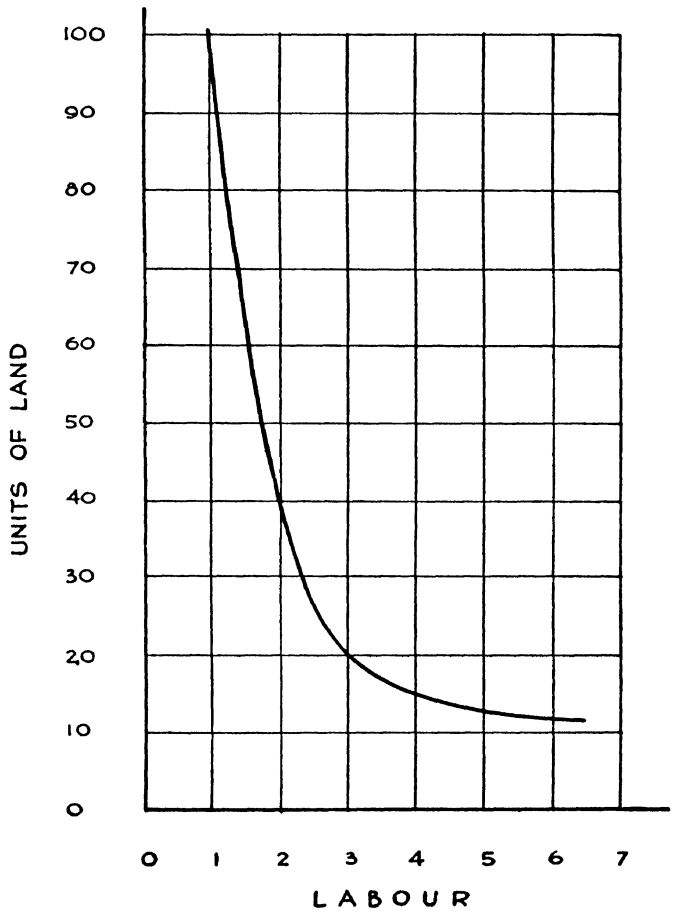


FIG. 10

We call this curve an *iso-product* curve. It represents all possible combinations of the two factors that will give the same total product. It must be convex to the origin because of the law of diminishing returns. If it were concave to the origin it would mean that when one factor was intensively employed and one factor extensively employed a diminution in the employment of the latter could be compensated for by a smaller proportionate increase in the former, which is nonsense. The iso-product curve is not a rectangular hyperbola, that is, it does not tend to approach the asymptotes at infinity. On the contrary, the two extremes of the curve become straight lines, parallel to the two axes, indicating that further substitution is impossible. After a certain point the addition of further units of labour could not replace any reduction in land and at the other extreme of the curve further use of land could not compensate for any reduction in labour. Graphically the extreme case of perfect non-substitutability would be represented by two straight lines, one perpendicular to the other like this:

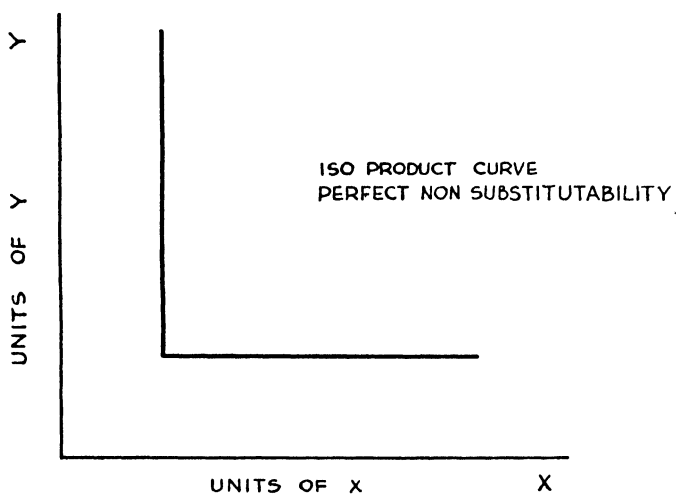


FIG. 11

Similarly, the case of perfect substitutability, where diminishing returns are absent, would look like this:

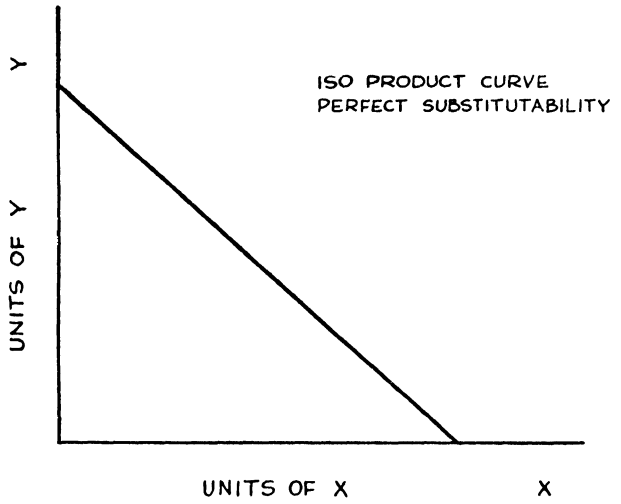


FIG. 12

The slope of the iso-product curve at any point measures the marginal rate of substitution and since the slope of a straight line is constant, this means that the marginal rate of substitution is a constant in the case of perfect substitutability. Where there is imperfect substitutability, the usual case, the marginal rate of substitution of factor Y for factor X, the slope of the curve, is a diminishing rate. In the case of no substitutability the marginal rate of substitution is zero.

Any given physical output for a firm will have a corresponding iso-product curve. Thus if we imagine output to be increasing along the axis of the third dimension OZ, we may graph the variables as in Fig. 13.

In Figure No. 13 the total output is measured upwards along the vertical axis OZ, rising perpendicular to the plane of the OX-OY axes. Thus the picture of total output is the solid mound $L\ N\ M$. The iso-product curves are consequently the contours of this mound showing the possible factor combinations for each height, or total product. They may, therefore, be set in a simple two-dimensional contour graph, as in Figure No. 14 and we may call this an iso-product map.

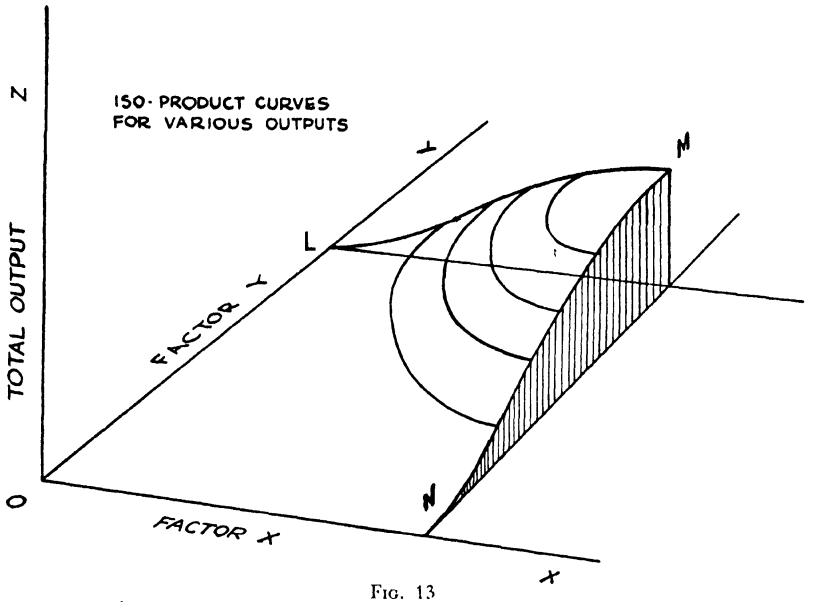

ISO-PRODUCT CURVES
FOR VARIOUS OUTPUTS
Z
TOTAL OUTPUT
0
FACTOR Y
FACTOR X
Y
X
L
M
N

FIG. 13

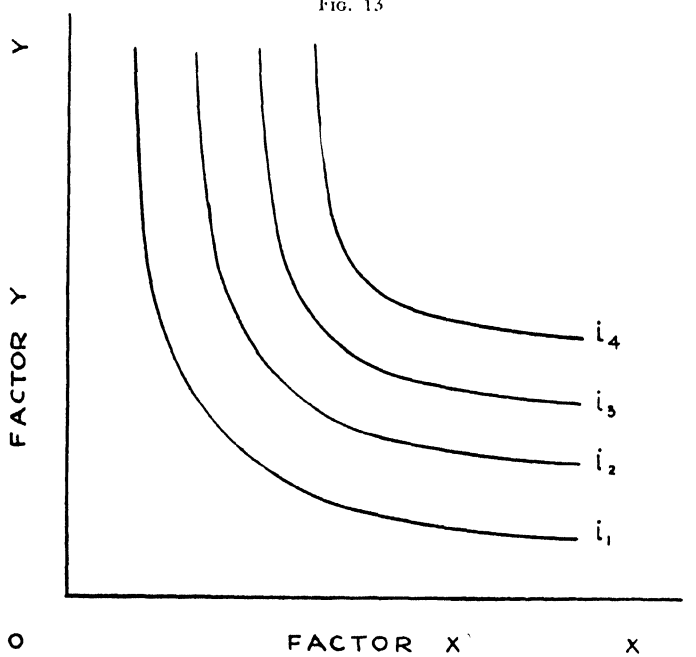

Y
FACTOR Y
0
FACTOR X
X
i_4
i_3
i_2
i_1

FIG. 14

The technical matters which determine the possibility of substituting one factor for another are no concern of the economist. His problems are to explain how, given the limiting technical conditions, the combination of the factors is determined by economic forces and how the level of output (and employment) is determined.[6] The degree of curvature of the iso-product curves is a given to the economist: it is determined by purely physical and technical forces. Thus, from an iso-product map, by itself, we cannot say how the factors will in fact be combined.[7] If, however, we know the costs of employing the factors, we can determine the quantities and proportion in which they will be employed. If perfect competition exists in the factors' markets, then the price which must be paid for each input is given to the single firm. Within the same set of axes as exist for the iso-product curves, we may trace a contour map of iso-cost curves. An iso-cost curve traces the various combination of factors that can be employed for the same total cost. Now, if the prices of the factors are fixed on the market so that one unit of X always costs the employer \$2 to employ, and one unit of Y \$1, the marginal rate of cost substitution remains unchanged and the iso-cost curve will have a constant slope, that is, it will be a straight line. Thus, under perfect competition, an iso-cost contour map will consist of a series of parallel straight lines, representing the constant market exchange ratio of the two factors for different total cost levels. If we turn our physical product into value product by multiplying the former by the price, we have both our iso-product and iso-cost curves measured in the same unit. The iso-cost curves will, somewhere, lie as tangents to the iso-product curves. This may be represented as in Figure 15.

The straight line "c" curves are the iso-cost curves lying in tangency to the "i"—iso-product—curves. The curve Q-Q^1 is the "expansion path" of the firm. This path is defined as

[6] Compare R. Triffin, *Monopolistic Competition and General Equilibrium Theory*, p. 110.

[7] Except in the case of non-substitutability, where there is only one technically possible combination. In such a case the entrepreneur has no economic problem of combination. Regardless of costs he must employ the factors in the given combination. His only economic problem is to determine the level of output.

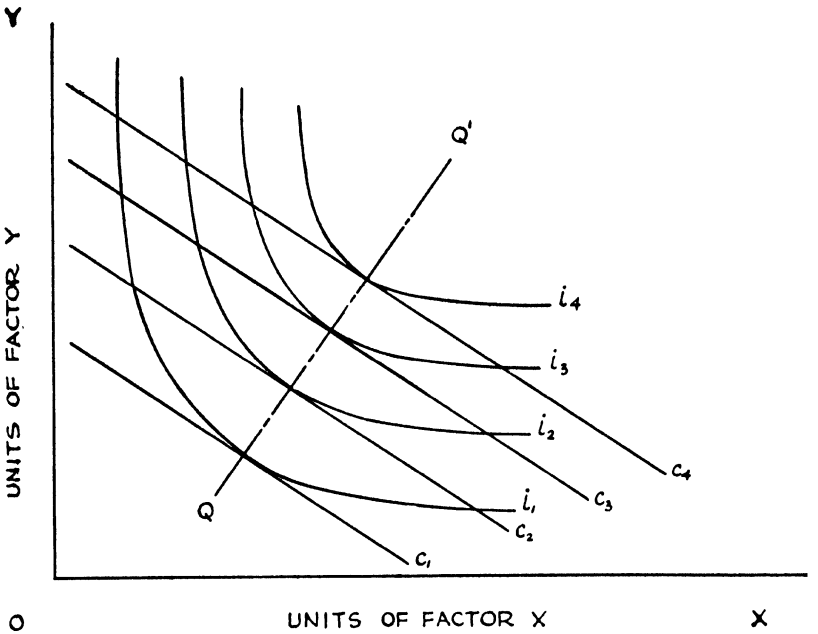


FIG. 15

tracing the loci of the minimum cost of combinations of the factors over an increasing series of outputs. The factors must necessarily be employed in the proportions indicated by the points of tangency. At any point of intersection the iso-cost curve must be cutting an iso-product curve at a lower level of output, i.e., the point of tangency marks the lowest total cost for which that total output can be produced, or, if you wish, the highest total output which can be achieved for that total cost. No entrepreneur could remain satisfied under any alternative condition. Only at a point of tangency can he obtain the minimum cost combination for a given level of output. This point of tangency, in concrete economic terms, defines the point where the ratio of the marginal factor costs (the marginal rate of cost substitution) is equal to the ratio of the marginal value productivities (the marginal rate of factor substitution), i.e., the point where the slopes of the two curves are the same. Hence the firm will expand along the curve which traces the loci of the points of tangency. This condition can be generalized

for the employment of more than two factors, though it cannot then be represented graphically.[8] In the simplest possible notation, if we call the factors w, x, y, z, etc., the marginal products of P_w, P_x, P_y, P_z, etc., and the marginal costs C_w, etc., the equation of combination would read:

$$\frac{P_w}{C_w} = \frac{P_x}{C_x} = \frac{P_y}{C_y} = \frac{P_z}{C_z} = \text{etc.}$$

In Figure No. 15 the expansion curve Q-Q^1 is a straight line. The conditions necessary for it to be a straight line are that the iso-product curves must always be of equal curvature to one another at the points of tangency (i.e., that the isocline be a straight line) and that the iso-cost curves be straight lines. The first condition means that there must be no innovations or technical changes affecting the quality of the inputs, or economies of scale in the use of one of the factors, any of which would affect the physical productivity of either factor. The second condition means that there must be perfect competition.

---

[8] "At the substitution margin a unit increase in the input of a service (v_k) will have a certain cost which we shall call its marginal unit cost (c_k). The addition in the quantity of output caused by a unit increase of the individual service we have earlier defined as the marginal productivity of the service (ϕ_{v_k}). The relationship between the marginal unit cost of the productive service and its marginal productivity gives us the rate of change of the costs of production as related to a unit increment of a single service. This relationship we shall define as the cost-productivity ratio of the service.

$$x_{c_k} = \frac{c_k}{\phi_{v_k}}$$

"Now it is evident that the substitution of one service for another service . . . will take place as long as the cost productivity ratio of the former service is smaller than the cost productivity ratio of the latter. Or expressed in a different way, as long as the marginal productivity of a service in relation to other services is relatively larger than its marginal unit cost, it is profitable to substitute this service for other productive services. . . . The minimum cost combination for a given output is thus obtained when the marginal unit costs of the different productive services are proportional to the marginal productivities of the services; that is when the cost-productivity ratios of the different services are equal.

$$C_1 : \ldots\ldots : C_n = \phi_{v_1} : \ldots\ldots : \phi_{v_n}$$

$$xC_1 = \ldots\ldots = xC_n$$

Sune Carlson, op. cit., pp. 32, 33. Dr. Carlson's treatment is based in part on the work of Dr. Hans Frisch, and English readers must be grateful to him for making the Frisch expansion thesis available to them.

This is simply to repeat for the case of the productive function of a single firm what we said in an earlier chapter about the production function in general. In order for it to be a straight line—constant returns to scale—perfect competition and static conditions must be assumed.

If perfect competition in the factors' markets does not exist, then the alteration in the employment of any factor by the firm will affect the rate of reward. This means that the iso-cost curves will be bent. Since, as you employ more of factor X and less of factor Y, the reward that must be offered X will increase, and that offered Y may be decreased, the curves must be concave to the origin, that is, they will have the opposite curvature to the iso-product curves.[9] These iso-cost curves will also be in tangency to iso-product curves and, as before, the expansion path of the firm lies along the points of tangency. See Figure 16.

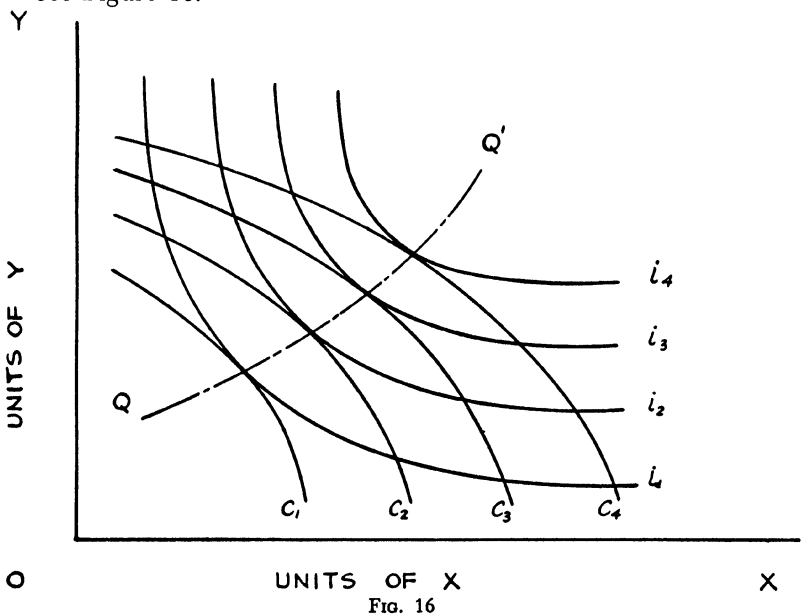


FIG. 16

[9] The convexity to the origin of the iso-product curves is determined by the diminishing marginal productivity of a factor as it becomes more intensively used. By opposite argument the concavity of the iso-cost curves can be shown to depend on the increasing marginal factor cost as the factor is increasingly employed.

The slope of the Q-Q^1 curve will depend on the relative elasticities of supply and of substitution of the factors X and Y, and will be the steeper the less elastic Y is in supply relative to X.

So far we have been dealing with the expansion path within a self-contained time period.[10] The self-contained or monoperiodic time area is a further limitation of the stationary assumption. Not only are no changes in tastes, population, techniques of production and so forth admitted to the model, but the process is conceived as being initiated and fully completed within the period without reference to the production process of previous or subsequent periods. In order to set the theory of the firm in a proper temporal order, even on the stationary assumption, it is necessary to link the periods so as to account for the continuous nature of the productive process.[11]

We shall define a unit of time as the period of the productive process, i.e., as the length of time required to bring the raw material from the moment of its entry to the process to its finished state. (The "finished state" is the state in which it is marketed by the firm, not necessarily a state of readiness for final consumption.) Some of the factors will be durable in the sense that their productive services will not be entirely consumed in the output of any one time period, others will be non-durable in the sense that they are entirely consumed by the production in a single unit of time. Machines are clearly durable inputs, in this sense, most raw materials, non-durable inputs. It should be noted that, while most durable services are "fixed" and most non-durable services "variable" in the sense that the cost of the former is supplementary cost and of the latter is prime cost, this relationship is by no means general. Labour is a variable but durable factor. Indeed the two classifications are based on quite different differentiae, the durable, non-durable classification depending on the physical fact of exhaustion or non-exhaustion of the input in a single productive

[10] It is called "monoperiodic" analysis by Dr. Carlson.

[11] We shall be dealing in the text only with the case, typical of manufacturing, of a continuous (in time) flow of inputs and outputs. In some sorts of production, agriculture, for example, each production period is distinct and the productive process is more or less self-contained within the period. It is this characteristic that lies at the root of such phenomena as the hog and poultry cycles.

process, while the fixed, variable classification depends on the shape of the unit cost curve of the factor in use.

We may illustrate the relationships within the first productive process over a group of time periods by the following diagram. The subscripts identify the time periods, t_0, t_1, t_2, etc.

Now, assuming that we are dealing with a process in time but on the stationary assumption which excludes all other temporal changes, the conditions determining the combination of the factors and the expansion of the firm are essentially the same as in the case of monoperiodic production. The modification which must be made is only that necessary to allow for the employment of the factors in alternative time periods. In order to avoid the awkward problem posed by the existence of durable factors we may think of the intensity of their use in any time period as being measured by the flow of productive services obtained from them. We may speak of the input of productive services and dismiss the distinction between durable and non-durable factors.

If we think of the problem of distributing the inputs of the productive services of any one factor between any two time periods, we see that the same principle applies as that which governs the combination of two factors in any one time period. The factors must be used so that the ratio of its marginal productivity to its marginal cost in one period is equal to the ratio of its discounted marginal productivity to its anticipated marginal cost in another.

If P_{w_0} is the marginal productivity of the service input from factor w in time o, and C_{w_0} the marginal cost of that input in the same time, then $P_{w_1} - hP_{w_1}$ is the discounted marginal productivity of the service input of the same factor in time 1, when h is the rate of interest, and $(C_{w_0} + rC_{w_0}) = C_{w_1}$ is the anticipated marginal factor cost of this input in time 1, when r is the rate of increase of cost in time from interest and storage charges (that is $r > h$, because it includes h and other charges), then the factor will be distributed in its employment over time so as to satisfy the equation:

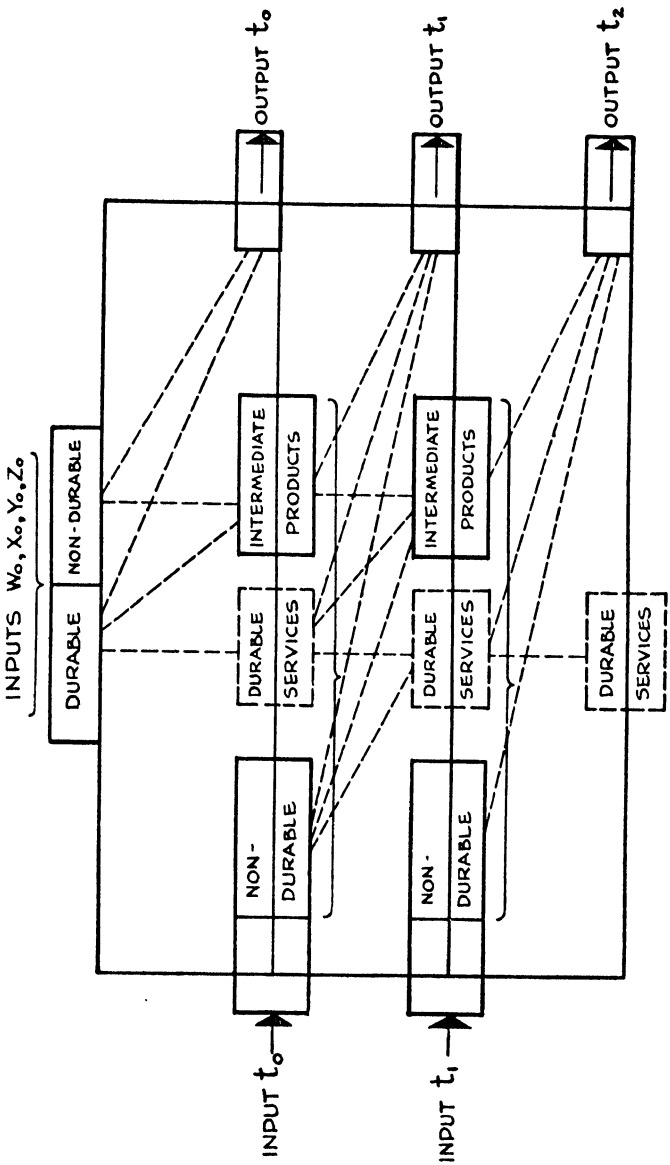


FIG. 17.[12]

[12] This diagram is adapted with slight changes from Fig. 17, Sune Carlson, op. cit., p. 107.

$$\frac{P_{w_0}}{C_{w_0}} = \frac{P_{w_1}\ (1-h)}{C_{w_0}\ (1+r)}$$

Similarly it may be shown that the factor χ will be employed so that

$$\frac{P_{x_0}}{C_{x_0}} = \frac{P_{x_1}\ (1-h)}{C_{x_0}\ (1+r)}$$

and so on with the other factors.

We have already shown, however, that in any period of time, t_0, the inputs of w and x will be adjusted so that

$$\frac{P_{w_0}}{C_{w_0}} = \frac{P_{x_0}}{C_{x_0}}$$

It therefore follows that, over time, the equation for the combination of the factors must be

$$\frac{P_{w_0}}{C_{w_0}} = \frac{P_{x_0}}{C_{x_0}} = \frac{P_{w_1}\ (1-h)}{C_{w_0}\ (1+r)} = \frac{P_{x_1}\ (1-h)}{C_{x_0}\ (1+r)} =$$

etc. [13]

Thus the firm, under the conditions here assumed, will increase its employment over time, so as to obey the conditions defined by the equations of minimum cost combinations.

---

[13] I have not been able in this section to follow the rule I have tried generally to observe in this work, viz. to exclude all mathematical symbols from the text. I have, however, used a highly simplified notation, one which unfortunately obscures the cost minimization aspect of the problem of combination. If we return to the notation used in fn.9, following Dr. Carlson, when $v_1 \ldots v_n$ are the service inputs and X the product, we may distinguish the temporal inputs as $v_{I_1} \ldots v_{I_1}$ in time I (output x) and as $v_{II_1} \ldots v_{II_n}$ in time II (output x_{II}). Then, "granting all our simplifying assumptions, it follows that . . . the output of one period depends both upon the input of variable services and upon the output of the other period. . . . We may therefore write the products x_I and x_{II} as functions.

$$x_I = \phi\ (x_I)\ \ (v_{I_1} \cdots v_{I_n},\ v_{II_1} \cdots v_{II_n},\ x_{II})$$

and

$$x_{II} = \phi\ (x_{II})\ \ (v_{I_1}, \cdots v_{I_n},\ v_{II_1} \cdots v_{II_n},\ x_I)$$

"The partial derivatives of these functions give us . . . the marginal productivities of the services with respect to the two products

$$dx_1 = \frac{\partial x_I}{\partial v_{I_1}} \cdot dv_{I_1} + \ldots + \frac{\partial x_I}{\partial v_{II_n}} \cdot dv_{II_n} \quad (x_{II} = \text{Constant})$$

and

$$dx_{II} = \frac{\partial x_{II}}{\partial v_{I_1}} \cdot dv_{I_1} + \ldots + \frac{\partial x_{II}}{\partial v_{II_n}} \cdot dv_{IIn} \quad (x_I = \text{Constant})$$

. . . "The prices of a certain service v_k in the two periods we write

$$q_{Ik} = q_{Ik} \ (v_{Ik}, \ v_{IIk}) \text{ and } q_{IIk} = q_{IIk} \ (v_{Ik}, \ v_{IIk}).$$

If, at the substitution margin, the service in the first period is changed by one uni t this change will affect the cost of the service not only in that but also in the second period, and *vice versa*. In order to find the marginal unit cost of the service we have, therefore, to consider the effect of the change on the total cost of the service in both periods, inclusive of interest cost. This total cost we write

$$C_{v_k} = q_{Ik} \ v_{Ik} + \frac{I}{I + i_I} \ q_{IIk} \ v_{IIk}$$

where i_I is the average interest rate of the first period. The partial derivative of this cost with respect to v_I and v_{II_k} give us the marginal unit costs of the service in the two periods.

$$cv_{Ik} = \frac{\partial Cv_k}{\partial v_{Ik}} = \left[q_{Ik} + v_{Ik} \frac{\partial q_{Ik}}{\partial v_{Ik}} \right] + \frac{I}{I + i_{m_I}} \ v_{IIk} \frac{\partial q_{IIk}}{\partial v_{Ik}}$$

and

$$cv_{IIk} = \frac{\partial Cv_k}{\partial v_{IIk}} = v_{Ik} \frac{\partial q_{Ik}}{\partial v_{IIk}} + \frac{I}{I + i_{m_I}} \left[q_{IIk} + v_{IIk} \frac{\partial q_{IIk}}{\partial v_{IIk}} \right]$$

where im_I represents the marginal rate of interest . . . The cost productivity ratios . . . we consequently write

$$x_I Cv_{Ik} = \frac{\dfrac{\partial Cv_k}{\partial v_{Ik}}}{\dfrac{\partial x_I}{\partial v_{Ik}}}, \qquad x_{II} cv_{Ik} = \frac{\dfrac{\partial Cv_k}{\partial v_{Ik}}}{\dfrac{\partial x_{II}}{\partial v_{Ik}}},$$

$$x_I cv_{IIk} = \frac{\dfrac{\partial Cv_k}{\partial v_{IIk}}}{\dfrac{\partial x_I}{\partial v_{IIk}}}, \qquad x_{II} cv_{IIk} = \frac{\dfrac{\partial Cv_k}{\partial v_{IIk}}}{\dfrac{\partial x_{II}}{\partial v_{IIk}}},$$

"On analogy with the case of joint mono-periodic production, the minimum cost combination for a given output is obtained when the marginal unit costs of the different services are proportional to the marginal productivities of the services with respect to each one of the consecutive periods, or, which amounts to the same thing, when the cost productivity ratios with respect to every product are equal . . . "

(*Fn. continued on p. 226*)

2. Ambiguity in the Theory of the Expansion Path

This "traditional theory of the firm" has recently been the subject of acute criticism by Dr. James Dingwall,[14] who argues that there are more than one possible expansion paths which the firm may follow. The theory we have just examined, Dr. Dingwall contends, rests on the assumption that the only "variable independently adjusted by the firm[15] is output and that "on its path to the profit maximizing output the firm will

---

(*Fn. 13 cont'd.*)

$$\frac{\dfrac{\partial C_{v_1}}{\partial v_{\mathrm{I}_1}}}{\dfrac{\partial x_{\mathrm{I}}}{\partial v_{\mathrm{I}_1}}} = \ldots = \frac{\dfrac{\partial C_{v_n}}{\partial v_{\mathrm{II}_n}}}{\dfrac{\partial x_{\mathrm{I}}}{\partial v_{\mathrm{II}_n}}}$$

and

$$\frac{\dfrac{\partial C_{v_1}}{\partial v_{\mathrm{I}_1}}}{\dfrac{\partial x_{\mathrm{II}}}{\partial v_{\mathrm{I}_1}}} = \ldots = \frac{\dfrac{\partial C_{v_n}}{\partial v_{\mathrm{II}_n}}}{\dfrac{\partial x_{\mathrm{II}}}{\partial v_{\mathrm{II}_n}}}$$

Dr. Carlson, by assuming no price interdependence between the periods, is then able to simplify the equation for a durable resource which yields variable services entering into the production of the same product in consecutive periods. "These services will be so distributed between the periods that their marginal productivities are the same. Thus, if $v_{\mathrm{I}k}$ and $v_{\mathrm{II}k}$ are the services of a durable resource, bought at the beginning of the first period, these services will be substituted for one another until

$$\frac{\partial x_{\mathrm{I}}}{\partial v_{\mathrm{I}_k}} = \frac{\partial x_{\mathrm{I}}}{\partial v_{\mathrm{II}_k}} \quad \text{and} \quad \frac{\partial x_{\mathrm{II}}}{\partial v_{\mathrm{I}_k}} = \frac{\partial x_{\mathrm{II}}}{\partial v_{\mathrm{II}_k}}$$

We should note finally, in comparing these conclusions with those in the text, that Dr. Carlson adds, "This relation assumes, however, that there exists no cost for the upkeep or storage of the resources". Also, "if a productive service which has entered into the production of the same product throughout two periods is bought successively in both periods, it will be so distributed between the periods that its marginal productivity in the earlier period exceeds its marginal productivity in the later period by an amount which corresponds to the interest cost".

The quotations in this note are from Sune Carlson, op. cit., pp. 108-14, and are made by generous permission of the publishers, P. S. King and Sons, Ltd., London.

[14] James Dingwall, "Equilibrium and Process Analysis in the Traditional Theory of the Firm", *Canadian Journal of Economics and Political Science*, Vol. X, No. 4, pp. 448-63.

[15] Ibid., p. 448.

attempt to produce each successive output at minimum cost".[16] The argument is then developed, following Dr. Sune Carlson, that this will be done by substituting one factor (input) for another until the marginal cost-marginal productivity ratios are equal. This argument, of course, leads to an expansion path of the sort we have explained and illustrated in the preceding paragraphs.

Dr. Dingwall argues, however, that there are other paths to the maximum profit point. Two of these he develops. If a factor of production is regarded as the given variable in the system the adjustment will be made so that "any given quantity of a given factor will be combined with the 'appropriate' quantities of the other factors. The 'appropriate' amounts of the other factors are defined to be those which will maximize profit under the given conditions."[17] If we were to present

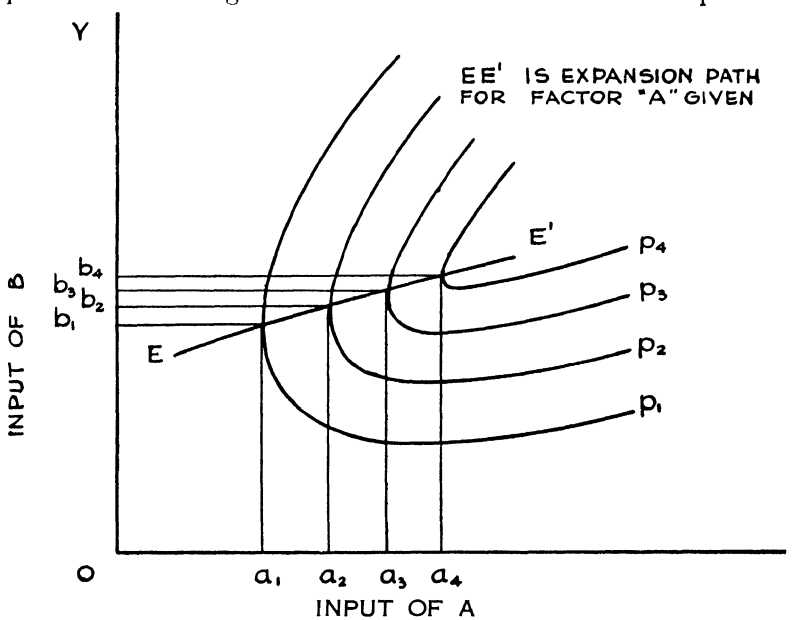


FIG. 18.

[16] Ibid., p. 449.

[17] Ibid., p. 449. Dr. Dingwall cites Joan Robinson, *Economics of Imperfect Competition* (London, 1933), to show that this is the view held by her. It is the logical development from classical assumptions, just as Dr. Carlson's theory emerges naturally from the Lausanne and Austrian schools.

this view graphically, using similar graphs and techniques to those employed in the illustration of the Carlson-Frisch expansion path, we should require "iso-profit" curves, tracing the combination of the two factors yielding the same total profits.[18] In Figure 18 these curves are the p_1—p_4 curves. Now, with A regarded as relatively fixed, "given" in the classical sense, let us take different levels of employment of A, and adjust the employment of factor B so as to give the maximum profit obtainable under the given conditions, i.e., the conditions defined by A being given. The points a_1, a_2, a_3, a_4 along the X-axis represent the successive given quantities of A for successive levels of output. It follows that the appropriate (i.e., profit maximizing) input of B, will be for that quantity of B determined by the point where a perpendicular erected from any given quantity of A is tangent to an iso-profit curve. Only at a point of tangency is profit maximized, for all intersections of the perpendiculars with iso-profit curves must be for lower iso-profits than the curve in tangency. In Figure 18 the "appropriate" amounts of employment of factor B are shown by the intersections on the Y-axis marked b_1, b_2, b_3, b_4. The curve $E-E^1$ is the expansion path of the firm, factor A given. By similar argument a quite different expansion path, factor B given, can be derived on the alternative assumption. Dr. Dingwall has no trouble in showing that these paths cannot coincide, and that neither will coincide with the Carlson-Frisch path, Q-Q^1 in Figure 16 (page 220). Thus, he concludes, "the accepted theory of inter-equilibrium adjustment yields . . . not one maximum-profit path of expansion or contraction but three. The doctrine of the trinity is clearly false! . . . In the absence of some restriction explicitly involving time, there is no limitation on the path which the firm, following the dictates of profit maximization may take on its way to the ultimate maximum-profit combination of the factors, other than that the requirement that each successive change shall increase profit."[19]

[18] These iso-profit curves must, by definition, be tangent to the iso-product curves at the same point the iso-costs are tangents. Thus the Carlson-Frisch expansion path follows the points of tangency of the iso-profit and iso-product curves.

[19] Dingwall, op. cit., p. 453.

In a word, that the entrepreneur will seek the maximum profit point, and that the firm will be in equilibrium there, follow from the assumptions made about entrepreneurial motives but, unless some further assumption is made there is no reason to believe that the firm will follow one path rather than another. One may assume that the entrepreneur will follow the least cost combination of the factors, but there is no necessary reason why he should do so. It may well be that, on empirical grounds, this assumption is justified, because it is easier for the entrepreneur to think of output as his independent variable and to adjust his inputs to this in terms of information about costs, the kind of information he may consider as his most reliable data. It is equally possible, however, that he may regard one input as given, and adjust the other inputs to this. The input may be given in the sense of requiring time for adjustment, or simply in the sense that psychologically the entrepreneur regards it in his thinking as given. In addition to these cases, set out in illustration by Dr. Dingwall, there are an indefinite number of cases, where, lacking a specific assumption about what the entrepreneur regards as adjustable, the path he may follow to maximum profit may be any path, so long as it obeys the one condition that the firm always moves upwards towards the summit of the profit mount.[20] Thus, we cannot believe that what is called "the theory of process" is a theory of process at all It is an elaboration of the conditions of maximum profit. What determines the "process", or the expansion path, is simply a further assumption. A theory of process ought to be, then, an examination of what are, indeed, the factors that determine how a firm does expand over time.

3. The Envelope Curve

Our principal task, therefore, is precisely to select and formulate the "explicit assumption" which we must make as to the nature of time. Now, in a certain sense, the method of the

[20] Mrs. Robinson, *Economics of Imperfect Competition* (London, 1933), Ch. VII' adjusts appropriate amounts of one factor, the other factor given, to achieve maximum profitability along the route. This clearly is not the minimum cost route, as Dr. Dingwall shows.

envelope curve[21] conceives the process of the firm as a temporal process. The theory is not truly dynamic, however, because time is defined as neutral; that is to say nothing happens in time except a series of adjustments to purely quantitative increases in "lay-out". The assumption usually made is that at each time period (short period) an adjustment is made to a short-period equilibrium by combining appropriate quantities of the variable factors with some given quantity of the fixed factor or factors.[22] In time the process consists of a series of alterations in the use of the fixed factor and a consequent adjustment in the use of variable factors, the whole process, in the long period, moving along the envelope curve towards a "long-run equilibrium".[23] Now, the alteration in the use of the fixed factor is generally conceived of as being continuous ("A long-period average total cost per unit curve may be drawn. In this case the amount of fixed equipment used is taken to vary as X varies, an appropriate amount always being used").[24] The curve is conceived as being a smooth parabola, at first falling, then rising, like the short-run average U curves. "The downward slope in the early ranges is due to the increasing economies of large scale which may be utilized as the output of the source expands. The upward slope in the later stages is due to the increasing difficulties of coordination

---

[21] Cf. J. Viner, "Cost Curves and Supply Curves", *Zeitschnift fur National öconomie*, Band III, Heft 1, Sept., 1931, pp. 23-46; R. F. Harrod, "Doctrines of Imperfect Competition", the *Quarterly Journal of Economics*, Vol. XLVIII, No. 3, May, 1934, pp. 442-71; A. P. Lerner, "Statics and Dynamics in Socialist Economics", *The Economic Journal*, Vol. XLVII, No. 186, June, 1937, pp. 253-70; J. Dingwall, op. cit., Lionel Robbins, "Remarks upon Certain Aspects of the Theory of Costs", *Economic Journal*, Vol. XLIV, March, 1934, pp. 1-18.

[22] This corresponds, therefore, to Dr. Dingwall's third category of theories of process. " . . . any given quantity of a given factor of production will be combined with the 'appropriate' quantities of the other factors . . . The 'appropriate' amounts of the other factors are defined to be those which will maximize profit under the given conditions." Dingwall, op. cit., p. 449. Dr. Dingwall cites Mrs. Robinson as holding this view, but it is really common to all followers of Marshall. Cf. the articles cited above of Professor Viner and Mr. Harrod, and also the contributions to the theory of costs of Professor Pigou and Messrs. Sraffa and Shove.

[23] For an explicit statement of "long-run equilibrium" see R. F. Harrod, op. cit., p. 454.

[24] R. F. Harrod, op. cit., p. 449.

and control".[25] This curve thus reflects external economies of scale as well as internal,[26] and what E. A. G. Robinson has called the rising marginal cost of management. It does not, however, include economies internal to the firm as a result of innovation or qualitative improvement of the factor.

Certain further assumptions have to be made or implied. Some relation between short-run realized and long-run anticipated average (and marginal) revenue must be assumed (i.e., some given value for the elasticity of expectations), which further implies some assumption about the degree of competition in the commodity market. Again, in order to determine the amount of fixed capital, "lay-out", in Mr. Harrod's language, appropriate to any foreseeable output, "it is necessary to assume a specific rate of profit on the fixed equipment. The appropriate rate to assume is the lowest rate which constitutes an inducement to the firm itself to invest in fixed equipment. This I shall call the 'standard rate of profit'."[27] This language would seem to suggest also an implicit assumption of competition in the factor markets.

On these assumptions a theory of the long-period adjustment of the firm may quite simply be erected.

In the following diagram let a_1, a_2, a_3, etc., be the short-period average unit cost curves, shown for appropriate increases in scale of lay-out to produce the series of quantities x_1, x_2, x_3, etc.

The A-A^1 curve is the envelope curve of the family a_1, a_2, a_3, etc.[28] It is the locus of the most economic of alternative

[25] R. F. Harrod, op. cit., p. 449.

[26] For Professor Viner the envelope curve reflects internal economies of scale only. External economies of scale are separately considered. See J. Viner, op. cit., pp. 34-38 and 39-40.

[27] R. F. Harrod, op. cit., pp. 449-50.

[28] The envelope curve by definition defines the lowest average total unit cost for any output. It is to be noticed that it does not transverse the lowest points on the curves to which it is the envelope, with the exception of the bottommost curve, but rather lies in tangency, while it is itself declining, to the declining sectors of the curves of the family, and while it is ascending, to the ascending sectors of the curves. This is not only graphically correct, as Professor Viner's draughtsman averred according to the famous footnote in which Professor Viner refers to this question (see J. Viner, op. cit., p. 36), but is also, as Mr. Harrod shows, correct in the concrete economic sense. "The lowest cost of producing X units is shown

(*Fn. continued on p. 232*)

lowest unit costs for a series of outputs over the long period during which plant is increased. It follows that the marginal curves corresponding to the a_1, a_2, a_3 curves, and the corresponding curve marginal to A-A^1 will coincide for the output values x_1, x_2, x_3, etc. That is to say, the marginal values will be the same for the outputs for which the long and short-run average curves are in tangency.[30]

Now, if there is perfect competition, and if all our cost curves are regarded as including rents, the long-term equilibrium of the firm, by definition, must be at the point of optimum use of resources, that is the point x_3 in Figure 19, where the low point on the long run A-A^1 is in tangency to the low point on the short-period average cost curve, and where both are, by the definition of perfect competition, in tangency to the average and marginal revenue curves.[31] This is the only point that fulfils the no-profit, no-incentive-to-change conditions of the hypothesis of perfect competition.

When, on the other hand, competition is not perfect, we have a less determinate situation. If demand is correctly anticipated, so that the realized marginal revenue curve intersects the marginal cost curves for the output x^1, then there is a kind of short-run equilibrium established at that output. It is not, however, a stable equilibrium, because the degree of excess capacity constitutes a destabilizing force, expecially if there is any degree of competition. In Figure 20, below, let the a^1, a^2 curves be short-period average total unit cost curves,

---

by the lowest point on any member of the family of parabolas for that value of X. The locus of these points is not the line joining the lowest points of the parabolas, but the envelope of the family . . . If the prospective demand is for an output less than such as to employ the optimum lay-out to its optimum capacity, a lay-out less than that of the optimum scale will be employed. Moreover it will not be employed to its optimum capacity. The amount of lay-out indicated as correct by the long-period curve is that, the short-period cost parabola of which is tangential to the long-period curve at that point" (R. F. Harrod, op. cit., pp. 450-51). In a word, it is obviously more economic, from the entrepreneur's point of view, to produce the output a^{x^1} (see Fig. 19) with the lay-out for which a_1 is the short-period average unit cost curve than with the lesser plant for which the average unit cost curve would be at its lowest point for the output x_1.

[30] See J. Robinson, *Economics of Imperfect Competition* (London, 1933), Fig. 7 and Theorem following, p. 33.

[31] See R. F. Harrod, op. cit., proposition 15 (viii) pp. 453-54; Also J. Robinson, op. cit., pp. 93-96.

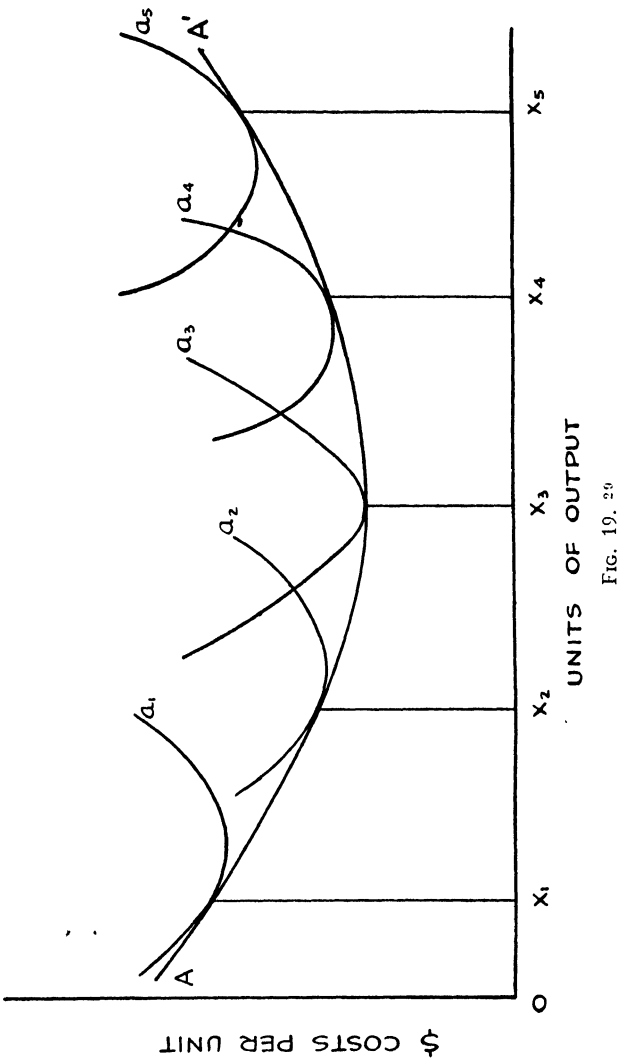


FIG. 19. [29]

[29] The relationship of these curves to the iso-cost curves of Chapter X may be explained as follows. The iso-cost curves represent different combinations of the variable factors yielding the same total cost. Given a certain fixed plant or "lay-out" (in Mr. Harrod's language), we are able to cut the total cost surfaces with a plane and the resulting curve will be a total cost curve for various outputs with the given lay-out. This may readily be reduced arithmetically to an average total unit cost curve and plotted in a two-dimensional output—cost per unit

(*Fn. continued on page 234.*)

as before, with m^1 and m^2 the corresponding marginal curves. Then A-A^1 is the long-run envelope curve and M-M^1 is the curve marginal to it. Let AR^1 be the correctly anticipated (therefore realized) sector of the average revenue curve, so that mr^1 is its corresponding marginal curve. Then the firm has correctly adjusted the appropriate lay-out for the production of x^1 units and is in equilibrium at that point, maximizing its profits by selling at the price P^1. But this solution is of the very sort to which Dr. Dingwall has so properly taken exception. It is the formal definition of a point of possible equilibrium. The beauty of Marshall's solution of the firm's long-run equilibrium was that he showed why, over time, the firm must move towards it and how, therefore, the equilibrium was in fact achieved. If a miscalculation was made, the incentives were such that the firm's behaviour with respect to output was to adjust output in the direction of the equilibrium position. Now, in fact, in the case pictured above, we have no such tendency. If the realized average revenue is less than was anticipated the firm will not adjust by returning along its cost curves and thus *raising* the costs at which it produces. It will, on the contrary, attempt to reduce costs by utilizing more fully and efficiently the capacity it has. If there are several firms in imperfect competition (as we have defined it) their joint behaviour will lead to negative shifts in their respective average revenue curves. The position becomes, thus, increasingly unstable. If, on the other hand, the realized average revenue curve exceeds what was anticipated, the firm will be induced to increase lay-out and expand capacity in order to maximize profits. Let us suppose the realized average revenue curve is AR^2 in Figure 20 (above), with its corresponding marginal curve mr^2, the firm would then increase capacity appropriate for the production of x^2 units of output. At that point it should be in a true equilibrium.[32]

quadrant. For a series of given lay-outs a family of such curves may be derived and the envelope of this family may then be drawn showing the lowest average total unit cost for any output.

[32] It is interesting to note that at this point excess capacity disappears, yet monopoly profits remain. The price (P^2 on the diagram) at which the firm sells

(*Fn. continued on page 236.*)

All this is surely just another way of dealing with the problem raised originally by Marshall and later by Mrs. Robinson of

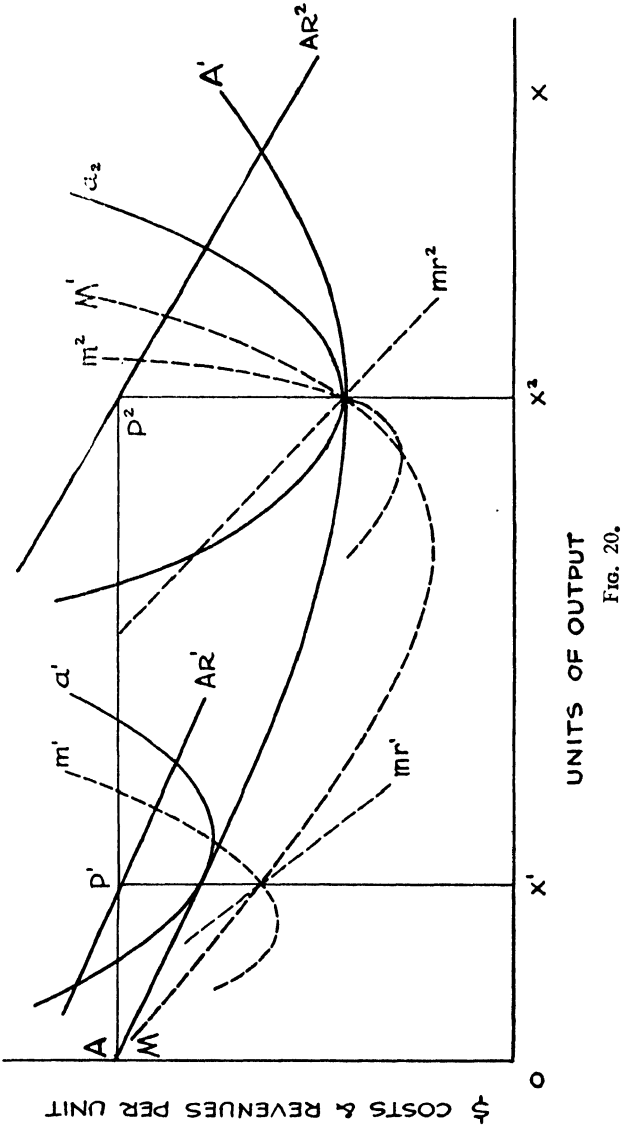


Fig. 20.

multiple points of equilibrium.[33] In their presentation what must be intended as long-run cost curves have more than one point of intersection with similarly long-run revenue curves, considered as changing in elasticity. Some of the points of intersection (where costs are declining) are found to be unstable, others (where costs are increasing) are regarded as stable.

What we are up against here, however the problem may be presented, is the unsuitably formal presentation of the time process. Any rigorous presentation of a "short-period" revenue or cost curve must recognize that two variable quantities are being presented in the form of a function. The average unit cost is a function of the output, average revenue a function of the quantity put on the market. These functions have certain properties or a certain form or shape. Economists have exercised great ingenuity in developing proofs or demonstrations of the true shape of these functions. Statisticians have shown, however, that these functions can never be measured as such, because the economist always must assume that the dependent variable changes only with alterations in the independent variable, "all other things remaining equal". All other things do not, in fact, remain equal, and price variations, for example, occur not only with changes in sales, but with changes in taste, style, supplies of other goods, etc. Statisticians, attempting to measure the demand function for any commodity, must assign by arbitrary means constant values to these other variables in order to achieve any even approximate measure of one function.

When we turn to the problem we are at present considering we see that the smooth "long-run" envelope curve is intended

considerably exceeds the average total unit cost of production, and the output of goods put in the hands of the public is considerably less than that output for which average cost and price are equal, although there is no excess capacity in the industry. This raises in acute form the question as to the socially optimum output. Mr. Harrod (op. cit., pp. 462, 463) suggests that the intersection of the marginal curves does not determine optimum allocation of resources, even if there is no excess capacity. Mr. Lerner and Mr. Dickinson ("Price Formation in a Socialist Economy", *Economic Journal*, Vol. XLIII, June, 1933), have agreed that the socially optimum output is determined by the intersection of the long-run marginal cost and average revenue curves.

[33] See A. Marshall, *Principles of Economics* (2nd ed. London, 1891), pp. 491-92 and J. Robinson, op. cit., pp. 57-59.

to convey the idea that the firm is expanding output over a long period of clock time.[34] During this period of time plant extension is involved. Nevertheless the temporal process is regarded as involving no other variables so that unit costs may be presented as a simple function of output.[35] The smoothness of the curve along which the firm's output-unit cost function travels in time implies an infinite number of "short-period" curves and a continuous adjustment of "appropriate lay-out", further implying the possibility of infinitely small increments of installed capacity. Moreover, demand is regarded as constant in time, the only temporal phenomenon admitted on this side being the possibility of inaccurate forecasting.[36] In fact the average revenue curve of any firm, even a monopoly, will shift in time, and, if the firm is in some sort of non-perfect competition, every alteration in its capacity and every alteration in the output of a competitor will alter the position of the average revenue curve. I do not believe there is any true determinate equilibrium under such conditions, only a relatively short-period stability held by agreement. Thus it is doubtful, under conditions of imperfect competition, if there is any long-period equilibrium point, towards which the firm approaches. The firm's process in time is, if it is in any kind of competition, to achieve some advantage over its rivals, or, if it is a monopoly, to enhance its monopoly power and, a less powerful motive, to increase its total net profits. Under any of these circumstances the process involves as its essential temporal attribute qualitative improvement in techniques.

As always the test of a valid assumption must be largely pragmatic. Does the assumption enable us to create a model which affords useful knowedge of the economic world, and does the hypothesis enable us to explain problems or clear up con-

[34] "It is more correct, therefore, to speak of long-run equilibrium in terms of the conditions which will prevail after a long-run, rather than during a long-run. Long-run equilibrium, once established, will continue only for an instant of time, if some change in the primary conditions should occur immediately after equilibrium in terms of the pre-existing conditions had been reached." J. Viner, op. cit., p. 29.

[35] "The theoretical static long-run, it should be noted, is a sort of 'timeless' long-run, throughout which nothing new happens . . . " Ibid.

[36] Cf. R. F. Harrod, op. cit., pp. 452-53.

fusions as yet unexplained and undissipated? The traditional theory of the market equilibrium of the firm does not, in fact, explain the process of adjustment and it follows that we cannot assert with safety that in the long run the firms adjust to the normal equilibrium price-output point.[37] Indeed, if the process of adjustment is such as to disturb the conditions governing (determining) that point, it ceases to be a static point, and itself travels along some curve in time so that the process of adjustment becomes a continuous one, and competitive market stability disappears.

In brief, we have here an illustration of the limitations of the concept of equilibrium and the method of comparative statics, limitations to which we referred in Part I. The real process of change gets left out and the relation of this process to the structure of the market is hidden. It is precisely to this that we wish to direct our attention.

---

[37] Cf. Dingwall, op. cit., pp. 454-56. "It is to be noted, however," he says (p. 455), "that when all illegitimate process propositions have been stripped away, the equilibrium 'analysis' is reduced to a purely formal level since it gives no explanation of the equilibrium. It is improper to picture the firm as moving through the successive positions given by the curves and as choosing by experiment between these positions as alternatives."

CHAPTER XI

THE TEMPORAL PROCESS OF THE FIRM

1. PROCESS UNDER THE USUAL ASSUMPTION OF PROFIT MAXIMIZATION

The notion of "process" that we believe is appropriate to the study of the temporal adjustments of the firm must be truly temporal. That is to say, it must be based on explicit assumptions about the concrete nature of time. The idea of "process" that we shall henceforth use, then, will include not simply the adjustments of employment and output to achieve an equilibrium over a period of stationary time, but also the qualitative changes in the factors, the appropriate adjustments in the firm's employment of them and the consequent alterations in the structure of the market. In this first section we shall retain the ordinary assumption that the firm aims at any moment of time to maximize net profits at that moment. This is a necessary stage in our studies, relating them to traditional firm analysis. We shall later modify this assumption to one more proper, as we believe, to the kind of temporal study we are undertaking.

Now there are various alternative assumptions we can make about the nature of time. Of these, the most fruitful and the most justified by empirical evidence, we believe to be that time, for the firm, consists in discontinuous improvement in the quality of the fixed factors with a consequent substitution of fixed for variable factors in production. Specifically longer and more efficient power units and machinery, requiring larger expenditure on fixed capital, tend to be introduced. It is on this phenomenon that we shall concentrate. It is not universal—there are many industries in which it can be shown not to be characteristic. Nor is it the only kind of improvement. Labour skills may be improved. In some instances, no doubt, materials deteriorate.

These we do not believe to be sufficiently characteristic of the modern industrial world to occupy us. A more important trend is the introduction of innovations that save the time and extend the practicable scope of management, such as improvements in accounting techniques, specialization of entrepreneurial functions, better trained secretaries, improvements in the use of dictaphones, teletype machines, comptometers and so forth. Since these innovations work in the same direction as those we select for discussion, we do not discuss them separately. They too, operate to increase the intensity of use of a fixed factor (management) to improve its quality and consequently to increase the mean size of industrial undertakings.

The drive for innovations in industry we do not believe to be peculiar to monopolistic industry. Under perfect competition, the competition is in terms of price. Each firm has equal average costs (including rents) at the point of output. Since no entrepreneur can assume that his competitors will behave as though the position is a static one, each strives continually to lower his average unit cost curve. This permits him to earn more than normal profits with an increased output, or he hopes that it will. Since any improvements are rapidly adopted throughout the industry (as long as it remains perfectly competitive) it is unlikely that any advantages accruing to any one firm will be long-lived, but at least the firm retains its position. How often have we not heard business men recite the credo "You can't stand still. Either improve your business or lose it".

When competition is non-perfect, it is unlikely to be price competitive, or, if it is, to continue long in that form. Prices are usually settled by agreement among the large firms; if small firms in some number remain, as under what we have called "unequal" competition,[1] they accept the price leadership of the big firms and adjust their output to the given price as firms do under perfect competition. The competitive element under these conditions takes the form of trying to increase the firm's share of the market. This is achieved partly

[1] The classification of market positions we have used throughout this work is given in Ch. VII, fn. 10.

through "commodity differentiation"[2] by means of advertising and partly from innovations intended to increase output capacity, selling attractiveness, etc. Indeed much effort goes into innovations in advertising appeal, merchandising, packaging, and "selling-point" gadgets.

The monopolist is also interested in improving his technique over time. Economists are accustomed to describe the monopolist's price as being determined for that output which equates his marginal cost and marginal revenue. This is no doubt the correct solution on the assumption that the monopolist fears no threat whatever to his position, because it maximizes his profit. However, the monopolist, far from the all-powerful and unchallenged tyrant he is popularly supposed to be, is in fact—as who is not these days?—a creature of hidden and awful fears. He fears public opinion, government "interference", he shudders at committees of investigation; even the most moderate utterances of the Commissioner of Combines send him trembling to cover.[3] Worst of all, he fears that someone else may volunteer to relieve him of all these dangers by undertaking to compete with him. Consequently he keeps his price as near the maximum point as he deems consistent with security. This means it must be sufficiently low to discourage any possible competitor from contemplating starting operations. Closure of entry is an important objective of monopoly price policy. This would be inadequate, however,

[2] "A general class of product is differentiated if any significant basis exists for distinguishing the goods (or services) of one seller from those of another. Such a basis may be real or fancied, so long as it is of any importance whatever to buyers. . . . Where such differentiation exists, even though it be slight, buyers will be paired with sellers, not by chance and at random (as under pure competition), but according to their preferences". Edward Chamberlin, *The Theory of Monopolistic Competition* (Cambridge, Mass., 1935), p. 56.

[3] I remember one day being called to the telephone by an acquaintance who is the secretary and moving spirit of a certain "pool" which undertakes to control the price of an important commodity. His voice was anxious. "Is it true," he asked, "that Mr. F. A. Macgregor (the Canadian Commissioner under the Combines Act) is in Montreal?" I told him he had been correctly informed. "Can you tell me," he went on, "without breach of confidence, what Mr. Macgregor is doing?" "Certainly," I said, "he's making a speech at McGill University." "You're sure that's all?" "Yes, quite sure." "Ah!" said my friend in tones of relief, "I shall be glad, though, when he goes back to Ottawa." As Mr. Macgregor himself has been heard to say, he cannot buy a railway ticket to any town in Canada without raising someone's blood pressure.

if the monopolist did not see to it that he would always have an advantage over any new rival. In Section 2 of this Chapter we shall examine the pricing policy he may pursue over a long period, and the conditions under which the pressure towards innovations may be removed.

In the meantime, let us agree that an essential characteristic of the firm's process in time is qualitative improvement of the factors it employs. This process, from its very nature, cannot be regarded as continuous, in the strict sense of the term. It is discontinuous. It leads us, once accepted, to a further point of some significance. The theories of process we have examined have all assumed that the process was towards larger firm units, i.e., that the firm expanded in size over time. Now, on the assumption that there were no qualitative improvements in the factors, there is no *a priori* reason why this should be so. Indeed, if the fixed plant remained qualitatively the same, and the appropriate, i.e., the optimum, combination of variable factors were always combined with it, there is every *a priori* reason to believe in constant returns to scale, so that expansion would be initiated from any short-run equilibrium point only by an increase in demand, or an anticipated increase in demand. There is some *a posteriori* evidence for expecting external economies of scale (as well as external diseconomies of scale),[4] but that is all. Now we know on empirical grounds that the facts of modern industry are that the number of firms tends steadily to decline while the modal firm increases in size. This phenomenon is not simply to be explained away by reference to a growth in demand. It has happened in industries in which there has been no significant growth in demand, and, even where there has been expansion of sales this is often not adequate, of itself, to explain the concentration of the industry.[5] The reason

[4] See J. Viner, op. cit., pp, 40-42.

[5] I am anxious to preserve the line of argument in the text at this point free from masses of statistics and qualifications on points of detail. There are many industries in which there has been no marked change in scale during the past fifty years, others in which increases of scale have not apparently brought increased efficiency. See Sargant Florence, "Economic Research and Industrial Policy", *Economic Journal*, Vol. XLVII, Dec. 1937, No. 188; and "The Relative Efficiency of Large, Small and Medium-Sized Business", Temporary National Economic Committee, *Investigation of Concentration of Economic Power* (Washington,

for this is to be found in the nature of the typical technology of our period. The process is to introduce new, more efficient machine units, involving the substitution of machinery for labour, and thus increasing the scale of undertaking while altering in favour of fixed capital the combination of the factors. Since the process necessitates an increase in total fixed costs, lower unit costs can only be achieved by a considerable expansion of output (it is from this that the contemporary entrepreneur's preoccupation with high utilization of installed capacity arises). Therefore the entrepreneur must anticipate either a positive shift in demand or a high elasticity at low price points.

The model in which we carry forward the present analysis requires a somewhat different graphical method, which we believe more suited to the study of temporal processes than the envelope curve, the implicit temporal assumptions of which we believe to be inappropriate. We plot an instantaneous or static or short-run position in the positive quadrant of the ordinary output-dollar coordinates, but we introduce a third dimension, time, which is defined as consisting of qualitative improvements in the fixed factor or factors of such sort as to permit increases of output with increases in scale at increasing returns to scale. (We shall also, shortly, introduce demand phenomena to our definition of time.) A unit of time, therefore, consists of the introduction of new, qualitatively different, machinery (or other fixed plant) involving, ordinarily, an increase in size. Over a period of time the process, consequently, may be described graphically as in Fig. 21. It is necessary to emphasize that these graphs are illustrative only, and that our models, like those of Part III, are inferential and probable and not rigorously determinate.

In Fig. 21 the Ot axis represents the passage of time, an independent variable measured in the unit of an innovation requiring enlargement of plant. The cost position of the firm at the start of the process is shown by the VC_0, FC_0 and AC_0

1941). The general trend, referred to in the text, is nevertheless so apparent in representative modern industries, such as base metals, electrical engineering, automobiles, construction, heavy chemicals, textiles, pulp and paper and so forth that it requires little statistical support. In Part V of this work some of the evidence of the process in Canadian industry is offered. We need not repeat it here.

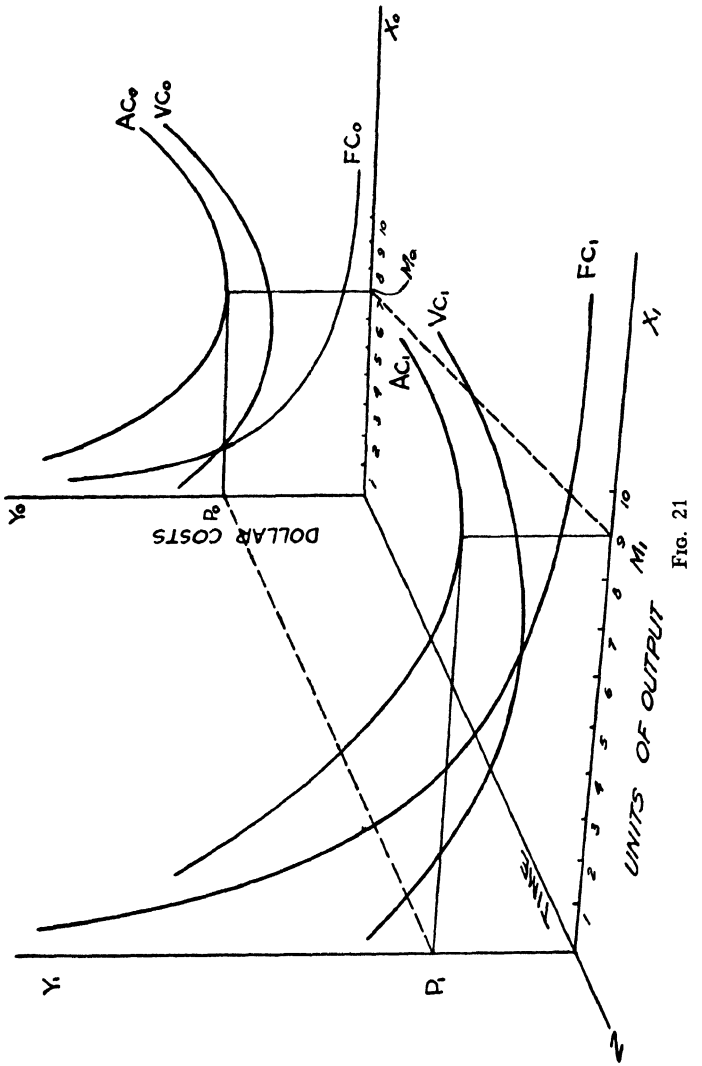


Fig. 21

curves, representing the static variable, fixed and total average unit costs at that moment. The position of the firm at the conclusion of the process is represented by the VC_1, FC_1 and AC_1 curves. Variable costs per unit are lower throughout, by reason of the substitution of fixed for variable factors and the

increased efficiency of the variable factors employed. Total fixed costs, on the other hand, are higher, so that the rectangular hyperbola representing fixed costs per unit lies higher in the plane. Thus the total average unit cost curve is higher for low outputs, but falls over a greater sector, falling more rapidly because of the greater influence of fixed costs, until for outputs of more than "four" units it is lower than the AC_0 curve. This is both analytically correct and empirically justified, and seems to me a preferable representation to any which implies that total average unit costs are lowered, by the process, for all outputs (see Fig. 19).

We have indicated in Fig. 21 the low ("optimum") cost points on the AC curves, illustrating the shift to greater output at lower unit cost that follows this process. This is the true nature of decreasing costs in time. The law of diminishing returns is a true law, but a static one. At any moment of time, AC curves must be U-shaped and, for most markets, equilibrium will be achieved under conditions of increasing costs. Costs do, however, decrease, over time, with increases in output, and this phenomenon must be given emphasis as a cause of unstable competition and the corruption of the perfect market.

This process, we have said, leads to instability. Indeed, one is forced to question both the appropriateness and the usefulness of the traditional equilibrium concept as an analytical tool in this context. This questioning must become ever more searching when we introduce considerations of demand.

There are two possibilities of temporal changes in realized average revenue. The first arises from the fact that the demand curve for any product may itself shift over time. The second comes from the fact that the firm, at any moment, knows only a small sector of its average revenue curve, and may incorrectly forecast the slope and shape of the curve for greater outputs.[6]

[6] In this sentence some sort of imperfect competition has been tacitly assumed. In perfect competition the firm knows the shape and slope of its average revenue curve, but it may not correctly forecast where it will lie, because if all firms undertake a general expansion they may incorrectly forecast the decline in the demand curve. Indeed it is exceedingly improbable, that in this case, they are capable of forecasting at all, because they will ordinarily presume that price will remain constant.

Let us first suppose that there is no temporal shift in demand and that the firm correctly forecasts its average revenue for future outputs. We shall take the case of a firm in imperfect competition, both because, as we shall show, this whole process is bound up with imperfection in the market, and because the

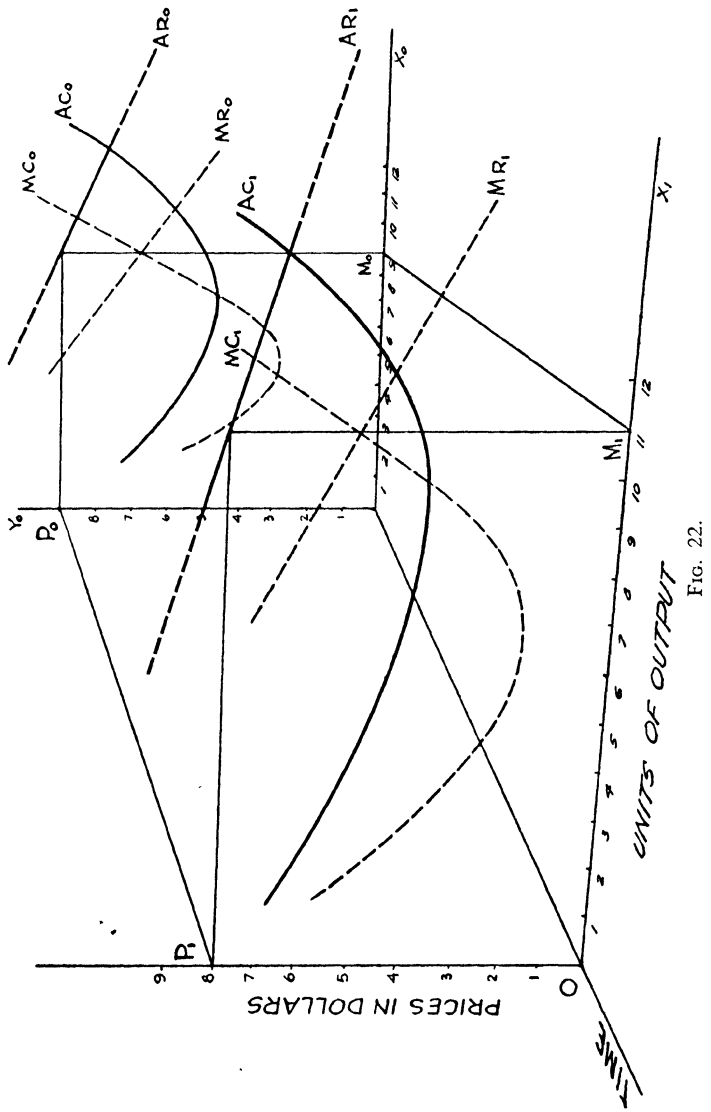


FIG. 22.

problem of forecasting by the individual firm in the perfect market is so much more difficult. In Fig. 22, we regard the firm as having been in equilibrium at time 0. It then increases scale, with the effect of raising its average unit costs for low outputs, but increasing the steepness of the negative slope of the curve, so that for greater outputs unit costs are reduced as compared with what they had previously been. The known sector of the average revenue curve in time 0 is shown as the heavily marked segment of the curve, which is believed by the firm (and correctly, according to our present assumptions) to have the shape for greater sales as indicated by the dotted extrapolation. The firm will then move, on the ordinary assumptions of economic theory with respect to market behaviour, to the equilibrium point at time 1, for an output of OM_1, selling at a price OP_1, thus maximizing its profit under conditions as given for t_1, and increasing profit over the maximum possible at t_0.

It is, of course, clear that in this case all we have done is to set out by means of a somewhat complicated graphical apparatus two static positions of equilibrium and to define the conditions under which the firm will move from one to the other. If it were not for the fact that more complicated problems lie before us, for the analysis of which a three-dimensional graphical system is helpful, it would be quite needless, at the moment, to use more than two simple, two-dimensional graphs set side by side, defining the conditions of the two equilibria. We have had nothing to say as yet about how the firm moves from one position to the next.

Let us turn to this question now. If, for a moment, we assume the process to be a continuous one we may then represent the constant average revenue correctly forecast by the firm, as a curvelinear surface lying within the OX-OY-OT area with a constant slope with respect to the OT-OX plane. Any point on the surface represents a potential point for a potential output at a certain moment of time. On the assumption of continuity[7]

[7] Continuity is quite illicit as an assumption within this model, and I am not attempting its surreptitious reintroduction. For illustrative purposes I am using it temporarily.

the cost surface is also a curvelinear surface, convex to the origin, falling and then rising, but the whole tilted downwards as the cost curves become more steeply inclined and as the optimum cost point lies lower as one moves forward in time. The two surfaces representing marginal revenue and marginal cost must intersect in a continuous curve. From this curve of intersection perpendiculars erected to the average revenue surface will give the price curve which can be carried across to the *OT-OY* plane describing the equilibrium price in time, while the extension of the perpendiculars to the *OT-OX* plane will result in the curve of equilibrium output in time. This is done in Fig. 23. A portion of the *MR* (diagonally shaded) surface is shown lying above the curved tilted surface of the *MC* curves, which, rising, intersects it in the curve E_o-E_t. Perpendiculars from this curve dropped to the *OT-OX* plane yield the curve of output

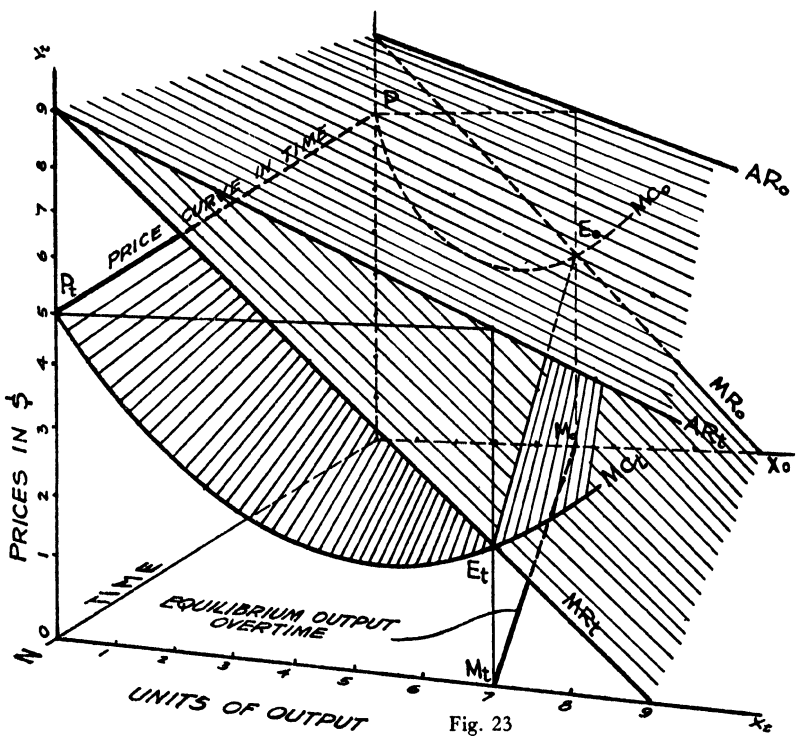


Fig. 23

over time, M_o-M_t, and extended upwards to cut the average revenue surface, the price curve P-P_t which is then projected on the OY-OT plane.

The output curve represents the expansion of the firms under the conditions assumed. Perpendiculars projected from it, intersecting the variable and fixed unit cost curves would show the appropriate lay-outs and appropriate variable factor costs for profit maximization at every point along the path of expansion.

Since, however, we have shown that we cannot believe the process of expansion to be a continuous one, we have to ask what happens when discontinuities and indivisibilities are present. In this case, still supposing demand to be constant in time and accurately forecast, the firm moves at any moment to maximize profits for the given lay-out. It will increase lay-out if it forecasts greater profits and/or greater competitive advantage with the larger fixed plant, and having built the new plant then moves to the new output equilibrium point. No determinate intra-lay-out path is possible. What the firm may do in the interim construction period it is impossible to say. It may operate with its original plant, continuing its previous output and price policy. It may be able to bring parts of the new plant into operation gradually and so expand in a more or less continuous manner towards its new equilibrium point. If the new plant is indivisible in operation, it is more likely to increase its merchandising and advertising costs and expand sales, even at a temporary loss, looking towards its soon-to-be increased capacity, and perhaps writing off temporary losses as part of the fixed cost of bringing the new plant into operation.

This enables us to deal, in *passim*, with the vexed question of reversibility to which we have earlier referred.[8] The problem here, it will be remembered, derives originally from Marshall. In treating the question of how equilibrium in the market is in fact attained, Marshall infers that in the long run the firms adjust their output so as ultimately to achieve the price that will cover costs, the firms "move back" along their cost curves, i.e., reduce their output until, by trial and error, the appropriate

[8] See page 209.

amount is reached. Now Professor Slichter and Professor Boulding have shown that the firms cannot "move back" along the same curve of costs as that which marked their expansion. This is because certain fixed costs are incurred in "the advance" which cannot be reduced in "the retreat", and because, temporally speaking, the rate of increase of cost in the advance and the rate of diminution in the retreat, will, for empirical reasons, be different. There is, in consequence, a certain sector of indeterminacy in the industrial supply curve. This difficulty arises, in our opinion, from the confusion inherent in looking upon a curve representing the functional relation of costs to outputs as of a temporal nature. Properly conceived such a curve represents a schedule of outputs and the unit costs that would be incurred for each possible output in the series of outputs, all regarded as potential in a moment of time. The same, as we have seen (Appendix to Chapter IX), holds true of the demand curve, so that equilibrium is a purely analytical concept, a potential point of rest at any moment of time. A cost curve of a firm, or a supply curve for an industry, so considered, is, of course, reversible in the sense that at any moment of time the costs attached to the series of outputs will not be altered whether we consider the firm moving forward or backward along the curve. Such movement, however, is a highly ideal concept and I am not quite sure what, if anything, it can mean in concrete terms. Actual movement of the firm, in terms of adjustments of output, is not along the static curve at all, but from some point on a static curve at one moment of time to a point on a different static curve at a different moment of time. This movement traces a cost curve and an output curve in time. These curves are in no sense reversible. The firm (or firms) will, if forecasting is perfect and there are no serious indivisibilities, achieve and move along an equilibrium path in time, but if demand shifts or if perfect forecasting is not achieved, firms may have to move first so as to increase, later so as to reduce, output. That is the situation anticipated by Marshall. Now the reduction in output, in this case, will result in the attainment of a cost point on a new static curve in a new moment of time. It will clearly not coincide, save by

rare accident, with the same cost-output point at some previous moment. Time is not a reversible variable, and constituted as it is of a certain kind of change, that which happens within it is not reversible. The reduction of output and the approach to an equilibrium price must be shown graphically as a process along the *Ot*-axis. The further analysis of this problem requires that we now drop the assumption of unchanging demand, accurately forecast.

It does not matter for our present purposes whether we regard demand as changing, with the changes themselves improperly forecast, or if we simply regard the forecasting of demand, changing or unchanging, as inaccurate. At any moment the entrepreneur knows only a small sector of his marginal revenue curve. If he is in perfect competition he will know the slope of his curve for any future moment, but he will not be able to forecast accurately how high the curve will lie above the *OX*-axis. If he is in some form of imperfect competition he will not know, though he may guess, the elasticity of his average revenue curve (and hence the shape and position of his marginal revenue curve) for outputs in excess of those currently being sold on the market. Indeed, if he is in monopolistic competition, and has a kinked average revenue curve, his marginal curve may be such as to leave him both without a determinate price point at any moment and without any notion as to the probable shape and position of his marginal revenue curve for different outputs at future moments.[9] Let us, however, put such difficult cases aside. The single firm now has a marginal cost surface in time, the shape of which it may be conceded to know both for the past and foreseeable future. There will, however, be two average revenue and two marginal revenue surfaces, one set of revenue surfaces representing the series of forecasts as to future revenues made in successive moments, the other set representing realized revenues as future moments successively flow into the present.

---

[9] See J. Robinson, *Economics of Imperfect Competition* (London, 1933), pp. 37-38, 81. Also M. Timlin, "General Equilibrium Analysis and Public Policy", *Canadian Journal of Economics and Political Science*, Vol. XII, No. 4, p. 484.

Thus in Fig. 24 the MC_0, MC_1, and MC_2 curves represent the firm's marginal cost curves for three moments of time during which the characteristic expansion of lay-out has been carried out. The AR_0, AR_1, and AR_2 curves represent the average

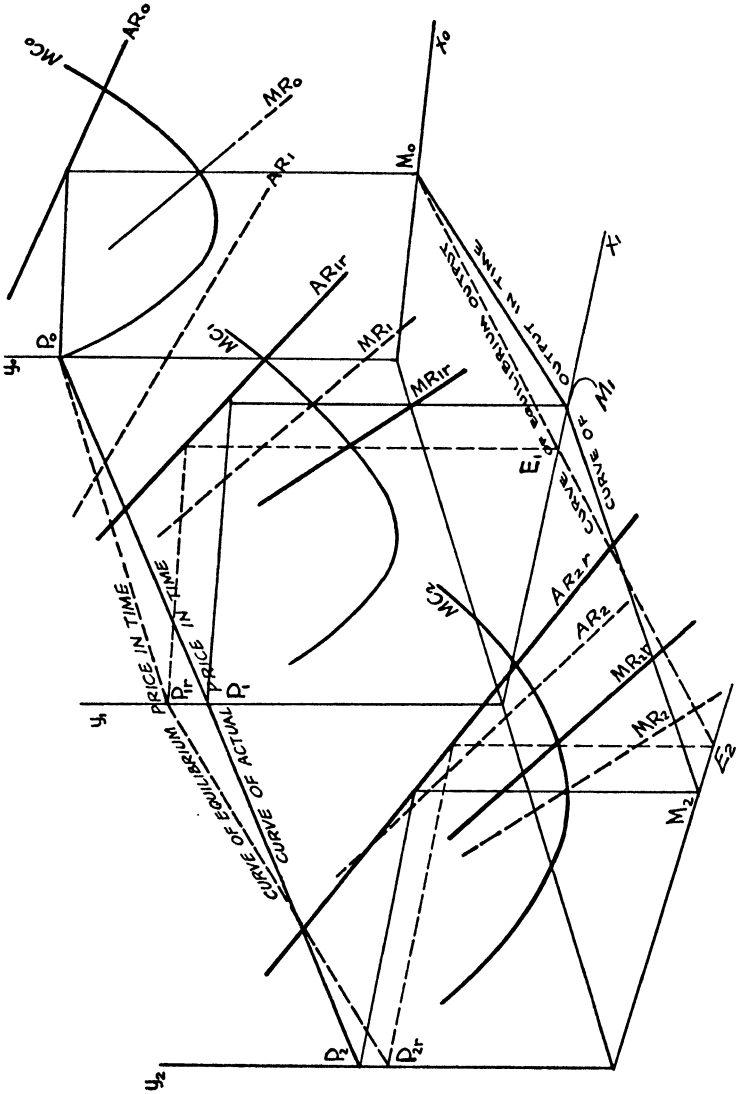


FIG. 24.

revenue as realized in time 0, and as anticipated, the AR_1 curve being anticipated at time 0 and the AR_2 curve at time 1. Anticipations are here represented graphically as simple extrapolations of realized revenues. The MR_0, MR_1 and MR_2 curves are the corresponding marginal curves. The AR_1r and AR_2r curves are the actually realized average revenues at times 1 and 2, and the MR_1r and MR_2r are their corresponding marginal curves. Now had the firm forecast correctly so that perfect adjustments to the conceptual equilibrium had been achieved, the outputs would have been determined by the intersections of the MC and MR_r curves yielding an equilibrium output curve as shown (M_0-E_2) with a corresponding curve of equilibrium price (P_0-P_2). In fact, however, because of inaccurate forecasting, the firm adjusts its output in t_1 for the amount OM_1, the amount, that is, which would equate its marginal cost with its anticipated marginal revenue. This quantity it is forced to sell at a price of OP_1, considerably less than it had anticipated. The firm must now readjust its output. If there were no further possible improvements over time in costs, it would simply reduce its output to the point of equilibrium, but since, as we assume, the time process continues, it adjusts so as to equate its new marginal cost (MC_2) with the presently realized marginal revenue (MR_1r) which it now anticipates and which is projected from time 2 as MR_2. It therefore produces the amount OM_2, which it is able to sell for OP_2, a price in excess of the anticipated price, because once again it has incorrectly forecast demand. Presumably in the next period it will once again over-produce for the price it hopes to realize unless there is an actual positive shift in demand. Thus the firm's actual output and price curves in time tend to weave a narrowing criss-cross pattern with the true equilibrium curves, a conclusion, actually, which agrees very closely with the result expected by Marshall.[10] This is also, I believe, the true solution to the problem of multiple points of equilibrium, a problem

[10] These price and output functions in time are, or ought to be, capable of measurement, and can be expressed, of course, as functions of one another. I have hoped that statistical inquiry, based on time series, might be able to create mensurable price and cost functions that would correspond to the analytical functions derived here.

that can only arise when one neglects the question of forecasting demand over a time period.

2. Profit Maximization in Time and Senility of Firms

We have so far conducted this argument on the usual assumption that the firm attempts to maximize profits, but this assumption has not been defined in a manner fully consistent with our concept of time. Though the profit maximization assumption is common to all theories of price and of the firm, it is seldom precisely defined. What is implicit in the marginal cost-marginal revenue theory of a firm's equilibrium is that the firm operates so as to maximize profit at any moment of abstract time. It is a highly abstract idea and not intended as a description of what firms really do. When we define time concretely in terms of the processes of which it is constituted, we are obliged to ask whether the firm aims at a maximum temporal rate of profit (which is exactly what the orthodox theory says), or whether it aims to obtain the largest estimated profit over some (indefinite) period of time.

I am convinced that the latter notion is both more appropriate to, and consistent with, our level of discussion and at the same time more in accord with the real facts of actual situations. The difference between the two assumptions in the case of perfect competition is not important in the sense that it does not materially alter the conclusions of analysis. In the case, however, where monopoly or near monopoly exists, our whole view of the behaviour of the firm is affected. In this case the maximization of the rate of profit in time does not necessarily coincide with the maximum amount of profit over a period of time. This is because the maximization of profit at any moment may result in the appearance of competitors whose supplies would reduce price to the point where total net profits over a sufficient period would be reduced below what they might have been had a lower rate been accepted and the potential competitors kept out of the market.

My attention was drawn to this point by a description given me by a friend of his firm's pricing policies. His firm, a giant,

multi-product corporation, is a monopolist or near monopolist in many of its products.

"I do not believe," my friend said, "we ever sell or attempt to sell at the point where current marginal cost and marginal revenue are equal. In products in which we have competition, actual, known competition, we may be able to approach this point, though not to achieve it. Still it serves in such cases reasonably well as an ideal or analytical explanation of our behaviour. In commodities, however, in which we enjoy a monopolistic or quasi-monopolistic position, the equality of the rates of cost and revenue with output will not serve as an explanation of our behaviour.

"Let me tell you what we do. I'll use baubles[11] as an example. We manufacture over 90% of the baubles at present marketed in Canada. Perhaps 5% are imported out on the west coast, and a tiny competitor, whom we neglect in practice and who can be neglected in theory, supplies 1%-2%. You may, for all practical purposes, say we have a monopoly of the supply of baubles in Canada. Indeed in the great central market of Ontario and Quebec we have a true and perfect monopoly.

"Now we currently market, say, 24,000,000 units of baubles in Canada at a price of 12¢. Our average cost is around 9.3¢. This is very far from the maximum profit point as defined in economics. I cannot, of course, say exactly where that point would be, but I suspect it would be in the neighbourhood of 18,000,000-20,000,000 units, when price could be 14-15 cents. How and why, you may then ask, do we sell at twelve cents?

"The reason is perfectly clear to any industrialist. The John Doe Manufacturing Co. Ltd., which does not make baubles, quite easily could. In modern manufacturing, with the techniques at our disposal, no firm is ever free of such *potential* competition. Now we know, almost exactly, what it would cost John Doe to make baubles. We know the ap-

[11] "Baubles", of course, is an imaginary commodity. Under the laws of both the United States and Canada the policy to be described in the text might be regarded as illegal, and to name any real commodity might prove to be exceedingly embarrassing to the publishers, the author and even to the industrialist whose evidence I am presenting.

proximate size of plant he might profitably erect, so we know his fixed costs over an appropriate range of output. We can estimate to within a fraction of a cent his operating costs. We know that for the first three years of production he could not get his costs below 11.5¢ a unit. Thereafter he might get down to 10.5 cents, and our own costs on seriously reduced output would rise to something close to that.

"Let us suppose we sold at 15 cents, with costs around 9.2 cents. Our total profit would be, at sales of 18,000,000, $1,044,000 each year. This compares with our present actual profit per annum of $668,000. If, however, we sold at 15 cents, John Doe and Co. would find it definitely attractive to enter the market. They would sell at about 13 cents to begin with and would operate at around 6,000,000 units. This would reduce our sales to about 14,000,000, raise our unit costs, because of the distribution of a heavy overhead, to 10.5 cents, and reduce profits to $350,000. If after three or four years price was settled at about 12 cents the total sales would be 24,000,000 units as now, but our share would be only about 16,000,000 units, our costs close to 10.5 cents, so that our total annual profits would fall to $240,000.

"Consider consequently, our profit situation over a period of, say, 12 years, based on the average life of our patents. At present we earn $668,000 a year. For twelve years that yields a total profit of $8,016,000. If, however, we sold at 15 cents, our profits for perhaps two years would be $1,044,000 per annum. Thereafter they would fall for, say three years, to $350,000 per annum, and then, for the balance of the time, to $240,000. Over the twelve years we could expect to take a net of not more than $5,000,000.

"The choice of a twelve-year period is, of course, arbitrary. Business men cannot afford not to look ahead, yet they cannot look too far ahead, as business conditions and techniques change so rapidly.[12] It used to be conventional to select a twenty-year period as the standard time for such calculations, based,

[12] We shall call the period of time for which the firm plans, its "horizon". Cf. O. Lange, "A Note on Innovations", *Review of Economic Statistics*, Vol. XXV, 1943; and J. Tinbergen, "The Notions of Horizon and Expectancy in Dynamic Economics, *Econometrica*, Vol. I, 1933.

I suppose, on twenty-years' capitalization at five per cent per annum. There was some vague idea too, that that period represented some kind of average life expectancy of a capital good. The rate of invention has changed all that, if it ever had any real basis. I think our selection of average length of legal patent protection as reasonable as any such arbitrary selection of a time period for such estimates into the future. Some firms, I believe, operate on a ten-year projection, on the assumption that this covers the duration of the trade cycle and yields an estimate true for the period, if not for any one year. Whatever the basis of the selection, it is really arbitrary, but you will find in practice that the period chosen always turns out to be 10-15 years. This seems to me to prove that the conventional period for business thinking has been shortened from twenty to about twelve years, and that the various criteria for selection of a period are so many rationalizations."

This is, I think, a true description of how the profit maximization principle works in monopolistic situations. The situation is clearly determinate, once the time period is selected and the estimates of potential competitors' costs are made. These are arbitrary quantities, but, in economics, so are many of the quantities—anticipated marginal productivity, for example,—which are regarded as determining entrepreneurial behaviour. I have tried to present the situation graphically, but there seems no great advantage of this,[13] because the graph is merely illustrative in purpose. The arithmetic illustration offered by my informant seems to me to be perfectly clear.

Now we have to ask what happens, under the assumption of profit maximization over time, when new techniques become available. In other words, how does the characteristic technical process of modern industry affect the firm, when monopolistic conditions obtain?

First, will the innovation be offered to a potential competitor, to the monopolist, or will it be equally accessible to both? Certain innovations, such as those extending the scope of man-

[13] If profit is treated as a primitive function of time and if the period of entrepreneurial planning is regarded as definite, the two possible lines of profit could be graphed as in Fig. 25.

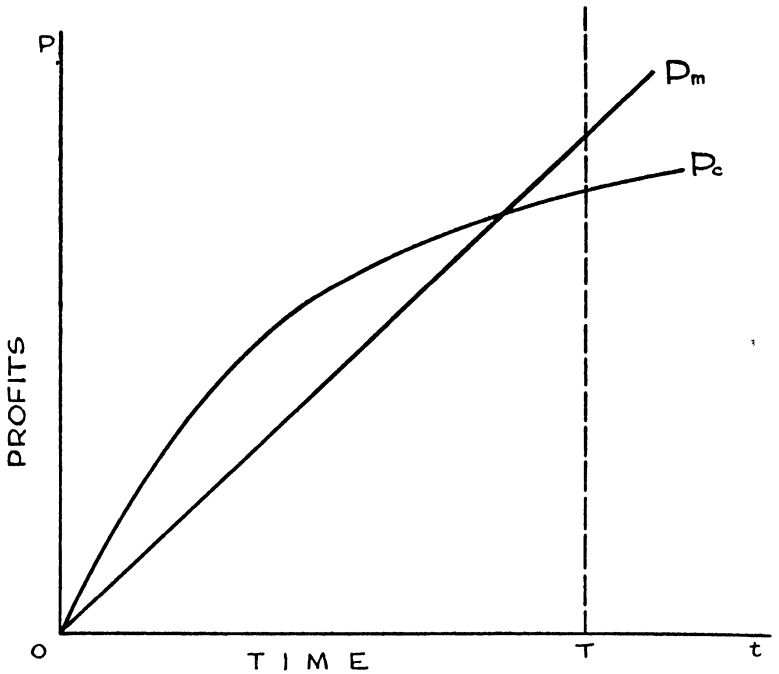


FIG. 25

In Fig. 25, the curve *Pc* represents the cumulative total of profits over time if the firm maximized the rate of profit, moment by moment, without regard for possible competition. The curve *Pm* is the projected profit curve on the expectation that the lower rate would prevent the entrance of any potential competitor. The perpendicular erected from *T* on the time axis marks the limit or horizon of entrepreneurial planning. If *Pm* crosses *Pc* to the left of the horizon curve, then *Pm* will be the profit path the firm will seek to follow. If the intersection is to the right of the horizon curve, the firm may follow the profit-maximizing path *Pc*, as static theory would imply.

agement or the efficiency of power units are clearly generally accessible. These affect both the costs of the monopolist and of the potential competitor more or less equally. The monopolist must therefore adopt the innovation and pass the economies along to the consumer if he is to continue to protect his privileged position. Other innovations, however, may be specific and protected by patent. Such an innovation, if offered to a potential competitor, may turn him into an actual competitor, and the whole situation may become fluid until a price agreement is achieved. This sometimes happens, but it is unusual

and unlikely. Far more probable is the offering of the specific innovation to the firm already engaged in this line of production, or the development of the innovation by the firm's own research. The firm then may, or may not, adopt the innovation, and, if it does adopt it, it may, or may not, pass on the economies to the consuming public.

We must remember that we are discussing only innovations specific to the line of production in which the monopoly is enjoyed. Such innovations are ordinarily patentable. Let us suppose the innovation is adopted. What circumstances may we infer would lead to its adoption?

First, it may be cost-reducing, which would mean that the firm's margin of superiority over any potential competitor would be increased, and its security as a monopolist strengthened.

Second, the innovation might improve the quality or marketability—not the same thing—of the product. This would have the same effect of enhancing the security of the monopolist's position. This would be an especially effective motive if there were some near substitute affecting the elasticity of the average revenue curve.

Third, the innovation, though specific to this line of production, might be of sufficiently general a nature with respect to its major principle that it could not be adequately protected by patent, and might therefore become accessible to a potential rival, even though patented and unused. Or cross-licensing arrangements and compulsory licensing legislation might affect the ability of the firm to protect itself against the use of the innovation by a potential competitor.

The firm can retain the benefits of cost reduction if its margin of security (its present cost advantage over the nearest potential rival) is great. If that margin is small and its monopoly position is insecure it will be less able to hold the price while reducing costs. This also depends, in part, on the degree of patent protection possible for the innovation. If it is not patentable, or, if patentable, must be used or licensed for use, its adoption is a necessity unless the monopoly is itself legally established. If, however, the innovation may be patented and held, without immediate use, it may be more desirable

to withhold it. In brief we may say that the more secure the monopoly, the more able the firm is to retain its gains from innovation, but, the less secure its monopoly power the more anxious it will be to increase that security by adopting the innovation. Thus the ability to retain the gains of an innovation does not seriously influence the firm's decision whether or not to adopt it.

Let us now suppose, however, that the firm does not adopt the innovation. Under what circumstances can we infer such a decision consistent with ordinary economic motives?

First, the innovation must be one that can be protected by patent, so that the firm need have no immediate fear of a potential competitor successfully exploiting it. If the patent law is incapable of yielding such protection in itself, because of compulsory licensing in cases of non-use, this protection may be found in the patent clauses of cartel agreements.

Second, the firm may be uninterested in cost reductions (although a secure monopolist under no compulsion to pass the gains on to the consuming public) because of taxation, which, like the Canadian wartime Excess Profits Tax, penalizes profits in excess of those of a base period. Trades union policy, too, may be such as to remove any incentive to decrease capital cost because labour costs would increase in proportion. This is alleged to be the case, for example, in the Cape Breton coal mines.

Third, the innovation, while reducing operating costs might have a serious effect on the rate of obsolescence of fixed capital. It is perfectly possible that it might render obsolete plant of a costly nature long before depreciation charges had covered the original investment. If there was any tightness in liquid resources this would become a serious consideration.

We may conclude, therefore, that rigidity towards innovations is a condition of monopoly or near monopoly.[14] It is unlikely

[14] "Under oligopoly an innovation cannot be output-increasing unless the diminution of marginal cost caused by it is sufficiently great to induce the firm to break the 'discipline' of the group. The last mentioned case happens when the marginal cost curve shifts to such an extent as to make it move out of the range of discontinuity of the marginal revenue curve. Thus, only innovations which reduce marginal cost to a great extent can be output-increasing under conditions of oligopoly." (O. Lange, op. cit., p. 24).

to be found when there is active competition or when the margin of security over a potential competition is not great. It does, however, exist, under certain monopolistic conditions. These conditions are that the monopoly is reasonably secure and the innovation unlikely to assist a potential competitor; that the benefits of the innovation cannot be retained by the monopolist but must be passed on to government or labour; or that the innovation would result in the wiping out of capital values prior to full depreciation. Such innovations, we may note, are more apt to be in production technique than in advertising, merchandising or management. That is why new devices in marketing and management continue in firms in which production techniques have become stabilized. Further increases of output over time, under these conditions, are likely to be accompanied by increasing unit costs unless the improvements in managerial techniques are very great.

We shall call a firm which has ceased to fear competition and which for one or several of the above reasons has slowed up the rate of introduction of innovations in productive techniques a "senile" firm. This senility is the last stage of the characteristic process of modern industry. The process begins with the rapid introduction of new productive techniques, the enlargement of scale, decreasing costs over time and unstable prices. It corrupts the market in the sense of reducing the degree of competition. This continues until price competition ceases. Monopolistic competition with price agreements follows and when a certain level of monopoly has been attained, say through cartel agreement, the firms may (not must) become senile, and experience increasing costs over time. Then the technical progress ceases, blocked by the very institutional pattern of which it was itself, in the first instance, the cause.

3. Conclusions

We come now to certain conclusions of the first importance for any future considerations of policy.

1. During a period when the firms in any industry are expanding lay-out with the accompanying technical changes

characteristic of modern industry, competition will become unstable because of the declining unit costs over time experienced by the firms. Under such circumstances an output "retreat" to a lower-cost, higher-demand price equilibrium, such as Marshall described, is not probable. Competition assumes the form of vicious price-cutting. Dog must eat dog until the remnant of the pack learns the wisdom of "co-operation". Prices will then be stabilized and monopolistic or unbalanced competition will result.

2. Expansion may continue, but there will be less urgent pressure for technical progress, and a good deal of excess capacity. Further increases in output over time may well be accompanied by increasing unit costs owing (a) to the rising marginal cost of management and (b) to the failure of the curved cost surface further to decline as technical improvement ceases. This is the condition of senility. It is clear that, as A. R. Burns argues in his *Decline of Competition*, empirical evidence of price stability is no guide to the competitive state of the industry. Consider the following possibilities:

(i-a) *Perfect competition* with little technical change in a manufacturing industry, such as Marshall so frequently postulates. *Prices will be stable.*

(i-b) *Perfect competition* in an agricultural industry where production must be laid down far in advance of marketing. *Prices will be unstable.* (Example, the poultry or hog cycle.)

(ii-a) *Imperfect competition* with decreasing costs over time. The market will be corrupted, or further corrupted. *Prices will be unstable.* There will be bankruptcy and concentration.

(ii-b) *Monopolistic competition* with a price agreement. *Prices will be stable*, as long as the agreement lasts. The slow trend of prices will probably be downward, there being no marked general price movement in a contrary direction.

(ii-c) *Monopolistic competition* in a senile trade. *Prices will be stable*, with a long-run upward tendency, there being no marked general price movement in a contrary direction.

Thus the process we have been considering in this chapter tends at first to disturb the stability of price, as well as to corrupt the market, to bring about a period of falling prices, sometimes

causing serious general disturbance in the economy, then to a period of renewed stability at sharply advanced prices held by agreement, with, however, a generally downward long-run trend, and finally to price stability with a generally upward long-run trend.

3. Apart from the disequilibrating effects of the process, is it, by reason of the effect on price, positive in its effect on welfare? To answer this question we must distinguish between the purely technical aspects of the process and its institutional entirety. Clearly the technical advance makes for welfare gains, in the strict economic sense, because it enables the society to produce more economic wealth with less effort, that is to say at a reduced real cost. We must ask, however, how far the institutional changes involved in the process affect the social distribution of the welfare benefits.

During the first part of the process the tendency will be for prices to fall at least as rapidly as costs. Thus the gain from the technical process will be fully passed on to the consuming public at large. The ill effects on welfare in this period will consist of the effects on employment of the instability and of the probable bankruptcies and amalgamations, and the consequent misdirection of resources. Shifts in demand during the process do not affect this general conclusion. If there is a positive shift in demand, so that there is some check on the decline in price, the process of concentration may be somewhat delayed, but the public will still be receiving more of the commodity at any price than it would have received had the process not been initiated. The same is true of a negative shift of demand, except that in this case the decline in price will be accelerated and the movement towards concentration speeded up.

In the second phase of the process when a price agreement is achieved, prices jump sharply,[15] and excess capacity appears in the industry. It is now apparent both that the full gains

[15] An excellent example is shown by prices per unit of potash delivered at United States ports. From 1924 to 1933 potash was controlled by a Franco-German cartel and prices were stable. Spanish and Russian supplies appeared on the market in 1934 and there was a brief period of intense "cut-throat" competition. At the end of 1934 a new, wider agreement was reached, prices jumped sharply, though not to their former level, but rose steadily as markets improved

are not being passed on to the consuming public and that there is a considerable degree of misallocation of resources. Public attention has generally been focused on the tendency to exploit the consumer by artificially maintained prices, but it is probable, though not demonstrable, that this evil has been exaggerated and is in fact not always very serious. Economists, however, will agree that the misallocation of resources and excess capacity herein involved, a condition which is very general throughout western society, has most unfortunate implications. In effect, it means that we, as a society, are failing to produce by a large margin what our techniques, skills, and productive capacity would enable us to produce, and that in a world still existing, for the most part, at a level far below the minimum necessary for even a bare subsistence. The degree of waste entailed is not positively known. Dr. Reynolds[16] estimated it at around twenty-five per cent of possible profitable production, but the experience of wartime production, even admitting the fact that wartime pressure led to a socially undesirable sacrifice of leisure and an undesirable employment of very young and very old persons, suggests strongly that twenty-five per cent is an exceedingly conservative guess.

If an industry moves through the whole process into senility, the welfare effects become increasingly bad, because not only do the gains fail to be passed on to the public, they fail to appear at all with the cessation of technical progress. The misallocation of resources continues, and is even accentuated if the firms remain, despite their senility, sufficiently powerful to enforce closure of entry.

We reserve further comment on the policy implications of this argument for Part Six of the present work.

---

with the upward thrust of the cycle. Here are the prices:

1932	1933	1934	1935	1936	1937	1938	1939
.619	.630	.300	.392	.440	.471	.471	.471

See "Canada and International Cartels", the *Report of the Commissioner on Combines to the Minister of Labour* (King's Printer, Ottawa, 1945), p. 3.

For further empirical evidence see A. R. Burns, *The Decline of Competition* (New York, 1936); Lloyd Reynolds, *Price Control in Canada* (Cambridge, Mass., 1940), and the various evidence before the Kilgore Committee of the United States Senate.

[16] Lloyd Reynolds, *Price Control in Canada* (Cambridge, Mass., 1940).

PART V

TEMPORAL SHIFTS IN LOCATION
THE CASE OF THE MARITIME PROVINCES

INTRODUCTORY NOTE

This Part is based on Chapter III of my *Economic Effects of the War on the Maritime Provinces of Canada* (Halifax, 1944). I am indebted to the Chairman and Board of Directors of the Institute of Public Affairs, Dalhousie University, for permission to reproduce some of the material published in that study. I should explain, especially to foreign readers, that my study of the Maritime economy was not widely distributed and is no longer available and I have been urged by some Canadian economists to reproduce the third chapter in some other, more readily accessible, form. This is because, fumbling though it might be, it was the first considerable attempt at applied location theory in the study of Canadian economic regionalism. I had considered attempting to have it reissued in monograph form, but the general trend of the argument was so close to the subject matter of this book that, on the advice of Dr. H. A. Innis and others, I have decided to include it here. It may be that foreign readers, uninterested in the special Canadian problems raised, will wish to skip this Part. The argument of Part VI is so arranged that they may do so without any break in continuity. This Part consists of a brief statement of the theory of temporal shifts in location equilibrium followed by an exercise in the application of the analysis to the problems of the economy of the Maritime Provinces.

CHAPTER XII

THE THEORY OF LOCATION EQUILIBRIUM

The gist of the theoretic thesis may be very simply set out. It is that the regional specialization and location of industry is brought into a certain balance or equilibrium, called "location equilibrium" by the forces associated with the law of comparative advantage. The causes of economic change, especially technical innovations, working on the size of firms, and finding certain institutional expression as well, upset this equilibrium and bring about shifts in location until a new equilibrium in location, satisfying the new conditions, is achieved. These shifts in location create serious social frictions and political problems and are pregnant with important issues of policy. The economic pressures have to become great to overcome social and, sometimes, political resistance. Thus the process is apt to be an explosive one, fraught with social peril.

Our historical thesis, which we shall attempt to support, may be briefly advanced as follows. The basis of Maritime manufacturing in the days of its prosperity was its transportation advantage with respect to the New England States and expanding purchasing power in the domestic market, protected by distance and poor transport from other possible North American competitors, and a natural advantage and accumulated managerial and labour skills in the chief industry, basic to the other development, namely lumbering and wooden shipbuilding. Confederation led to the loss of the transport advantage, not because of "unfair" freight rates, but because of the adoption after 1879 of a tariff system which changed the nature of the Canadian market, shifting the balance westward, and this coincided with a technological advance which destroyed the wooden shipbuilding industry. These changes were followed by other technological advances which increased the power of

the "agglomerating factors",[1] and led to a rapid disappearance of small-scale establishments, 1900-1920, a process that was brought to equilibrium by the "deglomerating factors",[2] exaggerated by the depression of 1929-35, and offset by "particular" factors in certain industries. These particular factors we cannot measure, but we can infer their presence by a comparative examination of certain industries, and this examination, based on the Dominion Bureau of Statistics returns, enables us to speculate with some degree of confidence about the probable ability of specific Maritime industries to retain the gains of 1939-45. Naturally this speculation has to rest on bold assumptions as to the nature of post-war trading.

1. The Theory of Location Equilibrium

We wish to preface this phase of our study by setting out the general theoretic conditions of location equilibrium. Let us suppose an area which we shall call "Westland" with a concentrated market of 5,000,000 people, and a second area, some 600 miles distant, which we shall call "Eastland", with a market of about 1,000,000 people; and let us suppose further that there are generally no differences in the income purchasing power of the people, their spending or saving propensities or tastes. Let us suppose that there are no marked differences in labour skills or in the rewards offered to labour, so that the labour costs per unit of output are essentially the same when comparable proportions of capital of like quality are combined with the labour. Again we shall assume at first that there are no "particular" factors arising from the possession of specially efficient or scarce natural resources giving a natural advantage

[1] Cf. C. J. Friedrich, *Alfred Weber's Theory of the Location of Industries* (Chicago, 1929), Ch. V. The "agglomerating factors" are those which, once there has been some concentration of industry in one area, give, from the mere fact of concentration, advantages to other industries in that area. They are the built-up supply of skills, both managerial and technical, the large-scale development of power, economies of size and natural industrial links, such as the growth of an engineering industry where there is a demand for machines, and so forth. The deglomerating factors are those which have a reverse action leading to industrial devolution, the rising cost of management, labour and material shortages and so forth.

[2] Friedrich, op. cit., p. XXVII.

in either area to the development of any specific industry. Finally we shall suppose economies of scale to be general to all manufacturing industries and to be such that, for a market of more than 1,000,000 and less than 5,000,000 increases in output will be accompanied by diminishing unit costs.

If the two areas together are protected by sufficient tariffs to close them to all outside competition, what will be the effect on the situation of manufacturing industry? Will it all gravitate to "Westland"? If not, at what point of concentration will an equilibrium be established? What are the conditions of that equilibrium? That is the theoretic problem underlying the question of localization of industry in general and in Canada in particular.

To solve it, even theoretically, we should need to know the shape of the cost functions of the different industries and the shape also of the joint demand functions, data, which, of course, it is impossible to obtain. We might, however, indicate briefly the general lines along which a solution would be sought. Industry would naturally gravitate to the "Westland" area. In this area it has a transport advantage amounting to 600 x the per ton mile freight charge on every ton of produce sold to 5,000,000 people. If, as we have assumed, economies of scale exist, then it is apparent that all industries in which these economies are realized will gravitate to the area of transport advantage in the bigger market. If the industries originally had been dispersed, those first to gravitate to "Westland" will be those in which economies of scale are sufficient to overcome the transport advantage so that the developing large-scale "Westland" firms will be able to undersell their "Eastland" competitors even in the latter's home market. In industries in which economies of scale are not sufficient to offset transport charges, the "Westland" firms will grow large while their "Eastland" competitors remain small, but the "Eastland" firms will not necessarily be driven out of individual existence. Thus the greater the economies of scale or the lower the proportion of transport to other costs, the greater the degree of concentration.

Let us ask, however, if there are any limits to this process of concentration. The limits which we shall discover emerge on

three levels of abstraction, (a) on the level defined by the assumptions already made, (b) on the level involved by allowing for the process to be a temporal one in which the concentration goes on industry by industry, affecting the wage rates in the industries in the two areas and (c) on the level discovered by dropping the assumption that there are no particular advantages. We must also question the generality of the assumption of economies of scale.

Even if wage rates are unaffected by the process of concentration and even if there are no particular advantages for specific industries, all manufacturing establishments will not gravitate to the larger area. Two limits are imposed on the concentration. The first arises because concentration involves imperfection in the market. The large "Westland" establishment grows and dominates the market and sets the price. It may become the owner of the small "Eastland" establishment, or it may not. However the matter of ownership is settled, in the imperfect market the higher cost establishment may continue in production, once competition is imperfect. If it is owned by the "Westland" firm and operated as an "Eastland" branch, the fact that it has higher unit costs may not result in its being closed. The market price is set at the point yielding highest profits to the parent firm. If at this price the "Eastland" establishment can operate for a quantity of production sufficient for the whole or the residue (in case the best profit point of the parent firm gives some spill-over for the "Eastland" market) of the "Eastland" market, it would pay to operate this establishment rather than to close it down and to increase output at increasing unit costs in the parent establishment. When we consider as well the important factor of the "good will" that would be lost if the "Eastland" establishment be closed down, it can be seen that in all cases where maximum profit points, and many cases where optimum cost points (not the same thing), are achieved in producing for the "Westland" market by the parent firm, the "Eastland" plant will survive, as long, of course, as it can produce at a cost that does not exceed the "Westland" price plus transport. If the "Eastland" establishment is independently owned the

same principle applies as long as it accepts the price leadership of the "Westland" firm and as long as, of course, the assumed conditions about rising unit costs hold for the "Westland" establishment for outputs beyond those required for the "Westland" market. The degree of inertia of the "Westland" firm will also affect its response to small-scale "Eastland" competition.

The second limit is imposed by the fact that economies of scale will not in all cases be sufficient to offset transport advantages. Though industries in such cases would concentrate in "Westland" to the extent that all large-scale firms selling to the "Westland" market would go there, small-scale firms would continue in the "Eastland", selling only to the "Eastland" market, but protected therein by their transport advantage within that market. That their security would be a tenuous one must be admitted, because the development of new techniques would be more possible to their rivals, so that, at any time, they might find themselves forced to meet the competition of rivals able to undersell them at home. As long, however, as the conditions we here assume existed, namely, that economies of scale were insufficient to offset the transport costs, the small-scale "Eastland" establishment could continue.

Once we admit that the concentration is not an immediate, extra-temporal adjustment, but a gradual process over time, we see that, even if wage rates were equal in the two areas at the beginning of the process, it would be impossible to assume that they would remain so. The concentration of industry and employment in "Westland" would increase the marginal productivity of labour and its demand price in that area and the slackening of demand for labour in "Eastland" would depress the price of labour there. Thus eventually, as the process continued, wage rates would grow apart to the point where labour costs had been so reduced in "Eastland" and so increased in "Westland" as to create a labour cost differential between the two areas equal to the transport advantage and scale economies enjoyed in "Westland". At such a point of cost equilibrium there would be theoretically no further movement of industry from one area to another and such industries as had not moved from "Eastland" would remain there.

Again, of course, there are bound to be particular advantages for certain specific industries in each area. When the particular advantages favoured "Westland" the concentration of that industry in which the advantages were enjoyed would be most complete. When the particular advantage in an industry favoured "Eastland", concentration would take place in "Eastland", "Westland", or there would be dispersal, according to whether the particular advantage was greater than, less than, or equal to, the general advantages enjoyed by "Westland". Indeed this case follows the general principles of specialization according to comparative advantage as familiarly developed in the theory of international trade.

Finally we must notice that the generality of the assumption of economies of scale is not justified by the facts. Professor Sargant Florence has found that in all three leading industrial nations, the United States, Germany and the United Kingdom, economies of scale are found in industries like engineering, heavy industry and the motor car industry, that, in textiles, boots and shoes and others, economies of scale appear up to what might be called "a medium scale of establishment", and that thereafter they do not appear, and that in brewing, brick-making and others there appear to be no advantages accruing with increases in scale.[3] The failure of economies of scale to appear with increases of size of establishment in certain industries may sometimes be because the most economical machine combination can be achieved on a moderate scale, as in the case in the boot and shoe industry;[4] sometimes the limit is imposed by what Professor Robinson calls the "rising marginal cost of management"; sometimes, as in the cases of baking, brick-making, "custom built" or "bespoke" tailoring, the very nature of the industry itself is opposed to large-scale organization. It would seem that the growth of the very large-scale establishment, with consequent concentration of industry, was

[3] Sargant Florence, "Economic Research and Industrial Policy", *Economic Journal*, Vol. XLVII, Dec. 1937, No. 188.

[4] H. C. Hillman, "Size of Formation in the Boot and Shoe Industry", *Economic Journal*, Vol. XLIX, June 1939, No. 194. See also Dr. Stern (ed.) *Labour Productivity in the Boot and Shoe Industry*, W.P.A. National Research Project, Washington.

most marked during the 'twenties, and that in the following decade the trend towards increases in scale was far less pronounced, was limited, in fact, to certain mass production industries, and, in some cases, was actually reversed.[5]

Thus there are certain types of industrial occupation in which economies of scale do not appear, or, if they do, appear very early in the process of enlargement, so that the area of the smaller market will continue to have its local establishments enjoying a location freedom from outside competition, but themselves unable to compete in the larger market. These industries will continue to be located in "Eastland", once developed there, or may even come to locate there, and will flourish along with other small-scale industries in which the economies of scale and other agglomerating factors have not been sufficient to enable the larger "Westland" concerns to wipe out their "Eastland" competitors and win a complete concentration. In addition, industries with particular advantages in "Eastland" will locate there and, under certain conditions, develop to large scale. Moreover, rising labour costs in "Westland" and falling labour costs in "Eastland" will result in a balancing of labour cost advantage against transport and capital cost advantages and will slowly apply the brakes to the general transfer or concentration of industries in "Westland", so that only those with the most marked advantages of scale and transport will completely locate in the more populous area.

No true equilibrium can therefore be achieved in the location of industry. The shift of population, in response to the wage differential, on the one hand, and technological change, on the other, continuously over time affect the conditions of each succeeding adjustment. Theoretically the movement of population and the characteristic technical advance should affect location in the same direction, making for greater concentration in the area enjoying the general advantages of the "agglomerating factors". The shift of population to the area in which

[5] Cf. *Investigation of Concentration of Economic Power* (The Temporary National Economic Committee for the 75th Congress), Monograph 13, "The Relative Efficiency of Large, Small and Medium-Sized Businesses" (Washington, 1941); also Thorp, *The Integration of Industrial Establishments*, U.S. Census Monograph (Washington, 1940).

higher wages are paid not only increases the disparity in the size of the markets, it ought, on the basis of static theory, to reduce the wage differential by lowering wages in "Westland" and raising them in "Eastland". This, however, does not happen. As long as there is technical progress in "Westland" and the market continues to expand, the "Westland" industry can absorb the increased labour force without reducing wages.

In the "Eastland" area, from which the immigrants come, their exodus does not necessarily lead to higher wages. The empirical evidence suggests the contrary, and this is so, because the migrants are ordinarily from the more aggressive, intelligent, far-sighted and skilled of the population. Their loss tends to reduce, rather than to increase, the productivity of the labour force left in "Eastland". (I am talking here of the industrial labour force. As I have elsewhere suggested, see *Canada and Peaceful Change*, ed. H. F. Angus, Toronto, 1937, the migration of farm workers from the Maritime provinces (1890-1914) probably increased the per capita productivity of Maritime agriculture.) Thus the further concentration of industry does not tend to have a positive effect on welfare levels on the area of emigration, but it does not necessarily increase the competitive advantage of the "Westland" area by reducing the labour cost differential.

Technological progress, on the other hand, generally works to improve the "Westland" advantage and to increase geographic concentration as well as certain exceptions. Some innovations, such as those which improved the efficiency of electrical generation from coal-powered turbines, have benefited the least-favoured industrial areas. The general process, however, which we have examined in Parts III and IV of this work, favours the development of the large-scale firm, and, in geographic terms, this means concentration in the area of the biggest market. Further, many of the important innovations have been in transport. As we have already seen anything which reduces the proportion of transport to total manufacturing costs favours geographic concentration. The history of Maritime manufacturing industry reflects in a very special way this influence of the technological process.

2. Technical Change and the Small Firm in the Maritimes

A. The Period of Reciprocity

Prior to Confederation, Maritime trade was largely with Great Britain, the West Indies and the New England States. By 1850 the seaboard colonies had recovered from the trade stagnation noted by Lord Durham[6] and after the Reciprocity Treaty of 1854 this recovery took the form of a prosperous business activity. The chief items of import were hardware, clothing and miscellaneous manufactures from Great Britain,[7] sugar, rum and molasses from the West Indies and the United States.[8] Exports were agricultural produce and fish to the West Indies and New England and lumber and wooden ships to Great Britain and the United States.[9] There was a small trade in fish and coal with the Province of Canada. To a large extent the provinces were self-contained. Many farms produced their own fuel, clothing and food, with small surpluses, usually from the woodlot, for cash sale or trade, which paid for the purchases of shoes, tools, sugar, tobacco and rum. Each village had its small grist mill, often a sawmill, a blacksmith shop, and even villages like Oromocto and Shediac had their shipbuilding yards. The unfavourable visible balance of trade was large and represented not capital imports but the value of ship exports and the carrying trade and was larger than any single item of visible exports.[10]

This flourishing economy was based on a great natural resource, white pine, which was the particular factor making for the success of the lumber and shipbuilding industry. The general location factor was the cheapness of sea-borne transport in Maritime bottoms of lumber, fish, mineral and agricultural produce to the chief markets, all of which lay across the seas; the ports, open the year round, were served by numerous

[6] Lord Durham, *Report on British North America* (Methuen ed., London, 1930), p. 142.

[7] Dominion Bureau of Statistics, *The Maritime Provinces Since Confederation* (King's Printer, Ottawa, 1927), pp. 81, 82, and Ch. IV, Table I.

[8] *The Maritime Provinces Since Confederation*, loc. cit.

[9] Ibid., loc. cit.

[10] Ibid., loc. cit.

river systems along which lay the farms and down which the lumber could be floated. There is no doubt, in spite of the lack of comparative transport cost figures, of the great general advantage enjoyed in the transport factor, adjudged by Weber as the principal determinant of localization.

At the then existing stage of technique, manufacturing was neither so specialized as it has since become, nor were the economies of scale so developed. In textiles and heavy industry the United Kingdom had developed pronounced advantages in which the colonies participated from the nature of their triangular trade with Great Britain and the West Indies. But imported fine textiles were for the well-to-do; local industry in homespuns supplied the mass of rural folk. Heavy industry in Britain might sustain the needs of Saint John and Halifax for engineering works and of the first railways, but local demands would be met by the improvised products of local blacksmiths, foundries and machine shops. Nor were these improvisations necessarily crude. The shipbuilding industry, and life in the isolated inland settlements had together developed an independent, inventive, jack-of-all trades skill that could improvise, initiate and adapt. One of the first successful steamboats of history was entirely engineered and built in Fredericton and sailed down the Saint John River to Saint John. The general situation advantage over other North American competitors in this naturally protected domestic market, the adaptable labour skill, the particular advantages arising from the West Indian trade in the manufacture of sugar, rum and molasses, early led to the establishment of variegated small-scale manufactures, of which, after the great trades of lumber and shipbuilding, the most important were tanneries, grist mills, distilleries, breweries, iron foundries, and, a little later in time, sugar refineries.

B. THE PERIOD OF NATIONAL POLICY AND MANUFACTURING GROWTH

Confederation and a national policy of tariff protection were believed, even by Maritime politicians,[11] to provide the conditions for a general development of manufacturing industry in

the new Dominion. It was confidently believed that manufacturing in the Maritimes would grow along with that of Ontario and Quebec. In the '80's, in spite of trade depression, new establishments on what was then a large scale of organization were started in textiles, shoemaking and sugar refining.[12] In spite of the trade recession which followed the abrogation of the Reciprocity Treaty in 1866 and the decline of shipping and shipbuilding, a decline which is apparent in the foreign trade figures,[13] the first two decades after Confederation were a period of great industrial expansion and venture. In Nova Scotia alone the number of establishments doubled[14] and capital invested increased from $6,000,000 to nearly $17,000,000.[15] In the Maritimes as a whole the expansion in manufacturing plant and capacity was of a similar order of nearly 300% from $12,000,000 or thereabouts,[16] to $33,000,000 in 1891.

The probable reasons for this expansion are not far to seek. There was first of all the spirit of optimism and confidence in entrepreneurial circles which was engendered by the National Policy, a confidence which was not shared by the traders, many of whom had been and remained consistent enemies of Confederation. In the second place there was the discovery of the Wabana fields, the rapid development of the Nova Scotian steel industry[17] and subsidiary steel products industries under the joint stimulus of cheap ore and the demand from rail construction. In the third place was the high protection afforded under

---

[11] "I am not, I think, over sanguine when I say the day is not far distant when the population in the Western country will be greater than in Ontario, and when the Maritime Provinces, with their coal, iron and water power will be the manufacturing centres for this vast Dominion . . . " Sir Leonard Tilley in *House of Commons Debates* (1879), pp. 1306-08. Quoted by N. McL. Rogers, *A Submission on Dominion-Provincial Relations and the Fiscal Disabilities of Nova Scotia within the Canadian Federation* (Halifax, 1934).

[12] Cf. Wilson C. MacKenzie, "Manufacturing in Nova Scotia", Report No. 32, *Reports of the Nova Scotia Economic Council*, Vol. IV (Halifax, 1939), p. 29.

[13] Exports declined from an all-time high of $16,332,341, with a further $10,000,000 of invisible exports in 1866 to $9,743,798, with less than $5,000,000 of invisible items in 1867. See Table I, Ch. IV, *The Maritime Provinces Since Confederation.*

[14] Wilson C. MacKenzie, op. cit., p. 29.

[15] *The Maritime Provinces Since Confederation*, p. 76.

[16] Prince Edward Island figures are not available for 1871.

[17] Cf. Ch. 5 of my *Economic Effects of the War on the Maritime Provinces of Canada.*

the new tariffs[18] to the domestic market, which, already developed, was of first importance in days when the typical scale of establishment was small and when neither communications made possible nor advertising desirable mass-produced "national brands". So a diversified manufacture of small scale, except for iron and steel, grew up in the Maritime provinces, based on particular natural advantages in the case of lumber products, iron and steel and sugar refining, and, for the rest, on general advantages of tariff protection and transport advantage with respect to the domestic (Maritime) market.

But during these twenty years of expansion, employment and output hardly kept pace with the investment of capital. Even in those early days were signs of the disappointments to come. While capital investment increased threefold in Nova Scotia and nearly as much in New Brunswick, employment in Nova Scotia increased by only 40% and in New Brunswick by only 5%.[19] With three times the capital Nova Scotian entrepreneurs increased their output value by only 80% and New Brunswick by only 16%.[20] The domestic market by itself was unable to take the produce of the new industry in anything like proportionately increased amounts, and the new policy exaggerated the effects of the cutting out of white pine and the death of the wooden ship in depriving the Maritimes of their previous advantages and markets.

C. THE THIRD PERIOD—THE NATIONAL MARKET AND CONCENTRATION OF INDUSTRY

The next stage of development was the period of the opening of the Canadian West and the creation of the national market.[21] The opening of the West caused a boom, on a national scale, which made all trade prosperous, and, particularly the steel trades. This effect was pronounced in the Maritimes, especially

[18] Before Confederation the Atlantic colonies had had revenue tariffs only, averaging around 10% ad valorum. The duties after 1879 were protective and ran three to five times as high.

[19] *The Maritime Provinces Since Confederation*, p. 76.

[20] Ibid.

[21] Say 1891-1914.

of course in Nova Scotia steel. The Great War accentuated this. The more permanent effect has received less attention. The whole centre of gravity of the Canadian economy was shifted westward. The Maritimes, from having had an ideal situation with respect to markets, with every transport advantage, now found themselves on the periphery, on the far extreme, of an unnatural, tariff-created national economy, with every general advantage in industrial situs enjoyed by Ontario and western Quebec competitors. Only the unusual and impermanent stimuli of rail-building and war could sustain high demand in this market for Maritime manufactures.

This new national economy developed its own characteristics, many of which were inimical to the prosperous development of a manufacturing industry in the Atlantic provinces. With these characteristics Canadian readers are too familiar for us to analyse them at length here, yet we may note them, because, sometimes, it is of the very obvious and familiar that the significance escapes us.[22] The new national system was based on an east-west flow of trade, linked by great transcontinental railway systems which had to be directly or indirectly subsidized by the public. It was a capital-importing system, accepting capital in the form of products of heavy engineering industry in Britain during the stages of first development (1871-1911) and in the form of manufacturing plant and machinery from the United States in its mature period (1921-1939). These capital imports were financed by the export of staples of extractive industry, wheat, later wood-pulp and paper and minerals. The miscellaneous products of mixed manufacturing were supplied by a domestic industry which grew up under the protection of a tariff wall. Contrast such an economy with the mercantile, free-trade economy of the Maritime provinces in the fifteen years before Confederation. Their exports were of raw materials and processed goods based on their natural resources, their imports of other manufactures, engineering products and tropical foodstuffs came from those to whom they sold. The Canadian economy rested on the

[22] At least so Hercule Poirot, who ought to know, repeatedly assures the inimitable Captain Hastings.

tariff system in the domestic market, a publicly subsidized transport system to enable manufactures to flow to the new lands and the grain to flow to the ports and hence to the world markets, and the production from virgin soil of great excellence of a highly specialized cash crop. The profitable agriculture was on the new land. Immigration flowed in not only from Europe but from the old provinces. During this period of the opening of the West the Maritime provinces had all their surplus population, nearly all their natural increase, drained off. The rural population, 1891-1911, actually declined.[23] Whereas in 1871, 18% of the Canadian population dwelt in the Maritime provinces and the centre of population gravity was in the vicinity of Montreal Island, in 1911 only 12½% of the population of Canada lived in the Maritimes and the centre of gravity had shifted westward towards Toronto. The domestic market for manufactured products was becoming a national, not a regional market, railways and low freight rates, and the development of new advertising techniques and media had destroyed the immunity of the isolated community market and had brought the small-scale productive unit into competition with large-scale outside firms. The Maritimes were unlucky in the conjuncture of technological and political events; the abrogation of Reciprocity and the creation of the Canadian national economy which destroyed their general situation advantages with respect to their chief markets coincided with the development of the steamship which destroyed their particular advantage in one industry, shipbuilding; with the opening of the Panama canal, which destroyed their particular advantage in another industry, sawn lumber; with the railway development in Canada, which broke down the isolation of the domestic community market for the product of small-scale manufacturing units, and with the rapid expansion in economies of scale in mass-production industry. Thus the period from 1891-1911 did not see the fruition optimistically expected by those who had undertaken the establishment of new mixed manufactures in the Maritimes.

[23] *The Maritime Provinces Since Confederation.*

During this period the number of manufacturing establishments in Nova Scotia decreased from over 10,000 establishments to about 1,100.[24] A similar decline occurred in New Brunswick. This decline must be properly understood. It does not mean an absolute decline in the total output or employment of manufacturing establishments. A similar decline in the number of establishments took place in Canada as a whole,[25] and it was happening also in the United States and Great Britain. It was a manifestation of the technical process within the structure of capitalist industry. New inventions, requiring large fixed capital, greatly increased the proportion of fixed to variable costs and made possible the increase of production at decreasing unit costs. This process, as we know, renders competition unstable, and competition in the market becomes imperfect. It is in part this process which leads to what Weber calls "agglomeration", because the general factors which lead to concentration of one industry in any given area will, unless there are offsetting particular factors, lead to similar concentration of other industries. Clearly the form of large-scale investment necessary can occur only if there is the promise of a sufficiently large market to absorb the greatly increased output. In Canada, the railways and tariffs made possible entrance to the entire national market, a rapidly expanding one, too, between 1891 and 1911, and the central concentration of population gave tremendous advantage to Ontario and western Quebec. This area possessed as well abundant water power which began in this period to be a cheap and efficient source of electricity. The Great Lakes, moreover, enabled Ontario to have cheap transport of American coal and ore for the development of a heavy metal industry. In that area, then, the new large-scale manufacturing industry was concentrated. Whereas in Nova Scotia the decline in the number of firms was of the order of 90%, while expansion in the value of output was about 150%, in Canada as a whole the decline in the number of firms was only about 50% (indicating that many new

[24] Wilson C. MacKenzie, op. cit., p. 30.

[25] From 1871 to 1931 the decline in the number of establishments in Canada was from 41,259 to 23,083.

establishments were begun, offsetting the absorption of small units) and the value of production increased by more than 1,000%.[26] The Nova Scotian increase in output, moreover, was more than accounted for by the expansion of the steel and sugar industries, in which there were particular advantages. Thus, allowing for these two exceptions, there was even a small absolute decline and, clearly, a most tremendous relative decline in the manufacturing industry in Nova Scotia as compared with that of Canada as a whole.

This process was not peculiar, as we have said, to Canada. In the United States where a similar extension of the western frontiers had had a corresponding effect on the national market, the New England States suffered in a manner comparable with the Maritimes.[27] We must clearly not be understood as making here a case for the sort of grievance that is sometimes voiced in the Maritime provinces against Confederation. The development of the nation state and the pressure of the technological processes combined alike in Canada and the United States to deprive the extreme east of general location advantages for mixed manufacturing industries and to bring about a more western concentration. This was inevitable, but it can, perhaps, be said that the incidence of the tariff was felt not only, as the late Rt. Hon. Norman Rogers showed,[28] in the form of increased costs in the Maritimes, but also in its adverse effects on industrial situs for Maritime manufactures.

That the Maritimes were aware of the nature of the change, after Confederation, that it was perceived, if not always clearly, by their people that it was primarily a transport advantage that enabled industries in Ontario and Quebec to obtain the further advantages of scale—the agglomeration factors—is evident from the attention given to the question of rail freight rates. But the pressure for reduced freight rates worked two ways. It

[26] These percentages are based on figures taken from official sources and quoted in my *Essentials of Price Theory*, p. 123.

[27] Cf. *The Maritime Provinces Since Confederation*, p. 24. See also unpublished thesis by Miss Dorothea Cox, "A Comparison of Population Changes in the Maritime Provinces and the Three Eastern New England States", in the University of New Brunswick Library.

[28] Cf. N. McL. Rogers, *Brief for the Province of Nova Scotia*, Ch. VIII.

might—and after 1926 did—bring more traffic to Maritime winter ports. It might enable Nova Scotia coal to enter the Quebec market, it might enable Maritime producers to effect slight reductions in cost from cheaper transport charges on materials, feeds, and machinery; it might help Maritime manufactures to compete in the national market, though here the assistance was too slight and the Maritime Freight Rates Act came too late to make any appreciable difference to the process of concentration we have just examined. But reduced rates also made possible increased access to the Maritime market for the products of large-scale plants in central Canada and increased the difficulties of Maritime enterprisers attempting to carry on on the basis of the local market. The adjustment of freight rates in the inter-bella period is a question that concerns the establishment of an equilibrium between the general location factors, the agglomerating forces, and the particular industrial and deglomerating factors (the finding of an equilibrium we believe to have occurred after the opening of the West and the First Great War in the period 1921-1939) and in that connection will be further examined.

A final general factor is labour cost. Our figures show that in the period when a temporary equilibrium was achieved, the adjustment occurred with an appreciably lower scale of wages in the Maritime provinces than in Ontario, or even Quebec. To some extent cheap labour has been an offsetting and deglomerating factor in favour of the Maritimes in the localization of Canadian industry. But this has been effect rather than cause. There is no reason to believe that the rates at which labour was offered in the Maritimes—the supply price of labour—was higher at Confederation than in Ontario. It was definitely lower than in the West. The provincial differentials, except for Quebec where a different culture and habit of thought did create a different supply price of labour, are the creation, not the cause, of industrial concentration. Industry did not gravitate to Ontario because labour was cheap there. But wages became higher there because industry concentrated there with success and profit and the productivity of Ontario labour increased so that higher wages could be paid. It is possible

that today lower wage standards in Quebec and the Maritimes may be an important factor in putting a limit to the concentration of industry in Ontario and may militate in favour of an equilibrium in which further Canadian development may be more regionally dispersed.

Thus we see that in the first stage of industrial development after Confederation, the period from 1871-1891, there was a considerable and optimistic, tariff-nurtured expansion of Maritime manufacturing plant. In the second stage, that of the opening of the west, 1891-1911, there was a shift in transport and general advantage in favour of the central provinces and a great concentration of manufacturing there. During the Great War Maritime plant, particularly in steel and lumber, was fully utilized. In the post-war period there was an adjustment and the establishment of an at least temporary location equilibrium in Canadian industry. It is to the study of that equilibrium period, between two wars, to which we now turn.

3. Location Equilibrium—The Maritimes and the Central Provinces

It is some such balancing of forces as we described in Section I, which leads to an equilibrium in the distribution and location of industry between regions, and, during the inter-bella stage in the evolution of Canadian industry, some such balance was struck in Canada. We do not suggest that it was permanent, any more than the balance that existed under Reciprocity was permanent. Great technological or political changes might disturb the balance and lead to a new seeking of advantageous situs and so to a further process of adjustment. But there is a distinction to be made, just the same, between the type of flux that went on during the development stages in the Canadian economy and the processes of growth and modification during what we have called the period of equilibrium and adjustment. There was a kind of balance in the Maritime economy based on lumber and sea-borne commerce in the Reciprocity period. This, in the next stage, the first twenty years of Confederation, was upset, although manufacturing sought locations in the

Maritimes. In the third period, 1891 to the Great War, when the West was opened up, an entirely new balance, for reasons we have shown, became necessary. Different causal factors were present affecting location advantage, and the Canadian economy adjusted in response to these with a heavy concentration of manufacturing in Ontario and western Quebec. In the period from 1921 to the second World War technical changes continued, industries grew and developed, there was change and flux, but the basic factors affecting location were unchanged, so it is a period in which, in spite of many other notable disequilibria—consider the cycle movement, for example—there was a comparative equilibrium in industrial location. There was no great tendency for a general dispersal of industry, as in France and England, nor for further concentration. The industries that had become concentrated remained so. The units which had continued to exist in the Maritimes were able to maintain their individual existence. Even the depression failed to lead to any great concentration, such as, under other circumstances, might have been expected.

The rapid expansion of the pulp and paper industry and the exploitation of the mineral resources of the Northern Shield, important developments of this period, giving Canada two other great export staples, one of which surpassed, the other rivalled, wheat as a commodity of export, but slightly modified the general advantages of Ontario as the situs of large-scale manufacture. The mineral developments were themselves chiefly in Ontario and went to increase the domestic market and to provide new raw materials and a stimulus for further diversification of its general and heavy manufactures. Pulp and paper mills were more widely dispersed, concentrating chiefly in Quebec, but also in Ontario and the Maritime provinces. In this industry particular situs advantages are very important and its development is based chiefly on raw material and power resources and independently of general manufacturing location advantages. At the same time its development in Ontario and Quebec was associated with an expansion of hydro-electric power generation which was also linked up with some of the

new metallurgical processes and, incidentally, made cheaper power generally available to industry.

The Maritime Freight Rates Act was a consolidation of the equilibrium rather than a condition of it. It was primarily designed to help Maritime ports by encouraging shipping via Saint John and Halifax, rather than to assist Maritime manufacturers to sell in the central Canadian market. Maritime manufacturers selling abroad, for example, and shipping via Montreal, as cargo liner schedules frequently made necessary, did not get the benefit of the 20% reduction. Again, there was a considerable period when the railways' pick-up and delivery service, introduced in Ontario and Quebec to meet truck competition, did not apply in the Maritimes and so left Maritime shippers under a handicap, equal to cartage charges, in competition on the central Canadian market. But on the whole the rates established by the Act enabled Maritime producers to compete in Central Canada, with their transport disadvantage partially offset by preferential rates, and thus acted as an equilibrating factor. Thus on pig iron the rate (1934) from Sydney to Montreal (956 miles) was $4.30 per ton, from Sault Ste. Marie (622 miles) $5.60 was the normal rate, the rate to meet water competition in summer was $4.25. On steel billets the Sydney-Montreal rate was $4.30, the Sault Ste. Marie rates were $5.60 and $4.50. On enamelware the rate from Amherst to Montreal (701 miles) was $8.60 per net ton, from Toronto (334 miles) it was $8.00. On bunker and forest products the rate from Campbellton, N.B. to Montreal (461 miles) was 19½¢ per hundred pounds, from Sudbury, Ontario (442 miles) the rate was 28¢, from Orient Bay, Ontario (870 miles) the rate was 34¢.[29] These preferential rates helped Maritime industries with market connections in the central and western provinces to maintain their position there. It must be remembered, however, that in reverse, so to speak, they had the effect of putting the weaker Maritime industries in a somewhat more difficult position in the domestic market. On the

[29] Rate calculations are from *A Submission to the Nova Scotia Royal Commission of Economic Inquiry*, 1934, by Enamel and Heating Products, Ltd. I am obliged to Enamel and Heating Products Ltd. for permission to use their material.

whole they would seem to have stabilized rather than offset the trend set up by the other location factors towards a location equilibrium.

Thus this period is an ideal one for our study. Our theoretic analysis suggests that such industrial establishments as existed in the Maritimes during this period did so because (a) they enjoyed particular situation advantages there, as did iron and steel, pulp and paper and sugar refining, or (b) because they were units tied in through ownership integration with central Canadian firms and were operated by them in an imperfect market, as was probably the case of cotton textiles, or (c) because as small units who accepted price leadership it was cheaper to permit them to compete, as long as they did not upset the price, than to destroy them, or (d) because they were units in an industry in which economies of scale were not sufficient for the large-scale firms to reach down to the Maritimes and sell cheaper in the local market than firms on the spot, or (e) because they were units in an industry in which there were no economies of scale, an industry naturally dispersed in widespread, small-scale units, such as bread-making, creameries, custom-built interior fittings for stores or offices, potteries and handicrafts.

If, during this period, a Maritime industry showed signs of growing as fast as, or faster than, its competitors in Quebec or Ontario there would be *prima facie* evidence for the view that it had, for one of the reasons outlined, off-setting advantages, at least up to a certain scale of size, in the Maritimes. If it showed growth, but at a slower rate than in the central provinces, it could be supposed that it was developing by permission, or because it did not pay to interfere, to serve the Maritime market; a growth strictly limited by the size and comparatively static state of the Maritime population. If an industry failed to develop in the Maritimes during this period, or declined, while at the same time it grew in the central provinces there would be *prima facie* evidence to support the belief that the general factors favouring location in those provinces were still operating. If we could obtain some such indication of comparative rates of growth we should have a criterion for assessing

the probable trends generated by the second World War. Subject to modifications from changes in the European market and alternative sources of supply, we should suppose that gains in employment and sales registered by industries which had no wartime increase in plant capacity could be retained only in cases where the industry was one which had shown during the twenty years of the "adjustment period" that it enjoyed some special advantages in the Maritimes. Increases in output in other cases would, if unaccompanied by plant extension or improvement, appear to be passing and ephemeral, the result only of the vast demands of war, to disappear when those demands were past. New equipment and increases in scale of plant, brought about by war, would be important in a permanent way if (A) they occurred in industries possessing particular advantages in the Maritimes, or (B) if they increased the freedom of firms selling in the local market from central Canadian competition under the conditions of (c) and (d) above, or (C) if they were of sufficient scale to give a hitherto small-scale "domestic" Maritime establishment sufficient physical size to enjoy economies of scale adequate to enable it to compete nationally with large-scale establishments in other provinces.

CHAPTER XIII

THE COMPARATIVE RATES OF DEVELOPMENT OF CERTAIN MARITIME INDUSTRIES IN THE INTER-BELLA PERIOD[1]

1. SOME GENERAL CONSIDERATIONS

The object of this chapter is to examine certain industries that have become localized in the Maritime provinces to see (a) how these industries have grown in size in comparison with their growth in the provinces (Ontario or Quebec) in which they tended to concentrate, (b) how far the growth in size has been correlated with changes in efficiency, (c) what has been the comparative development of labour and power costs and (d) what the indications, if any, are as to the balance of general and particular location factors with respect to these industries in the Maritimes. In the cases of pulp and paper and the steel industries we have given here only the briefest mention indicating points of contrast with other manufactures. (These industries are fully studied in Chapters 4 and 5 of my *Economic Effects of the War on the Maritime Provinces of Canada.*) We have not attempted to define "an industry", but have accepted the Dominion Bureau of Statistics classification of Canadian industries.[2]

[1] This chapter is based on the *Census of Industry Returns* which were made available by the Dominion Bureau of Statistics, which also, most generously, aided by making extensive compilations and running many series of correlations. I am particularly indebted to Mr. A. L. Neale, sometime Chief of the Industrial Statistics and Social Analysis Branch of the Bureau, who gave me office space in his Branch and whose staff helped me enormously, and to Miss L. J. Beehler who directed the work of compilation. This section is published by permission of the Bureau and has been edited in conformity with the Dominion Statistics Act, and I take this opportunity of thanking the Dominion Statistician for permission to publish this material. Because of limitations of time and staff we were unable to make the survey a comprehensive one, and were able to study only the following industries: Pulp and Paper, Boots and Shoes, Primary Iron and Steel, Castings and Forgings, Woollen Cloth, Woollen Yarn, Cotton Yarn and Cloth, and Biscuits and Confectionery.

[2] Something, too, should be said of our methods in measuring size of establishments. The data of the *Census of Industry* exist originally in so-called "schedules"

The *Census of Industry* does not provide any data from which comparative costs may be calculated, and, because of the nature of its returns on capital invested, it is not possible to calculate actual or realized profits on investment. Hence we cannot make from our data any direct study of costs or any comparison of the efficiency of establishments according to the generally accepted business criterion of efficiency, viz., profitability. We

---

which are annual returns made by each individual establishment and which contain detailed information of the firm's employment, showing monthly variations, its wage bill, in more recent years the wage distribution, short-term capital costs, fixed capital investment, power installation, details of power costs, and finally, details and totals of physical and gross value production. From these figures it is difficult to select any that give a satisfactory measure of size. The Bureau of Statistics measures establishments both in terms of labour employed and value productivity, but both these measures are open to most serious objections. Both change with the cycle, and thus indicate that establishments grow smaller during depression periods and larger during good times. This is only true if we define the size of establishment in terms of the units by means of which we measure it. If, however, we think of a large establishment as being one of a large plant, measurement in terms of employment is most misleading, as is measurement in terms of product. Even if we could eliminate the cyclical trend by well-known and somewhat arbitrary methods of adjustment, there would remain almost insuperable objections to both these units. As Dr. Spurgeon Bell has shown (Spurgeon Bell, *Wages, Productivity and National Income*, Brooking's Institute, 1940), the long-term trend has been towards larger plant establishments with greatly increased productivity but with reduced labour employment.

For our purposes, therefore, we needed a unit of measure which corresponded to the popular and proper economic definition of the size of establishment. Unfortunately fixed capital investment as shown on the schedules is not a trustworthy unit of measure, because the firms reporting used different accounting methods in estimating both capital worth and depreciation.

Power installation and power consumed is a guide in certain industries to the physical capacity of the plant, but in certain other industries it is no guide at all. For example, in the pulp and paper industry, the power installation of a plant manufacturing ground wood will be very much greater than that of a plant using one of the chemical processes, yet the two plants may be equal in size according to any other measure, capital investment, physical capacity, value product, or employment.

In certain industries the returns show the physical capacity of the plant in terms of either tonnage capacity, as in the case of pulp and paper, or of machine installations, as in the case of textile plants and primary iron and steel. In such cases we had, of course, an ideal stable unit of measure of physical size. We could not use such a measure for all the industries examined, but when we had it we used it. After all consistency in the unit of measure from industry to industry was not important. In each case we were comparing the industry in the Maritimes with the same industry, measured in the same unit, in Ontario or Quebec. As long as the same unit was used consistently in application to any one industry it did not matter seriously if different units were used for different industries. Thus, where possible, we used a stable physical unit of size. Where such a unit was not available we used employment figures partly corrected for cyclical trend.

have had to assume, therefore, that wages paid and the net value productivity of labour were indications of efficiency. The business man will be quick to object that these are not measures of efficiency at all, and we must admit that they do not necessarily correspond to or coincide with profitability. But it is probable that both our criteria are fairly reliable indications of profitability. Profits have to be paid out of what is left after the wages bill is met, and the higher the net value productivity of labour, the greater, other things being equal, this residue will be. Of course high capital costs may leave little of this residue for profits, but generally speaking the rate of return on capital is higher the more productive the labour employed. Also high wages are generally paid when the employer can "afford" them. Thus a high rate of reward to labour is usually associated with good returns to enterprise and capital.

In any case we may well challenge the general acceptance of profitability as a criterion of efficiency. From our point of view the ability of an industry to maintain employment at good wages is one test of its social efficiency; its ability to get a high net value return for a unit of labour employed is surely another, for this means that its labour is of good skill and efficiently (productively) combined with its capital instruments. We are consequently prepared to justify our use of these measures of efficiency on grounds other than that of expediency, although it is also true that these tests of efficiency were about the only ones that could conveniently be made with the *Census of Industry* returns.

2. Number and Size of Establishments

An industry may grow in one of three ways: it may increase its productive capacity by the multiplication of manufacturing units, it may increase capacity by a great expansion in the size of plant of existing units, or it may, without increasing capacity, use much more intensively the plant that it has. In modern manufacturing the last method is a characteristic of short-term expansion in the boom period of the cycle, but it is not the sort

of growth in which we are interested, for it does not represent a permanent development, nor indicate the long-term trend towards concentration or dispersal of plant. The increase in number of plants, if pronounced in an area, may be an important indication of concentration, but since this type of development is not characteristic of most modern manufacturing, as, on the other hand is the expansion of plant in existing establishments, it is this last which is the most important manifestation of regional industrial development.

The statistical evidence of growth of plant in the eight industries we have studied is summarized in the following composite tables:

Growth in Number of Establishments in Certain Industries in the Maritime Provinces Compared with Growth in Ontario and Quebec

Industry	Number of Establishments by Years													Net Growth 20 Years	%age Growth	Industry
	1920	'26	'30	'31	'32	'33	'34	'35	'36	'37	'38	'39	'40			
1a Pulp & Paper Maritimes	12	13	13	—	—	11	—	—	—	—	—	—	11	-1	-8½	1a
1b Pulp & Paper Quebec	45	50	49	—	—	42	—	—	—	43	44	—	45	1	2	1b
2a Boots & Shoes Maritimes	5	—	6	—	—	—	—	—	—	5	6	—	6	1	20	2a
2b Boots & Shoes Ontario	59	63	70	—	—	78	—	—	—	79	82	83	86	41	71	2b
3a Primary Iron & Steel Maritimes	3	2	4	4	4	4	4	4	4	4	4	4	4	1	33⅓	3a
3b Primary Iron & Steel Ontario	17	13	15	15	14	14	14	14	15	15	15	15	15	-5	-111 3/17	3b
4a Castings & Forgings Maritimes	26	—	21	—	—	19	—	18	—	19	—	19	19	-7	-30	4a
4b Castings & Forgings Ontario	65	—	87	—	—	93	—	91	—	87	—	84	86	21	32⅓	4b
52 Woollen Cloth Maritimes	3	3	4	4	4	4	4	4	4	4	4	4	5	2	66⅔	5a
5b Woollen Cloth Ontario	31	31	34	34	34	38	38	38	38	32	32	32	37	6	19⅓	5b
6a Woollen Yarn Maritimes	5	7	7	7	7	7	7	8	8	8	8	9	9	4	80	6a
6b Woollen Yarn Ontario	15	17	—	18	18	20	23	21	21	20	19	19	20	5	33⅓	6b
7a Cotton Yarn & Cloth Maritimes	5	—	—	4	4	4	4	4	4	4	4	4	4	-1	-20	7a
7b Cotton Yarn & Cloth Quebec	11	14	14	14	15	16	16	14	14	14	14	14	15	4	36-4/11	7b
8a Biscuits & Confectionery Maritimes	13	—	16	—	—	17	—	15	—	15	—	11	12	-1	-7-2/13	8a
8b Biscuits & Confectionery Ontario	104	—	102	—	—	82	—	78	—	77	—	78	82	-22	-21	8b

Growth in Size of Establishment in Certain Industries of the Maritime Provinces Compared with Growth in Ontario or Quebec

Industry	Unit of Measure	*Size of Median Establishment by Years 1920	1926	1930	1933	1935	1937	1939	1940	%age Growth	Industry
1a Pulp & Paper Maritimes	Capacity tonnage per annum	10,000	11,500	29,200	31,200	—	37,400	37,500	37,500	375	1a
1b Pulp & Paper Quebec		19,500	16,400	37,500	37,500	—	62,500	62,500	50,000[1]	320.5	1b
2a Boots & Shoes Maritimes	$ value of gross output[2]	302,000	228,000	185,000	159,000	—	127,000	122,000	—	-59[3]	2a
2b Boots & Shoes Ontario		181,000	225,000	212,000	166,000	—	175,000	158,000	188,000	4	2b
3a Primary Iron & Steel Maritimes	Employees[4]	663	613	494	192	408	579	563	645	-3	3a
3b Primary Iron & Steel Ontario	Employees[4]	324	238	343	212	403	540	562	752	132	3b
4a Castings & Forgings Maritimes	Employees	27	—	21	17	25	17	13	11	-59	4a
4b Castings & Forgings Ontario	Employees	20	—	17.5	17	19	19	20	17	-15	4b
5a Woollen Cloth Maritimes	Looms	21	—	—	8	9	7	Not calculable		-66	2/3[5] 5a
5b Woollen Cloth Ontario	Looms	20	—	30	25	25	35	—	30	50	5b
6a Woollen Yarn Maritimes	Spindles	200	—	220	225	200	240	—	260	30	6a
6b Woollen Yarn Ontario	Spindles	1,064	—	2,200	3,200	3,168	3,500	—	3,650	247	6b
7a Cotton Yarn & Cloth Maritimes	Possible Spindle Hours	—	—	57	72	—	60	—	61	6.6	7a
7b Cotton Yarn & Cloth Quebec		—	—	198	235	—	220	—	178	—	7b
8a Biscuits and Confectionery Maritimes	Employees[4]	144	—	105	70	83	96	129	126	-12.5	8a
8b Biscuits & Confectionery Ont.	Employees[4]	68	—	56	69	66	77	79	81	19	8b

Notes: *For reasons to which we refer in the next text it was found impossible to use the modal establishment and so the median was ordinarily used.
1. Approximate. 2. Corrected for general price change. 3. 1939 Last year of calculation. 4. Average not median establishment. 5. 1937 Last year of calculation.

An examination of these tables and of the data on which they are based reveals the great variety of development experienced by the different industries. For the most part the "disappearance"[3] of the small-scale firm was complete so that there was not any great change in the numbers of establishments in the twenty years under examination. On the whole there was a decline in the number of establishments in both regions in biscuits and confectionery only, but, in the Maritimes, there were declines in pulp and paper, castings and forgings and cotton yarn and cloth, and, in Ontario, there was a decline in primary iron and steel. But Summary Table II reveals more significance in the comparative trends. In pulp and paper the growth in size of establishment was marked in both the Maritimes and Quebec. The detailed figures of establishments showed further that three distinct modes of size appeared when the establishments were classified. There were the small groundwood mills, which, on the whole did not increase in size but did diminish in number over the twenty-year period. Then there were the medium-sized pulp and paper mills, the mode of which shifted regularly to the right, that is became larger, during the period. Finally the very large newsprint mills became more numerous and grew ever larger.[4] This period was one of great technical development in this industry. Thus it shows, rather uniquely among those we have studied, the process of formation of large-scale units, a process pretty generally completed in the other industries. Both in the Maritimes and in Quebec the process was similar, small units closing up, new and larger units coming into production, and a rapid expansion in the capacity of the existing units. The Maritimes, it is true, had only one of the very large

[3] We recall that we are using this term rather inexactly to mean the process in which small-scale firms went out of existence. When the process came to an end the "disappearance" as a process was over, but the disappearance was complete in this sense only, not in the literal sense.

[4] We cannot here reproduce the detail of the data. The scatter diagrams showing by individual establishments the changes in size over the twenty years would both delay the general reader and expand this volume beyond reasonable limits. We must ask the reader to accept our statements of fact about the detail of the statistical data as being accurate statements of or fair inferences from the base data.

newsprint establishments of the sort which sprang up in Quebec, but the general trend was similar.

The pulp and paper industry was the only one, among those we studied, in which in the Maritimes there was any important general plant expansion. In boots and shoes, an industry in which the most economical size is reached at a moderate scale of establishment and in which that scale was achieved in most Maritime establishments by 1920, there was one very small new establishment started, but otherwise there was no new capital formation; on the contrary there was a marked depreciation in capital invested and a diminution of employment and average output. In the Ontario industry the basic data show no "over-all" change in average size, measured by output, but they do show that this seems to be because there was no important expansion of demand. There was some new capital formation and the substitution of capital for labour, a process which, in contrast to what happened in the Maritime industry, greatly enhanced the productivity of labour.[5] Castings and forgings and small foundry business was a small-scale industry generally declining in numbers and size, and there was little difference in the decline in both areas. But in primary iron and steel—and in the other secondary steel products—there was an expansion, in Ontario, in the number and size of establishments, a great amount of new capital formation, a general development of the industry, while in the Maritimes there was little change until the war.

In the textiles the Maritime industries remained about the same in size and numbers, and there was a little capital formation—one new large establishment in woollen yarn in the late '30's and a change-over (qualitative change) in machine equipment in one cotton yarn and cloth plant. On the whole the textile group has not flourished in Canada since the last war, and has depended on low wages and high tariffs,[6] to maintain minimum profitability. Nevertheless, notably in woollen yarns,

[5] Dr. Bell shows that during the same period in the United States the characteristic function of new technical processes was to substitute capital for labour rather than to increase output. Cf. Spurgeon Bell, op. cit.

[6] Cf. *The Report of the Royal Commission on Price Spreads* (Ottawa, 1935).

there was a growth in size of establishment in Ontario and Quebec, very much more marked than in the Maritimes, and capital formation, often of the sort substituting for labour, was on a much greater scale. In biscuits and confectionery the basic data show the appearance of the small-scale unit of production of the "Laura Secord" type in Ontario, a type of high-grade confection-making which seemingly combines the advantages of large-scale firm organization with those of small-scale production units. The appearance of numerous small-scale imitators make the average figures record a change in size of establishment which is misleading. The large-scale chocolate and candy makers in Ontario grew larger, and there was a development there of chewing-gum plants and a greater variety in the industry than in the Maritimes where there was a general decline in the size of the industry.

In general the only Maritime industry to record real growth, comparable with that of its Ontario (or Quebec) rival, was pulp and paper, in which very marked particular advantages are enjoyed in the Maritimes, advantages not the least of which is the transport advantage to the New England, New York and British markets, as in the "good old days". In iron and steel, another industry in which the Maritimes have a particular advantage, the Maritime industry was static while the Ontario industry grew and became diversified. In boots and shoes the Maritime industry became much smaller while the Ontario industry grew. In the textile groups the Maritimes showed slight growth; there was greater growth in Ontario and Quebec. In biscuits and confectionery, while in both the Maritimes and Ontario there was a decline in the scale of enterprise this was accompanied in Ontario by diversification of enterprise and development of a small-scale but highly remunerative luxury trade, while in the Maritimes there was no such offsetting factor to the general decline. Thus, though there was no marked further concentration of the industries we have examined after 1920, it was apparent, with one exception, that established industry in the Maritimes was not growing, that there was still a gradual drag exerted by the agglomeration factors toward the centre; and, we must remember, new industries, automobiles,

radios, etc., were always appearing in Ontario and Quebec, seldom in the Maritimes. If the Maritimes were not being further drained of their small-scale enterprises in favour of the central provinces, if a kind of equilibrium was established, it was one which gave to the Maritime industries, in general, a markedly slower rate of development than to the industries of the central provinces.

3. Productivity and Wages

The willingness of business entrepreneurs to expand their plants or to venture capital in new plants, resulting in physical growth, may be an indication of their expectations of profit and consequently an indirect index of efficiency. Thus the relative rate of growth of industries in the Maritime provinces is some guide to the comparative efficiency of the various industries and to their probable location advantages. But it is a very uncertain guide. Social efficiency, as we have seen, is not necessarily the same thing as profitability, i.e., the ratio of revenues to costs, and, in any case, efficiency even as we propose to measure it does not arise solely from location factors. In general the more efficient establishments, in the sense of wage-paying ability or net value productivity of labour, will be those well situated with respect to resources and markets and the other location factors, but the skill and judgment of management will always be a variable and partially independent governing factor. We say "partially independent" because it is part of our thesis that the tendency of managerial ability to migrate to the areas of industrial concentration is one of the general location advantages of these areas. Nevertheless there will always be individual exceptions to this rule and these exceptional cases, occurring in particular industrial establishments, will give rise to special instances of high efficiency independent of the other location factors.

Finally there is the fact that sheer growth in size has not always been associated with efficiency. The expectation of economies of scale may be realized up to a certain size of plant. Beyond that size further expansion may be accompanied by higher costs and a lower ratio of revenues to costs. The

Federal Trade Commission's Report on the Relative Efficiency of Large, Medium-sized and Small Businesses, to which we have already referred,[7] presents evidence strongly suggesting that efficiency, measured by unit costs and by earnings per unit of capital invested, is less in the largest scale of establishment than in units of medium size. The Committee's Report further suggests that the motives for increasing scale are numerous and may be directed towards increased power, towards creating an imperfect market or other objectives which are not associated with business efficiency as defined by the Committee. Unfortunately the measurement of efficiency and the unit of measurement of size of unit chosen by the Committee do not satisfy the most rigorous requirements so that their evidence while suggestive is not conclusive.

We need, however, more direct evidence than statistics of growth, to enable us to draw any conclusions as to the relation of size of plant and efficiency and as to the comparative regional development of location advantage between central Canada and the Maritimes. The *Census of Industry* returns enabled us to calculate the value productivity of labour, which may be defined as $\frac{V-C}{L}$, when V is total value produced, C is cost of materials and fuel, and L is the number of labour units employed. We observed (a) the correlation between value productivity of labour and scale of establishment and (b) changes over the period in the value productivity in Maritime and central Canadian establishments.

The regression equations of net value productivity against size were not always very valuable because frequently there were too few instances to yield valuable results when treated mathematically. Scatter diagrams, however, were in such cases highly suggestive. Our results may be summarized in the following table.

[7] 76th Congress, Investigation of Concentration of Economic Power, op. cit.

Summary of Comparative Development of Efficiency of Establishments Measured by Net Value Productivity of Labour, and Correlation of Efficiency with Size of Establishment

Industry	*Regression Equations* 1930	1940	*Results of Scatter Diagrams*	*Change of Efficiency of Median Establishment* 1930	1940	*Industry*
1a Pulp & Paper Maritimes	r = .87	r = .81	Superiority of larger mills over groundwood mills of small scale	1800	3000	1a
1b Pulp & Paper Quebec	r = .61	r = .74	Three modes all shift to right, indicating superiorities with increase in scale	2100	3250	1b
2a Boots & Shoes Maritimes	Insufficient instances		Medium scale superior to small	1700 (1926)	1490 (1939)	2a
2b Boots & Shoes Ontario	r = .14 ('26)	r = .35 ('39)	Medium scale and large scale by 1939 had superiority over small	1900 (1926)	2250 (1939)	2b
3a Primary Iron & Steel Maritimes	Insufficient instances		No clear trend	2250	2200	3a
3b Primary Iron & Steel Ontario	Insufficient instances		No clear trend until '40, when medium scale was superior to both large and small	2100	3200	3b
4a Castings & Forgings Maritimes	No positive correlation		Superiority of medium scale	1500 (mode)	1375 (mode)	4a
4b Castings & Forgings Ontario	No positive correlation		Superiority of medium scale	2000 (mode)	1900 (mode)	4b
5a Woollen Cloth Maritimes	Insufficient instances		No trend	900 (av.)	1600 (av.)	5a
5b Woollen Cloth Ontario	r = .32	r = .54	Superiority of medium and large scales	1375	2150	5b
6a Woollen Yarn Maritimes	Insufficient instances		Medium scale superior	1275	1350	6a
6b Woollen Yarn Ontario	r = .49	r = .38	Medium scale superior	1300	1675	6b
7a Cotton Yarn & Cloth Maritimes	Insufficient instances		All much same size	Not calculated		7a
7b Cotton Yarn & Cloth Quebec	Insufficient instances		All much same size	Not calculated		7b
8a Biscuits & Confectionery Maritimes	Industry too diversified		Large establishments superior	1400	1450	8a
8b Biscuits & Confectionery Ontario	Industry too diversified		No trend	2000	1800	8b

In general these results, as far as they go, tend to support the findings of the Committee of Investigation of Concentration of Economic Power[8] that superior efficiency was most usually enjoyed by plants of medium scale. Our charts, however, indicated that in all industries except iron and steel, cotton textiles and biscuits and confectionery[9] there were economies of scale—or at least advantages—that came with increased scale up to a certain optimum point of size. But the optimum size, differing of course for each industry, might be reached, as in castings and forgings or boots and shoes, at a fairly moderate scale of enterprise, or, as in the case of pulp and paper, it might be at a large scale. But these conclusions, though of general economic interest, are less to our immediate purpose, than the evidence from this analysis that in all the industries, except cotton textiles, for which the evidence was inconclusive, and biscuits and confectionery, the central Canadian industries maintained or increased their efficiency over the period examined, and this improvement was associated with some increase in the scale of establishment. In all industries where there was evidence of an optimum scale, i.e., all except primary iron and steel, cotton textiles and biscuits and confectionery, the Ontario or Quebec industry developed to the point whereby 1940 the median establishment approached the optimum scale of size. In no industry, except pulp and paper, did the Maritime median or average approach optimum size. Moreover in boots and shoes, iron and steel, and castings and forgings, the Maritime industries actually lost in efficiency during the period, and only in pulp and paper and woollen cloth did the Maritimes register gains comparable to those of the central Canadian establishments. In biscuits and confectionery the Maritime establishments appear to have gained while those of Ontario lost in efficiency, but this is somewhat illusory. The Maritime establishments were chiefly candy and biscuit makers and they seem to have held their own with their

[8] Vide supra.

[9] These exceptions are noted not because the contrary principle obtained in them, but because there was no positive evidence as to the relation of scale to efficiency revealed in the statistics of these industries.

Ontario competitors, though on the whole they were slightly inferior. The Ontario average for net labour productivity was reduced in 1940 by the appearance of a group of miscellaneous small firms engaged in various associated confectionery trades, with low productivity. In every industry the inferior productivity of the Maritime labour-capital combination is pronounced, running from as little as 7% inferior in pulp and paper where near equality is approached to as much as 80% in primary iron and steel. Thus the general advantages of central Canadian industry are pervasive; they are apparent in all the industries examined, and they are almost invariably associated with a larger scale of enterprise, which our theoretic analysis suggests arises from their larger domestic market and the agglomerative location factors. But in pulp and paper the general advantages are slight and the Maritime industry has progressed in efficiency as it has developed in size at a rate comparable with that of the industry in Quebec. In biscuits and confectionery and in woollen cloth the Maritime industries have at least held their position. They are inferior in efficiency as we have measured it, but over the period we have examined they have narrowed rather than widened the gap of superiority enjoyed by their Ontario competitors. In castings and forgings there was little change in the relative positions as between the Maritimes and Ontario. In boots and shoes, primary iron and steel, and woollen yarns the gap was widened, the general advantages were becoming more pronounced. When we relate these findings to our statistics of growth we see, in effect, that pulp and paper was the only industry in which the particular advantages in the Maritimes were such as to permit actual growth and development. In biscuits and confectionery and in woollen cloth the Maritime industries are safe to carry on on their present moderate scale, even to develop slowly with growth of the domestic market. In primary iron and steel the Maritimes have a static, but not necessarily a decaying industry. In cotton textiles our study showed qualitative rather than quantitative changes. The Quebec industry had a lot of new capital formation in the period after 1937, most of it to enable plants to specialize in particular products and to diversify the manu-

Comparison of Average Annual Earnings of Labour in Certain Industries in the Maritime Provinces and in Quebec and Ontario

Industry	*Average Annual Earnings of Labour*						
	1920	*1926*	*1930*	*1933*	*1937*	*1939*	*1940*
1a Pulp & Paper Maritimes	1,142	1,018	1,244	995	1,288	1,246	1,392
1b Pulp & Paper Quebec	1,341	1,293	1,322	1,026	1,370	1,377	1,564
2a Boots and Shoes Maritimes	1,008	881	956	587	618	763	—
2b Boots and Shoes Ontario	1,020	1,025	961	839	918	914	1,069
3a Primary Iron & Steel Maritimes	1,837	1,086	1,303	1,176	1,443	1,442	1,644
3b Primary Iron & Steel Ontario	1,910	1,750	1,740	1,312	1,499	1.564	1,720
4a Castings & Forgings Maritimes	1,240	—	1,248	978	1,138	1,146	1,180
4b Castings & Forgings Ontario	1,477	—	1,210	764	1,176	1,183	1,355
5a Woollen Cloth Maritimes	—	—	712	718	780	—	853
5b Woollen Cloth Ontario	—	—	730	690	830	—	990
6a Woolen Yarn Maritimes	744	433	415	458	563	569	686
6b Woollen Yarn Ontario	—	—	610	580	700	—	880
7a Cotton Yarn & Cloth Maritimes	887	765	725	726	885	837	1,029
7b Cotton Yarn & Cloth Quebec	691	667	757	740	850	853	832
8a Biscuits & Confectionery Maritimes	704	—	896	714	809	795	799
8b Biscuits & Confectionery Ontario	881	—	960	929	1,002	1,075	1,099

facture. There was a similar development on a much smaller proportionate scale in the Maritimes, but enough to suggest that the Maritime plants were intended to remain in operation as part of the national industry. In the other industries, up until the war, the Maritimes were losing ground in efficiency and size, and there is, from our evidence, little to justify optimism about any new growth or development of these industries. This is not to say that certain old-established individual firms may not continue indefinitely to enjoy a measure of success.

The study of wage payments, which was based on average annual earnings rather than wage rates—which are deceptive in industries in which there is much part-time employment—further supports the findings of the previous analysis.

In every industry except cotton textiles there is a regional wage differential between the Maritimes and Quebec and Ontario; in cotton textiles annual earnings are about the same in the Maritimes as in Quebec. Over the period of twenty years earnings in pulp and paper have varied in the same way and in the same proportions in both Quebec and the Maritimes, with, perhaps, a slight reduction in the regional differential. In woollen cloth the trend was also upwards both in Ontario and the Maritimes with, however, an increase in the regional differential. In cotton textiles the trend was about the same in both regions and very slightly upwards. In woollen yarns and biscuits and confectionery while Maritime earnings held about level, with cyclical fluctuations, the Ontario trends were upwards, increasing the regional differential. In iron and steel and castings and forgings the trends were downwards, with a recovery in 1940, and the regional differential unchanged. In boots and shoes while Ontario wages held constant Maritime wages declined. Once again the industry with the strong trend in the Maritimes was pulp and paper; woollen cloth, cotton textiles, the steel industries about held their own; the others lost.

This question of earnings will bear looking at another way. The degree of wage differential and the comparative depression of wages are a guide to the amount of labour cost differential necessary to offset the general location advantages. Thus in 1939 and 1940 the wage differential between Ontario (or Quebec

as the case may be) and the Maritimes was not great in pulp and paper, iron and steel, cotton textiles (in 1940 it favoured the Maritimes) and was moderate in castings and forgings and woollen cloth. In woollen yarn, biscuits and confectionery and boots and shoes it was high, indicating the general greater inferiority of these industries in the Maritime provinces. In the depression year it is again noticeable how earnings in boots and shoes and woollen yarn were particularly sharply reduced in the Maritime provinces. Even in war years boots and shoes, woollen yarn and biscuits and confectionery could not pay an annual wage of $800—the very minimum of subsistence—in the Maritime provinces.

From the point of view of the business manager the efficiency of an enterprise is determined by its earnings per unit of cost. Such figures for groups of industries are impossible to obtain in Canada in reliable form. But something approximating them may be derived from our data. The "net value added" of the *Census of Industry* returns represents the gross value of output less cost of materials and fuel. This figure, divided by labour employed, gives us our "net value productivity of labour". If we divide this by the average annual wage we have the net value product per wage dollar, or the net dollar return for every dollar of labour cost, out of which capital costs, taxes and other fixed costs and profits must be paid. In general this figure will correlate very high with profitability of enterprise, though there will be some exceptions. In Summary Table V we present the earnings' efficiency of the wage dollar for the different industries in the Maritimes and Ontario in 1940.

The reader will be struck by the very low figures for iron and steel and castings and forgings, industries in which there are high capital costs. Though we present these figures in our table we believe too much attention should not be paid to them, for the calculations of net value added in these industries are exceedingly complicated, particularly where several operations in a series of productive processes are carried out in one plant, and the reporting of "value added" in the *Census of Industry* returns seemed highly unreliable. The other

Comparison of Net Productivity of the Wage Dollar in Certain Industries in the Maritime Provinces and Ontario, 1940

Industry	*Net Value Productivity Per Annum of Labour Unit*	*Average Annual Wage*	*Value Productivity Per Wage Dollar*
	$	$	$
1a Pulp & Paper Maritimes	3000	1392	2.16
1b Pulp & Paper Quebec	3250	1564	2.08
2a Boots & Shoes Maritimes	1490 (1939)	763 (1939)	1.95 (1939)
2b Boots & Shoes Ontario	2250 (1939)	914 (1939)	2.46 (1939)
3a Primary Iron & Steel Maritimes	2200	1944	1.34
3b Primary Iron & Steel Ontario	3200	1720	1.85
4a Castings & Forgings Maritimes	1375	1180	1.17
4b Castings & Forgings Ontario	1900	1355	1.40
5a Woollen Cloth Maritimes	1600	855	1.87
5b Woollen Cloth Ontario	2150	990	2.17
6a Woollen Yarn Maritimes	1350	686	1.96
6b Woollen Yarn Ontario	1675	880	1.90
8a Biscuits & Confectionery Maritimes	1450	799	1.82
8b Biscuits & Confectionery Ontario	1800	1099	1.64

Reporting not reliable

figures, however, are sufficiently reliable and seem significant. They very strongly suggest that the wage differential between the Maritimes and Ontario, at least in the industries we have examined, is now fixed at a point where it offsets the general advantages enjoyed in the area of concentration. Whereas the productivity of labour in Ontario (or Quebec) exceeds that of Maritime labour in all industries, and sometimes by considerable amounts, the earning efficiency of the wage dollar is very nearly equal in all industries in which the figures are reliable, except boots and shoes. In pulp and paper, woollen yarn and biscuits and confectionery the Maritime industries actually have a higher earning efficiency of the wage dollar than the central Canadian industries. Thus the greater efficiency and general advantages of the industries in the area of concentration are held in a sort of balance or equilibrium by the lower wages of the Maritimes. This equilibrium does not seem to obtain in boots and shoes, where one might expect a further decline, absolute or relative, in the Maritime industry, or in pulp and paper, where one might expect a continued development, if the market permits, of the Maritime industry.

Power costs have often been regarded, particularly in industries like pulp and paper, as of decisive importance in determining location advantage. Our evidence suggests the possibility of exaggerating the importance of power as a location factor. The general inferiority of Maritime power resources is well known. The rivers in the Maritimes have an insufficient flow of water in the late summer and sometimes in winter to maintain a steady, adequate supply of power. Most of the pulp and paper mills, in consequence, have to maintain steam standby plants to supplement their hydro. Yet power costs in most industries, though they averaged somewhat higher than in Quebec or Ontario, were such that the disadvantage thus resulting was only a small proportion of the unit cost. Even in pulp and paper the power cost advantage enjoyed by Quebec was, on the average, small. Power costs per ton of output for the Quebec median establishment were, in 1940, just 5% less per ton than for the Maritime median.

Since power costs are roughly one-fifteenth of total costs, this works out to an over-all advantage for the Quebec industry of .33% of the total per ton cost. This is an unimportant fraction, but it should be admitted that there are individual mills in the Maritime provinces where the disadvantages of high power costs and unreliable hydro-electric power are of some importance. In other industries the power poverty of the Maritime provinces is probably not a decisive location factor.

4. Effects of the Second World War—General Conclusions

The stimulus given by the second World War to manufacturing in the Maritime provinces may result in permanent gains in employment and the level of activity in the pulp and paper and, possibly, the iron and steel industries. These industries, enjoying as they do particular location advantages, constitute the most important industrial occupations in New Brunswick and Nova Scotia. The expansion of the domestic market may be such that the better times now enjoyed by woollen cloth and biscuits and confectionery manufacturers may not be entirely lost after the war. We need not fear any serious decline in these industries in the Maritimes if there is any normal post-war business activity. Similarly cotton textiles in the Maritimes will not suffer from central Canadian competition, but, of course, will endure the vicissitudes of the national industry if there are large imports of cotton textiles admitted to this country as part of a post-war international trading settlement. In the other industries we have examined, the competitive Maritime position is not such as to lead us to expect them to retain any gains if, indeed, they are able to maintain their pre-war position in the face of increased pressure from their rivals in the central provinces. The history of the past twenty years would suggest that the general location factors strongly favour the central areas and, in these industries, the offsetting factors seem inadequate to give them any equality in competition even in the Maritime market.

Wartime capital formation in Maritime industries has been slight and has been chiefly in iron and steel and shipbuilding. It has been of two sorts, large additions to basic plant, and numerous small additions to small foundries and shops.[10] There has also been some expansion of plant in woodworking industries.[11] For the rest, plant expansion has been negligible. In the other industries we have studied here expansion of output has not been accompanied by plant expansion, so that there is no problem of evaluating new capital formation apart from the iron and steel and forest products industries.

We have finally to consider the wartime effects on the general location factors. Has the war upset the general location equilibrium we have been studying? That equilibrium, our evidence suggests, is maintained on the one side by the general advantages of the area of concentration, on the other side by lower wage costs in the Maritimes and specific advantages in particular industries. It is an equilibrium that has been opposed to the further development of manufacturing industry in the Maritimes with the exception of pulp and paper. Has the war changed either of the weights in the balance? It has tended to some equalization of industrial wages on a national basis, particularly in steel, and it has added much more to the machine and technical equipment of Ontario and Quebec than it has to that of the Maritimes. To some extent the industrial training has improved labour skills in the Maritimes and this may to some extent offset the reduction of the wage differential, and some of the new plant in the central provinces will be useless after the war. But, on the whole, it would appear that both sides of the balance have been disturbed so as to produce a further trend towards increased concentration in the centre. Whether such a movement will occur after the war will depend on matters of which we can now have no knowledge, such as

[10] A study of these developments suggests little hope that there is much likelihood of their being permanent. See my *Economic Effects of the War on the Maritime Provinces of Canada* (Halifax, 1944), Ch. V.

[11] The expansion of the woodworking industry has led to some standardized precision cutting and considerable technical improvements. It is possible that a furniture and interior-fittings industry could grow up and could use these small woodworking plants as feeder mills for parts. See op. cit., Ch. IV.

the state of international trading and our foreign trade balance, so that it would be foolish further to speculate about it. All we can say is that, on our analysis, the effects of the war seem unlikely to bring any permanent benefits to the majority of Maritime manufacturing industries. They have maintained their position in a precarious balance of location advantages over the past twenty years, a balance the properties of which we have observed and one that, to the extent it has been disturbed by the war, has moved in a manner probably detrimental to Maritime manufacturing.

This conclusion seems to indicate that the second World War, like the first, has, in the main, accentuated rather than reversed trends already established. Both wars had temporary stimulating effects on the Maritimes, as on the national economy. The permanent effects of the second World War have, however, no more reversed the process of industrial concentration than those of the first World War. The over-expansion of the steel industry in the first war was followed by inevitable decline, just as the shipbuilding industry of the second World War is now declining. One should hesitate to generalize from this single area, but one may properly wonder if the permanent or long-run economic effects of war are not usually of this sort, viz., to accentuate and accelerate trends already established by prevailing and fundamental economic forces.

In brief, we may conclude that the process, initiated by technological change, towards larger plant units with the decline of competition, has the further effect of geographic concentration of industry as well. We have studied one instance of this process, but that it is not peculiar to Canada is evident. The "depressed areas" of the United Kingdom, the general distress of the North with the geographic concentration of new industries in the Home Counties and Midlands area, is testimony of a similar process in that country.

This change in the location of industry, which in many countries is a consequence of the temporal economic process we have studied in this book, results, in turn, in population shifts and in serious social and political strains and problems.

It raises in a most acute and difficult form the question of whether the "natural" market mechanism of classical economic and political theory can be regarded today as an adequate protection of the public interest and the general welfare.

PART VI

AREAS OF DECISION AND POLICY

CHAPTER XIV

INTRODUCTION TO PART VI

1. A SUMMARY OF THE ARGUMENT: PRESENTATION OF THE PROBLEMS OF PART VI

In this Part we plan to discuss some of the questions of economic policy which are raised by the analysis of the previous five Parts. The relevance and appropriateness of analysis to problems of policy we regard as one test—a pragmatic one—of its validity.[1] There is a tendency today for economists to be apologetic when they find themselves confronted with the practical issues of state policy. It is well, of course, that we should be cautious and should insist on the warning that economic analysis is an inadequate and incomplete guide to policy. We should ourselves realize the limitations of our science, its degree of abstraction from the real world, the impossibility of predicting the behaviour of non-economic variables that nevertheless affect the course of economic events, the incalculable vagaries of some physical factors (like the weather) and of some psychological factors (such as generate a political upheaval), and the final limitation that economics is, in the language of the textbooks, a science of means, not of ends. In spite of all this, surely no one will claim that economists have no concern with welfare—the enhancement of welfare has been the direct and primary purpose of all the great economists from Adam Smith to Lord Keynes—or that policy decisions made in the light of economic analysis are less apt to be sound decisions than those which are inspired by the prejudices and emotional stereotypes of the economically illiterate.[2]

---

[1] "The test we must apply to economic theory, therefore, is not whether its results are contained in its premises but whether it is a serviceable instrument in the study of concrete economic problems", E. R. Walker, *From Economic Theory to Policy* (Chicago, 1943), p. 47. This is saying in slightly different form what Professor Pigou said in the first chapter of his *Economics of Welfare* (London, 1920), " . . . the type of science the economist will endeavour to develop must be one adapted to form the basis of an art".

In Part One of the present work we argued that the nature of economic causation was complex. Certain causes of a physico-institutional sort are in operation and these causes involve a kind of necessity. Given ordinary and rational human motives they cannot be removed and they entail a certain chain of effects. Nevertheless, human purpose is a part of this causal system and can enter the process in certain areas of free decision to direct the process in a meaningful way towards socially desirable ends. It is the function of politics to decide what these socially desirable ends are. It is the function of economics to define the areas of free decision, and to explain the potentially effective lines of action which can be expected to result in the attainment of these ends.

In Part Two we investigated the exogenous or external causes of economic processes and examined the causal systems of representative theories of economic change, showing how economists have selected for emphasis some "given" cause, or causes, especially prominent in the changing order of their own day. We went on to show that these causes (population change, technical innovations and geographic discovery and expansion) could be treated neither independently nor as perfectly exogenous or external to the economic order. No single-cause theory of economic change is tenable. The general theory of economic change, as we have restated it, involves the operation of these prime causes through the institutional pattern of society, the modification of the prime causes themselves by the effects they originally produced, and the human purposive response to the challenge of the process with such further modifications as this response produces by altering the institutional channel through which the effects were manifested.

In the Third Book we studied more particularly how under varying institutional circumstances selected causes would probably work themselves out.

Part Four carries this analysis forward to the study of the

---

[2] Why do the economically unsophisticated and uninstructed often gravitate to editorial positions on newspapers? Is it simply that journalism is not yet a profession, with standards of integrity, and that editorial and leader writers write the sort of economic nonsense they do write, because it pleases the owners?

characteristic process of the modern firm, and Part Five is an examination of the effects of this process on the regional location of industry in eastern Canada. It is from the analysis of these three books that we are able to define certain areas of decision.

If the population of a country is increasing the economy will be an expanding economy or not, depending on its institutional form, expanding if the market is competitive, and failing to expand if the market is not. The real welfare effects will be negative if technical advance is not proceeding at a sufficient rate to offset the operation of diminishing returns, positive if it is. A declining population, however, will result in a stagnating economy without any increase in real welfare if institutions are monopolistic and there is no technical advance. A declining population without technical advance, but with competitive institutions, will usually lead to a stagnating economy but an enhancement of welfare may possibly eventually result if the original population was pressing too hard upon the means of subsistence.

A technical advance enhances real welfare in the sense that it increases the efficiency of the economy, that is, it creates a larger output of goods for the same or less real effort—the proportion of real output to real input is greater. The welfare gains of a technical advance are not, however, evenly distributed. If the economy is competitive, the welfare gains will be more evenly distributed than if the economy is monopolistic, but one of the effects of the advance is to corrupt the competitive nature of the market. A technical advance is not, in itself, stagnating, but it is related to the ten-year cycle. This relation is twofold. Cost-reducing innovations tend to be bunched at the bottom of the cycle so that the eventual recovery requires a smaller volume of employment than would otherwise be the case. Further, a linked advance "peters" out, unless new innovations occur in a constant stream, and this process, in conjuncture with the cyclical rise in liquidity preference at the crisis period of the cycle, accentuates the crisis and consequent decline. The suffering from cyclical disequilibrium is also unevenly distributed, bearing most heavily on those very

classes of the community who gain least from the technical advance.

The characteristic process in the economy of the western world is a rate of population growth declining to the point of a stable or actually diminishing population, with technical developments which have increased the degree of monopoly and which create surpluses of a differential sort in the distribution of the final shares. The failure to distribute these surpluses aggravates the stagnation effects of the declining population and is the true cause of what has been called secular stagnation. The analysis of the single firm in process supports these conclusions, and leads, further, to the concept of senility, a condition of monopoly in which technical advance ceases.

Many of the most urgent problems of the modern world are connected with this economic development. The instability of domestic economies is the cause of intense social strife. The workers who bear the brunt of cyclical depression and secular unemployment have learned to organize so as to confront concentrated power with concentrated power. Group interests, in the intensity of social strife, are given precedence over national interests. The basis of parliamentary democracy, which is the willingness to subordinate all ends to the end of community interest as determined at the polls, is destroyed. The small producer and the unorganized worker, chief sufferers in the economic process, are ready victims of the fascist type of appeal. Only a strong, united, politically sophisticated and self-disciplined people can keep parliamentary institutions working satisfactorily. Among such people, the essential agreement to settle issues through the established institution of parliament might exist, and some hope consequently might be maintained that discussion and experiment would lead to the working out of sound policy in the national interest. In most countries, however, irreconcilable groups struggle to dominate, and in the bitterness of battle do not hesitate to put group interest above the law and the constitution, or to believe that the achievement of group objectives justifies any means. A successful fascism requires (a) consolidation of economic power in the hands of a small group, and (b) the development of a dis-

contented and embittered mass that can be won to support the leader group. The economic process creates both these conditions.

In the international sphere, the uneven distribution of wealth and welfare and the instability of the western economies is one of the conditions breeding conflict and war. We do not assert that it is the sole cause of war. It is certain, however, that the suspicion that makes for "power politics" as popularly we call the constant jockeying for power among the major states, and which, in turn, leads ultimately to war, is stimulated and even engendered by economic rivalries and jealousies. This general condition is gravely aggravated if and when the concentration of economic power in any major state is followed by the possession of political power. Such a despotism, unstable in essence, looks for the consolidation of its power abroad and frequently pursues policies destined to increase international suspicion and distrust. The concentration of power is an evil thing, and the process of the western economy invariably leads to such concentration.

Now it may be argued that the "realistic" view would be to accept the fact of this concentration of economic power and to adapt our political institutions to it, admitting that parliamentary democracy was the political form suitable to the competitive stage of capitalism but asserting that the communist or fascist state is appropriate to the present stage. It is on this point, I believe, that communists and radical liberals or liberal socialists part company. Both are agreed that the conditions of modern economic society are socially intolerable and that radical economic change is essential if society is to progress or even to escape annihilation and destruction in atomic warfare. Both reject the solution that the small group of capitalists wielding a great concentration of economic power should have their position consolidated by the attainment of unqualified political power. The communist believes that the only solution is the creation, first, of a class-conscious labour organization capable of wielding economic power comparable to that of the capitalists and later the dispossession of the capitalist class and the organization of a state power in the hands of the leader group

of the working class, the "dictatorship of the proletariat". Eventually, of course, the communist professes to hope, this power will voluntarily be abdicated and a kind of stateless society will eventuate. This unhistorical and naïve belief, however, we may dismiss. It is unlikely that even communists sincerely entertain it.

Now the modern liberal, who may not differ greatly from the communist about the economic reforms which are necessary, regards the *political* views of the communist as straight reaction, as, indeed, in any strict sense of the word, they are. Since the only progress of which the race is capable is a progress in individual responsibility and judgment, the development of an improving system of values, any political system that limits or reduces this responsibility and which forbids the full participation of the individual in the process of social judgment is a retrogressive or reactionary step. Even if we make mistakes—and we must make them to learn—it is better to make them and learn from them than not to participate in the community of judgment. It is probably the result of the accident of history that gave us our first successful socialist revolution in a politically immature state, which has led to this association of advanced economic policy with a primitive political organization, an association in the necessity of which communists believe with the fervour and fanaticism of an anchorite.

The liberal, however, who believes in the slow advancement of man through the exercise of individual judgment and in the political institutions that make this exercise possible, must show that it is practicable to devise economic solutions of present issues and, by free choice, to guide the process of economic change towards socially desirable ends.

These socially desirable ends require definition, but definition can come only from the people themselves. I suppose, however, that no one would dispute the following propositions:

(a) That it is better, other things equal, to have an expanding than a stagnating economy;

(b) That it is better to have an increasing than a diminishing total real income;

(c) That it is better to have a more than a less equitable

distribution of this income, as long as the reform in distribution does not result in a diminution in the size of the possible income;

(d) That the concentration of economic power is undesirable if the same level of welfare can be attained without it;

(e) That any reduction in the amplitude of the cycle is desirable, if it can be accomplished without a sacrifice in real welfare and the rate of technical progress.

Some of these objectives are not strictly compatible under all circumstances, and it is when one has to be weighed against another that disagreement may occur. At this point we wish to make the assumption that these disagreements are not, in practice, serious and that, for all practical purposes, we can regard these objectives as encompassing our social goals.[3] As the argument develops we shall try to justify this assumption by showing that, in many instances, the method of achieving one of them leads, not to the sacrifice, but to the attainment of the others.

We may observe, to begin with, that changes in population are not wholly exogenous or external to the economic order, but are changes, as we have shown reason to believe, that are in turn subject to modification, even to reversal of sign, when acted upon by the very effects they produce. This indicates that over a sufficient period of time population trends may be changed if policy is successful in affecting the necessary economic conditions. Thus the present population trend in the West may be reversed if security and regularity of employment and improved living and health conditions can be achieved. However this may be, population is subject in another way to direct political control. The world population pattern at present is constituted of countries which suffer stagnation effects from the failure of population increase to maintain itself and other countries where welfare is below the world average because total population (and the rate of increase) press too

[3] I realize that I treat cavalierly a problem to which careful attention is usually paid. The problems of welfare economics are not as simple as I have represented them. It is partly because they have been so thoroughly discussed in the literature that I avoid here further discussion which I cannot believe would add in any way to what has been written.

severely on the means of subsistence. It is clear that world welfare would be enhanced if population were transferred, for example, from India to the United States, and not only would total welfare be increased, but it would be increased not "on balance" but as a sum of the welfare increases in both India and the United States. Of course, for this policy to be practicable, people in the United States and in the younger members of the British Commonwealth would have to overcome deeplying racial prejudices. The stereotypes of "white Australia", "a white man's country", and "no orientals" would have to be abandoned in Australia, the United States, and Canada. It would be unrealistic to suppose that this kind of prejudice can be rapidly conquered. Americans and Canadians at the moment have an amazing facility for perceiving the wickedness of British policy in Palestine, the problems of which they fail fully to comprehend, at the same time remaining impervious to the hard cruelty of their own immigration policies as applied to the starving, "displaced persons" of Europe.

Thus the transfer of population as a deliberate act of policy is something that people are free to decide to do, in the sense that it is economically feasible, as an immediate change of the rate of population increase is not. People are not morally free to do this, however, because they have not overcome irrational prejudices that prevent them. This is a moral failure, a lack of freedom attributable to incomplete rationality, not a restriction on freedom from the operation of inevitable and indefeasible economic circumstance. We may here observe, too, that successful population transfer requires other achievements in the way of economic reform. There is no purpose in moving Chinese to Australia or Indians to Canada if they are to remain unemployed, or to find employment at starvation wages replacing native Australian or Canadian workers. For the expansive stimulus of population increase to have effect, the ten-year cycle must be brought under some control and institutional restrictive forces, such as monopoly, eliminated.

When we come to consider population transfers between regions within one national state, as we shall do shortly, we shall see that there are always two aspects of this problem.

Population can be transferred to the other resources, or the other resources can, in part, at any rate, be transferred to the areas of present population. This is also true in the international field. If, for political and moral reasons, population transfers are today not immediately practicable on any adequate scale, there is no reason why certain capital resources should not be transferred to the over-populated areas, thus enhancing their productive power, increasing their welfare, and, at the same time, acting as an expansive stimulus to the economies of the capital-exporting countries. This is clearly an area of free decision and the potentialities of this programme we must examine in some detail.

When we turn to the question of technical advance, we see that this process is directly related to the concentration of economic power, the creation of undistributed and unearned surpluses, the existence of excess capacity, misallocation of resources and social waste, and indirectly to cyclical unemployment and secular underemployment. Given the ordinary springs of human behaviour and the inter-related nature of scientific discovery, we should scarcely expect or desire to be able to exercise complete control over the nature of invention or the rate of innovation. We may observe, however, that even this is not entirely beyond the influence of purposive control. Science is harnessed to specific social circumstances and the institutional nature of society profoundly affects scientific discovery. When, as in Canada before this last war, private industry is able to offer a million dollars for research into such problems as that of recovery from waste liquors in the manufacture of chemical pulp, as against every thousand dollars that public enterprise and private philanthropy make available for basic research in the physical and biological sciences, it is fairly likely that many of the more productive minds will be drawn into the service of limited commercial purposes rather than into the service of either pure science or of those applied branches of science that serve the needs of humanity as a whole. Even in the medical profession this influence is clearly exhibited. In America, whole sections of the country containing large populations are almost without medical services,

while in the cities the doctors compete with one another like horse-traders to build up lucrative practices. And the American Medical Association, discarding in this regard the proud traditions of the learned professions, maintains a propaganda organization for "private enterprise" in medicine which apes the lobbying techniques of big business. Certainly, in many of the applied branches of science, purposive social direction could influence the nature of scientific development and the social uses to which science is put.

For the most part, however, we should accept as desirable technical improvements in the productive process and strive, rather, to control their social effects. This possibility must be realized through (a) the control of monopolies and cartels, (b) the control of accumulated surpluses, (c) the control of the cycle, and (d) control of the location of industry.[4] It will be our task in this Part to define the areas of decision in each of these cases and to indicate the lines along which economic policy might appropriately be developed.

2. A Digression on Some Matters of Politics

In the last Section in a reference to the primitive political theories and practices of the communists, I adverted to the dogmatic fanaticism of these modern apostles of the new revelation. But this is a world of dogmas. Equally resistant to the appeal of rationality are the devotees of so-called "private enterprise", those who reject any proposal of social or economic reform which entails any degree of purposive political control over business enterprise regardless of its merit, on the grounds that it is "socialistic" and an "interference with private enterprise". If the communists resemble the early Christian

[4] The chapters which follow in this Part do not occur in the same order as that followed in the text at this point. The problems emerge in this sequence, so we have thus listed them. When it came to their arrangement for an orderly discussion of policy they had to be differently grouped. The location of industry is so related to problems of population that we have discussed it along with the question of the transfer of population and the provision of investment opportunities in Ch. XV. Cycle control as a special problem of the comparatively short run is treated in Ch. XVI, as are purely fiscal remedies of accumulated supluses. The problem of monopolies and cartels is discussed in Ch. XVII along with certain aspects of the problem of surpluses.

fanatics who destroyed the library of Alexandria because it was pagan, these are the modern idolators, clinging to the outward forms of an out-worn faith. And, alas, of their idol not the feet only are of basest clay. Private enterprise, as the neophytes sing it, is a creature of harmony and nice proportions, so maintained by initiative and competition. In fact, closure of entry, monopoly, trusts, cartels and tariffs have so engrossed and distorted him that his ancient visage now betrays none of the fine lineaments of his youth.

Adam Smith, himself, and all respectable economists since, have been concerned not with dogmas, but with welfare. The defence of the competitive private enterprise system was that it made for an optimum allocation and use of resources. The decline of competition, the appearance of the senile firm, and the growth of inertia and stagnation on the one hand and of cyclical instability on the other, prevent today the attainment of any approximation to the level of welfare technically possible. The economic system defended today as "free enterprise" is an inefficient one. Our analysis has shown that forces within itself must make it, if left free of purposive direction, increasingly inefficient and unstable. It is dangerous, too, to social stability and political progress. It develops intense social conflict, divides the body politic and renders difficult if not impossible any rational formulation of a concept of national or community interest. It sets nation against nation, class against class and community against community. It leads to great accumulations of wealth in the midst of most dire poverty and makes, too frequently, a continuation of that poverty, of unemployment and human wastage, a condition of the continued accumulation and tenure of great wealth. The accumulation of wealth and the concentration of ownership is the basis of great power, wielded often with disregard of any sense of community interest and to the detriment and even danger of the State. The ethics of modern business enterprise is an ignoble one. Anything goes. People may be bullied, browbeaten, frightened, cajoled into buying things they neither need nor want. Public taste may be degraded, public morality corrupted, to serve the ends of profit and power. The acquisitive instinct is nurtured to

the exclusion of all others. The value system of a whole society, its entire culture, reflects the baseness of taste, the ignobility and venality of those men who set its tone and whose own standards are fixed and qualities determined by the necessities of the commercial world.

No one can seriously maintain that this halting and inadequate organization of our economy, with its topsy-turvy scheme of values, its base, ignorant and uncultivated aristocracy, and its growing menace of concentrated and irresponsible power is an end in itself, a social good in its own right to be retained at whatever cost to social security and human welfare. Institutions are never good or evil in themselves, absolutely; they are meant to serve human purposes and are appropriate or inappropriate to those purposes. The purpose that economic institutions are intended to serve is human welfare, the provision of an adequate material basis for the good life. Our present economic institutions are inappropriate. They are inadequate to serve the purpose they are intended to serve, and they are inappropriate to that purpose in the sense that they actually corrupt our social formulation of it. The grave crisis of our race is a challenge to its adaptive genius. Grave crises demand heroic action. But the measure of boldness in social policy is not the degree of novelty: it is the acceptance of intelligence as the guide of action. Intelligence in social judgment comprises both the technical knowledge of the economist in the weighing of the consequences of different courses of action and the wisdom of the historian in relating new institutional developments to the accustomed way of life. Changes that are successful are those grafted on the main trunk of social life and experience. Revolutions change a single aspect of social life, a single institution is overthrown to make way for another. A king loses his head and a "Commonwealth" succeeds to sovereignty; British governors are sent packing and a federal congress succeeds to sovereignty over the states; an aristocracy has its powers and privileges taken over by the merchant class, or by the organized hierarchy of the embattled proletariat. But the Commonwealth adapts itself to the traditional constitutional usages of the English state, and reflects the economic and religious

pressures of seventeenth-century England. The states are the old English colonies in custom, law, social organization and commercial enterprise. The culture of the bourgeoise salons of the Consulate is in the direct line of the culture of the *ancien regime*. The dictatorship of the proletariat remains a police state essential to the governance of a far-flung, populous and primitive domain. The changes may eventually be felt throughout a nation's whole life, they are not lightly to be dismissed when the institution overflung by revolution is so basic as that of sovereignty or social hierarchy, but the changes are never clear-hewn, abrupt, complete, a "tearing-down" of an entire civilization and a "making of a clean start". History is a continuous process. The forces and ideas that made the revolution have their origin in the structure that is destroyed, are a fundamental part of what is created in its place. We of the West who inherit the revolutionary tradition of liberal parliamentarianism have in our liberalism the source and impetus necessary for economic reform. Our economic system has grown under the pressure of technical change to the place where already its typical units are national, continental, even world-wide in scale. The adaptation of the parliamentary system to harness and control in the public interest these colossal agglomerations of power is perfectly possible and completely in harmony with the essential principles of the liberal faith. Already the most embracive and essential of these undertakings, communications, transportation, power supply, etc., have in country after country been brought under public control or ownership. Different devices have been experimented with, with varying success. Experience in responsible control and administration is being acquired. A gradual extension and amplification of the control system, according to administrative principles no longer strange to parliamentary systems, is all that is required. This means neither communism nor socialism of a doctrinaire sort, but the development of a "mixed" economy within the framework of parliamentary democracy.

Recently we have heard much to the effect that social control and ownership within the economy must corrupt democ-

racy, because of the great bureaucracy that would be created and because of the pervasive economic power that would accrue to the state.[5] This argument is falsely premised. It sets up the alternatives of a freely competitive economy existing in the framework of a *laissez-faire* liberal state and a concentrated, socially-owned economy—"state capitalism"—with a vast, all-powerful bureaucracy which would possess the reality of power of which only a feeble semblance would remain in the hands of parliament. These are not the alternatives. The true alternatives from which we have to choose are (a) a continuation of the present system of great corporations, "privately owned", i.e., owned by stockholders of whom a minority control policy, possessing great and erosive power within the state, with a large state bureaucracy such as we now have, struggling to exercise some minimum of control over these corporate monsters by means of futile anti-trust legislation and by supervising their battles with the almost equally powerful unions of organized labour, or (b) a state armed with more effective legal powers and with the more monopolistic economic units owned by the state and directed to the service of the public interest. Either way the problem of concentrated economic power has to be dealt with; either way the problem of bureaucracy remains.

The problems of concentrated economic power and of bureaucracy ought never to be envisaged as the problems of the socialist state or of the state in which there is, to be accurate, more public ownership rather than less. These problems are endemic in the vast and complicated apparatus of the modern state. I am not at all sure that they are more difficult under complete socialism. They may be, though that remains to be shown. But it is frivolous to suggest that they are peculiar to public ownership, and that if we avoid any increase in the public sector of the economy we shall thereby save our democratic soul. This is nonsense, and dangerous nonsense. These problems have to be faced and solved in any event. It is possible that they will be made no easier if public ownership is

[5] This is the view expounded by Professor F. A. Hayek in his *Road to Serfdom* (London, 1944).

extended. As to that, there is no evidence on which to found judgment. But if the lines of solution can be worked out they can be extended with any extension of the public sector of the economy. These are problems for the politist, not the economist. Already some indications are offered,[6] the development of the committee system, further economies in the use of parliamentary time, the delimitation of administrative responsibility, the development of technical advisors to the Opposition, and so forth. Wartime experience in England and the Commonwealth nations has done much both to impress on parliamentarians the dangers of administrative rule and to develop parliamentary techniques for maintaining the chain of responsibility and of acquiring parliamentary review of administrative acts.

However the solutions may come—and they will be easier in countries with a parliamentary system than in those with a congressional-executive system—we may be reasonably confident that they can be worked out. If they are not, democracy is damned in any case, whether or not the public sector of the economy is increased. If they are, our democratic society and the good and liberal elements in our civilization have a greater chance of survival if the process of economic change and development is brought under responsible social guidance than if it is not.

[6] Cf., for example, Professor Brady's suggestions in Ch. II of Brady and Scott, *Canada After the War* (Toronto, 1943).

CHAPTER XV
POPULATION AND INVESTMENT

1. POPULATION TRANSFER AND REGIONAL DEVELOPMENT WITHIN A NATIONAL ECONOMY

In many countries of the west today industry has been concentrated, new industries have grown up while old industries languished, and this economic process has tended to leave some areas or regions within the country languishing for lack of healthy economic nourishment while other areas have experienced all the growing pains of rapid development. In Parts IV and V of this work we have seen why this process of concentration occurs. We have now to crystallize the policy issues posed by it.

The problem is a common one. In England the concentration of the iron and steel trades, of engineering and shipbuilding, the loss of overseas textile markets, combined to create Jarrows and depressed areas throughout the old industrial north, while the new industries, motor-engineering, automobiles, aircraft, aluminum, electrical consumers' goods, rayons, food canning and motion pictures, all catering primarily to the home market and protected by tariffs, tended to concentrate in the home counties, around London, and in the south Midlands.[1] While towns like Bolton, Jarrow and Newcastle languished, Coventry, Oxford and the industrial suburbs of London grew like the "boom" towns of the American west in their flourishing youth. These new industries posed further problems of policy in that, while they demanded raw material imports, they failed to provide, as the old industries had done, foreign exchange from exports to provide for the service of Britain's heavy debit balance on current account.[2] They were luxury trades in

[1] See A. Plummer, *New British Industries in the Twentieth Century* (London, 1937); G. C. Allen, *The Motor Industry in Great Britain*, London and Cambridge Economic Survey (July, 1926).

[2] See A. T. K. Grand, *A Study of the Capital Market in Post-War Britain* (London, 1937).

two senses, both in the sense of catering to non-essential needs, and in the sense that only a people with large accumulations of foreign investments on which to draw could afford to import the raw material for these non-exporting industries. Today, with these foreign investments dissipated by war, Britain is making a deliberate effort to rationalize and revive her old export trades in order to provide an adequate exchange balance to sustain the level of imports necessary for the maintenance of her standard of life.

In the United States similar, if less acute, problems have been posed. The concentration of the new textile industry in the South has left derelict towns in New England. The new kraft paper industry, again in the South, has led to the decline of the northern sulphate industry, and, with the technical developments in bleached kraft, we may expect the northern sulphite industry also to decline. The motor car and machine tool industry has been centred in the Detroit area. Though there has been some decentralization of this industry, especially by General Motors, and some "diffusion" of light manufacturing generally, the newer industries like the old tend to be highly centralized and population has been left to gravitate to the new centres of employment.[3]

In Canada these problems are particularly acute. As we have already seen (Part V), technical advances destroyed the "wood and wind" mercantile economy of the Maritime provinces and geographic expansion later did similar disservice to their agriculture. Their smaller manufactories declined under the pressure of the process of industrial concentration. This present war has accentuated their comparative industrial disadvantage and the phenomenal capital expansion of Canadian industry has been almost entirely concentrated in Quebec and Ontario, and, to a lesser extent, in British Columbia.[4]

[3] See D. B. Creamer, *Is Industry Decentralizing?* (Philadelphia, 1935).

[4] See my *War Time Contracts Let in the Maritime Provinces* (Halifax, 1942); and Chs. II and IX of my *Economic Effects of the War on the Maritime Provinces of Canada* (Halifax, 1944). These studies show that, though there was considerable public spending in the Maritime Provinces during the War, little of it was for industrial expansion, and what little was for industrial expansion was for equipment of highly problematic value for peacetime uses. Subsequent experience has confirmed the

The prairie economy, similarly unindustrialized, has suffered from all the vicissitudes of a single-crop agricultural economy dependent on the world market, from soil erosion, and the failure to develop alternative occupations.

Since 1891 the Maritimes have experienced a steady emigration which became most serious in the three decades 1901-1931. This drainage actually exceeded natural increase in the province of Prince Edward Island, and was so great as to keep the population of Nova Scotia almost stationary in the period 1911-1921, and to hold its rate of growth far below the Canadian average. The fact that New Brunswick did experience some increase during this period is to be attributed to its higher birth rate, not to an unwillingness of its young people to move away from a stagnating to an expanding economy.[5] The prairies, which had seen the most rapid growth in population of all Canada, 1901-1931, in the following intercensus period showed stationary, and, in some cases even declining populations.[6]

In the past, we have always allowed population to gravitate to expanding industrial areas under the natural pressure of economic circumstance. It has been part of the ordinary operation of the economic system that labour should move to those occupations that offered the best rewards, and, if those most remunerative occupations should be in distant places, to those areas where the most fruitful natural resources were to be found. Sometimes the response to the economic stimulus was easy to make. Young people, from an overcrowded country in which opportunities were restricted, found it comparatively easy "to seek fame and fortune" in romantic new lands. It was especially easy when difficult cultural ruptures were not necessitated as in the case of Scottish immigration to Canada, or when political tyranny or the frustration of a stultifying caste system made exile easy and welcome. This was the case with

doubts I then entertained as to the peacetime value of the wartime additions to industrial plants in the Maritimes.

[5] W. B. Hurd and J. C. Cameron, "Population Movements in Canada, 1921-1931: Some Further Considerations", *Canadian Journal of Economics and Political Science*, Vol. I, p. 222.

[6] *Census of Canada*, 1941 (King's Printer, Ottawa).

much of the nineteenth-century migration to the United States, Australia and Canada.

All migrations of people, in response to economic stimuli, are not, however, so easy, or so happy in their results. People become attached to their communities, their culture, political institutions and way of life. The continuous drain of population, especially when the migrants comprise the most vigorous and aggressive youth of a people, can have a most serious effect on the communities of emigration. It is possible that early migration from the Maritime provinces, for example, relieved the pressure on a fully manned fishing industry and on submarginal farming during the first decade or so of the movement. The continuous drain over two générations has, however, deprived these provinces of a sufficient leaven of progressive and vigorous citizenry and has left many municipalities with heavy fixed costs for schools, hospitals and other community services and a large proportion of aged people as a burden on the reduced population of productive age and vigour. Very often, too, the passage of the immigrant is a hard one, his reception cold, his new life and his assimilation to the new community cruel and difficult. Not many immigrant boys became Marconis or Carnegies. Most of them remained poor and despised and exploited, "wops", "hunkies", "kikes", "limeys" till the day of their death.

Migration within countries is also often hard. The "natural" movement of population in Canada, in response to economic stimuli, has occurred only as life in communities of the poorer regions became intolerable, and it has had most deleterious results on the deserted communities, leaving, in the words of the Prime Minister of Nova Scotia, "towns which with their deserted streets and unpainted houses, look as though they had sunk, so to speak, below the margin of a decent subsistence".[7] And if the towns are derelict, the conditions of the countryside are often worse. "Here are found the most indescribable conditions of life. Crowded, stuffy, ill-made, unventilated

[7] The Hon. Angus L. Macdonald in his Foreword to N. McL. Rogers, *A Submission on Dominion-Provincial Relations and the Fiscal Disabilities of Nova Scotia with the Canadian Federation* (Halifax, 1934).

two-and three-room houses, draughty in winter, swarming with flies in summer, dirty, unpainted and ill-kept, are only too often what are thought of as home by children to whom the schoolhouse, a barren one-room shack, is a kind of prison that offers no escape or hope of escape. There is no beauty or promise in these lives; bred on a hard and difficult soil, neglect has left them to endure without hope of alleviation a bare and ugly existence lacking not only the minor amenities, but even the simple delights possible to a contented people not ground down by poverty and the struggle to survive."[8]

It does not seem to me that in a rich country like Canada we can or should permit again the growth of such dire poverty in the poorer regions as, in the past, has driven people to migrate, nor ought we to allow to continue the wretchedness in which some communities exist as a result of emigration of the most vigorous and youthful of the population. Population transfers must, no doubt, continue from the Maritimes and, perhaps to some extent, from the prairies, and people must, of course, move freely about the country. But something could well be done to ease the way of the migrant, to help him to find the new employment he seeks, to advise him of job opportunities, to help him find housing, move his family, and, most important, to see that a minimum of social and community services is maintained in the communities left behind.

Thus, though a continued mobility of labour in response to economic stimulus is desirable, I suspect we have reached the point where we have seriously to consider the condition and welfare of the less prosperous regions. A virtue of the old system, with all its toughness, was that it did bring the most efficient and vigorous labour to work the most productive and well-located natural resources. We should watch out that a state paternalism, which would subsidize sub-marginal industries working low-grade resources, and do so, perhaps on a basis of political horse-trading, region against region, should not be substituted. There is danger of that in federal countries like Canada, Australia and the United States, and some danger

[8] B. S. Keirstead, op. cit., pp. 212-13.

of it in industrially regionalized countries like Great Britain. The location of industry, however, as capitalism has become more monopolistic, has not always followed the natural law of comparative advantage. Concentration of ownership has led to geographic concentration, and has done so even when it resulted in the under-employment or complete waste of intra-marginal resources and even in the uneconomic use of resources.[9] There is convincing evidence that the ownership concentration of Canadian industry, for example, has left unexplored and undeveloped important resources in the Maritime provinces that might with economic wisdom have been exploited to the advantage of the people of those provinces.[10] Studies of industrial stagnation in English depressed areas and in the United States show similar results.[11]

There is therefore a case here for developmental spending by public authority. We say public authority because it is fairly obvious that it would not be undertaken by private venture. This is for two reasons. In the first place, private capital is already committed to the programme of concentration and is not interested in the welfare effects of this programme on the regions which suffer most from it. In the second place, what is required is not the direct establishment of new industries or the rehabilitation of old ones no longer economic, but the development and improvement of resources which would support economic undertakings whether these were of a public or private nature. Indeed, to avoid the danger of the uneconomic use of resources and their misallocation as a result of

[9] One illustration of this, selected at random from among many, is the operation of the nitrogen cartel. This cartel cut off the Canadian sulphate of ammonia industry, on the West Coast, from its natural markets in the United States and Pacific Islands, while, at the same time it prevented East Coast Canadians buying from their natural source of supply in Norway. "The control exercised by the cartel thus deprived Maritime farmers of the freight advantages which their location near tidewater ports had made possible." "Canada and International Cartels", *The Reports of the Commissioner, Combines Investigation Act* (Ottawa, 1945), p. 6.

[10] See my *Economic Effects of the War on the Maritime Provinces*, especially Ch. IX.

[11] See Ellen Wilkinson, *Jarrow* (London, 1935); V. W. Bladen, "The Economics of Federalism", *Canadian Journal of Economics and Political Science*, Vol. I, p. 398.

action purely political in inspiration, it is probably desirable that new industry should to a large extent be privately initiated and financed. In the Maritime provinces, for example, the forest resource, if properly managed and economically exploited, could sustain a far larger forest-using industry than it now does, and the Maritime forest-using industries enjoy certain location advantages in world competition.[12] To enable this natural resource to be profitably exploited by an expanding industry, however, a forest road programme, the development of new sources of electrical power, and the protection of waterways are needed along with the legal enforcement of more scientific cutting policies. These are clearly matters of state action. In some cases, as for example in the British coal industry, more drastic state action is required. There rationalization and mechanization require integration of ownership and policy, the provision of alternative employment necessitates a state plan of industrial diversification in the coal-producing regions, and the development of rival fuels and sources of power makes desirable large-scale research into the uses and chemical properties of coal. Very much the same can be said of coal mining in Cape Breton.[13]

Thus our problem is that the concentration process leads to depressed areas and regions. We may either decide to transfer labour from these areas or to bring capital into them for the development of their more economic resources, or we may do something of both. If we decide to allow population to move, we should ease and assist its movement, and do something to prevent the migration from leaving derelict and sub-marginal communities in its wake. This something should take the form of an economic development of the resources of the areas of emigration. We stress the qualifying adjective "economic", because a temptation to be resisted is the subsidization of uneconomic enterprises with a consequent reduction in the size of the national real income from resource misallocation. Modern science and technology, however, have put in our hands powerful instruments for controlling industrial location. Resources

[12] See my *Economic Effects of the War*, Ch. IV.

[13] Ibid., Ch. V.

can be improved, new uses for existing resources can be developed, new resources discovered. It is within the competence of the modern state, if its people so decide, to end forever the complete decay of derelict areas, that in the past have existed to the detriment of the whole body politic and to the ruin and suffering of the unfortunates who continue to dwell in them.

2. Economic Change and the International Order[14]

When we turn from the issue of population and industrial location within a single country to the same question in the world as a whole, our problem, though much vaster and more complicated politically, is, in one respect, simplified. International and inter-continental migrations of peoples on any scale adequate to the problem, or any scale comparable even to the great migrations of the past, simply does not appear as a practicable policy within the foreseeable future. Never were all the forces and motives which drive or lead people to leave their native lands in search of better conditions of life more powerfully operative than now on the hearts and minds of millions in Europe and Asia. But the countries to which they would wish to come will not receive them. This attitude is not to be explained in purely economic terms, though it is true that the fear of trades unions of low wage competition, and the general fear of cyclical unemployment both play a part in fixing it. When one considers, however, the stagnating effects of current population trends on the economies of such countries as the United States and Canada, one can see these fears to be unfounded, or, rather, one sees that the long-run stimulating effects of rapid population increase would offset and more than offset temporary dislocations of the wage structure in particular labour markets. (This proposition is postulated on the maintenance of the rate of technical advance. Compare Chapters VII and IX.)

There are, however, other and non-economic motives at work, racial prejudice, fear of politically undesirable groups,

[14] I am indebted to the editor of *The International Journal* for permission to use in this section material from an article of mine called "Economic Change and the International Order", published in the July, 1946, issue of that Journal.

the national hostility of a post-war world, and these are none the less powerful because irrational. A great French historian has said that moral facts are as much facts as those that are physical and tangible. We might add that "immoral" facts are facts, too, and that this consideration should command the careful attention of those good people who wish to build the world over to conform to their own particular vision of Utopia.

The alternative to population movement, the mobility of labour resources, is the transfer, in the opposite direction, of the capital resource. This we wish to consider not in terms of philanthropy but of soundly devised economic policy. Our problem is constituted on the one hand of the serious under-capitalization and over-population of great sectors of the world, with consequent dire poverty, and, on the other hand of mature and stagnating economies with stable or near stable or declining populations and with a lack of suitable investment opportunities. We seek a way of improving welfare in all the countries, by raising the productive capacity of some and by stimulating the economies of others. There is an area of free decision here, in that there are controllable factors, as we shall see, which, given the population structure, nevertheless could be so directed as to lead to the desired end. There are complications arising from the political relations of the great powers, but, as we shall attempt to show, it is quite possible that the economic policies, directed primarily to the end of material welfare, would make no negligible contribution to the solution as well of the purely political problems as of the economic ones.

Political pessimism today derives from the apparent fact that the great continental nation states, the United States and the Soviet Union, are not prepared to sacrifice to an international body any particle of national sovereignty. The smaller and undefendable powers, not by reason of moral superiority, but because their frailty in an atomic age leaves them no alternative of policy, are apparently willing to create an international sovereign with a monopoly of armed force. Jurists and political scientists are agreed that this is the only permanent adequate solution for international anarchy and war. Hobbes was right,

but there is no single superior force capable of establishing the world Leviathan. Instead, the two great continental powers, comparatively flexible in resistance to atomic bomb attack, continue a balance of power policy based on the concept of indefeasible and inalienable national sovereignty. Such a situation is only metastable and it is a matter of time before war comes. However, there is the matter of time. Realistic thinkers must inquire if anything can be done to break down suspicion, fear, mistrust, and intense nationalism during the present period of a precarious power balance.

A factor in the fears we entertain of the future is the likelihood of a major economic depression beginning in the mature capitalist economy of the United States and spreading throughout her economic orbit. This depression would have unpredictable consequences, the most dangerous of which would be that a reactionary American government might attempt to distract its people from their miseries by an aggressive foreign policy, and to stem growing unemployment by a gigantic rearmament programme. The elements of so-called "crypto-fascism" are already pronounced in the United States. Russian aggressiveness, which grows from an almost hysterical fear of the "outside" world, would know only one way to meet an ideologically hostile and equally aggressive America. The precarious balance would be overset: metastability, as always, would show itself truly as a mere illusion of stability.

The fears of a depression are not ill-founded. Not only is the United States a mature economy, in Professor Hansen's sense, suffering from stagnating influences, which include a high degree of monopoly in her industrial structure, all of which, as we have seen, lead through the aggregative mode to a decline in economic activity, but also she seems, at the moment of writing, committed to internal fiscal and monetary policies bound to lead through inflation to crisis and cyclical depression. The conjuncture of cyclical and secular decline, if uncontrolled, may well lead to the most serious depression in modern industrial history.

There is, however, no inevitability about secular stagnation as such. Professor Hansen, himself, believes that a sufficient

volume of public investment, properly timed, could maintain the national income and the level of employment more or less indefinitely. As we shall show when we come to discuss the cycle, we are not convinced that this in itself is an adequate solution. It is a kind of temporizing with the outward manifestations—as they appear in aggregate form in the constituents of the national income—while the real causes of stagnation, as Professor Hansen's argument admits, and as in this work we have tried to show, lie much deeper in the economic institutions of our society. Moreover, fiscal reform, at best, is most effective in a self-contained economy and has no significance for the still vast, under-developed, economically immature regions of the world. It is true that some discussion at the time of the Bretton Woods meetings involved the notion of international spending, for developmental purposes, and this is consistent with the general Keynes-Hansen thesis. This, however, would have taken the form of individual national borrowing from the international fund, and, for reasons we shall hope to make clear, would be inadequate economically and open to political objections.

The real causes of stagnation in a mature economy are, to review Professor Hansen's thesis, the failure of population to expand, the failure of new resources and new territories to be developed, and the failure of new innovations to demand large-scale new investment. The failure of population to expand is stagnating in its effects because it means that consumer demand is constant and entrepreneurial expectations are consequently modest and limited. There are great areas of the world where population is still expanding, where potential consumers' demand is great, where all that checks it from becoming effective demand is lack of purchasing power. This purchasing power could be created by the introduction of western innovations and the capitalization of these economies with the effect of increasing their per capita productivity. This process, at one and the same time, would mean the wide adoption of investment-creating innovations, the opening of new resources and ter-

ritories, and the creation of excellent investment opportunities.[15] However, before we feel that here indeed is the programme for North American capitalism we must pause to consider two things. The first is economic. The failure of the rate of innovation in the mature economy to be investment-creating comes from the fact that the real incentive in economic development is the existence of windfall profits or surpluses. In the purely competitive economy surpluses tended to be temporary. The rewards paid to the various agents of production exhausted the total value product, and the temporary special gains which new techniques or improvements created, were erased by the abrasive of competition. In our trustified economy, these surplus gains are accepted and necessary. The possibilities of investment in "backward" countries are only "good", from the private investor's point of view, if in the short run they promise high profits commensurate with the risk, and if the process does not reduce the total entrepreneurial surplus by establishing a competitive industry which would reduce the rate of profit more than it would expand sales. Hence the small use that has been made of the possibilities of Africa and Asia, and hence the "exploitation" of backward countries when western investment has occurred. The "exploitation" is a necessary safeguard for profits commensurate with risk. Clearly it requires someone who can afford to forego personal profits and who stands to gain from the general improvement in employment and real welfare to undertake successfully the development of the African and Asiatic economic frontier.

The second observation is that no such investment by either British or American capital or both together could be undertaken in many of the strategic areas without at once precipitating great trouble with the Soviet Union and, considering past records, a good deal of resistance from native populations. The development of an immature area by a mature metropolitan economy always leads to the "possession" of the re-

[15] "The process of raising the national incomes of relatively backward countries through development works will provide new or enlarged markets for the more mature economies." B. H. Higgins, *Public Investment and Full Employment*, International Labour Office (Montreal, 1946), p. 3.

sources and properties so developed by the metropolitan power and, if they are at all strategic, to the extension of the metropolitan's strategic interests and position. If we consider North Africa, the Levant and the oil fields of Irak and Iran we can see, at once, the real meaning of such investment. The solution in the Near East requires, as the London *New Statesman and Nation* has said, some internationalization of the control of the oil wells, not their further development by rival Russian, British and American "interests".

We come now to the synthesis of our two lines of argument. On the political side we see no hope of an immediate surrender by the great powers of absolute national sovereignty. The only way of working lies in removing specific areas of conflict and in establishing various, purely *administrative*, organizations of international character. On the economic side we want to develop and improve the welfare of the poverty-ridden areas of the world, at the same time providing an expansionary stimulus for the stagnating, mature economies of the west. But we cannot, for both economic and political reasons, hope that this will be successfully done by private enterprise or even by national governments. The true solution, then, is the establishment of international "crown" companies, i.e., developmental corporations of a non-profit sort, whose capital would be subscribed by the various powers with adequate capital-exporting ability. These corporations would order capital equipment from private enterprise in America, state enterprise in Russia, and whatever in-between-sort of enterprise the British genius for compromise develops in the United Kingdom. The boards of directors of the corporations should include representatives of the powers within whose territories the resources to be developed lie, as well as of the capital-subscribing powers. These boards should be responsible to the Economic and Social Council of the United Nations, and should report annually. Presumably they would possess the same degree of independence in administration that ordinary boards of directors do vis-à-vis the stock holders. The employed personnel should be international and should have a prime loyalty to the international organization for which they work.

The economic advantages of such an arrangement are apparent. The people of the economically under-developed countries would enjoy an advancement of welfare, an increase in productivity, and the capitalization of their economies without the dangers of exploitation and domination by a foreign power. The peoples of the mature economies would enjoy, as a result of effects through the real mode, a share of the increased wealth production of the world and at the same time, through the mode of aggregates, an expansionary stimulus to their own economies.

The political advantages—if we may permit ourselves in this Part full licence to roam across the accustomed fences into fields usually closed to the economist—are, first, the removal of certain causes of mistrust and conflict and, second, the working experiences of international collaboration and development of an international administrative personnel in a field of activity, the technico-economic, where it is notably easy for such collaboration to grow. Thus, in these two ways we should prepare the way to a more full-grown international community of a political sort. Apply, in imagination, the scheme to the more troubled areas, Iran, Malaya, Manchuria, even to the Ruhr. There is always the danger of believing in a "one and only" solution of any problem. Clearly there are political problems of the gravest difficulty that would still have to be worked out. Nevertheless, if one considers these problems, say in Iran, in the atmosphere engendered by rival British, French, Russian and American exploitation of oil, and in the atmosphere that would exist if that oil were owned and developed by a corporation directed by nominees of the various governments, including Iran and Irak, responsible to the Economic and Social Council and selling freely to all buyers, one sees how much easier the purely political settlement would be. It is, of course, no panacea. There would still be the temptation on the part of the large powers to dominate the governments of the weak powers. There might be, at times would be, political and national pressures on the directors to behave as plenipotentiary delegates of the governments who nominated them, rather than as colleagues associated to achieve an agreed economic object-

ive. The need for steady progress in the political and legal fields would remain. If another war is to be avoided the peoples of the world must advance together in the development of all sorts of international institutions, in the strengthening of the United Nations, in the control of atomic energy, in juridical methods of settling disputes, in political methods of achieving peaceful change of a rigid but outmoded *status quo*, in the promotion of intellectual and cultural relations among the masses of the people. All we claim is that some, if not all, these things might be made somewhat easier by reason of joint international effort to promote an expansionary high welfare economy.

Is this proposal a realistic and practicable policy in the political sense? This question involves the relation of these international corporations to existing international economic institutions and also their probable reception by the major powers. Clearly it does not in any way replace the need for an international bank and monetary fund. The Bretton Woods arrangements are designed chiefly to make possible and encourage international trading among the industrial powers of the western economy, to solve the problem of chronic shortages of dollar exchange, and to enable the war-devastated areas to restore their trading position and productive capacity in the reconstruction period. It is, perhaps, a fair criticism of the Bretton Woods compromise that it scarcely allows enough latitude to countries attempting to stabilize their domestic economies by compensatory fiscal policies, and that there is a fair prospect of the fund failing to be useful for expansionary purposes, once the reconstruction period is over. The international corporations might well rely on the bank for working capital, might even have recourse to the fund, and might become the channel through which short-term international funds would flow into developmental and economically stimulating use. Once again we emphasize that we are not pressing still another "single solution of all our problems", but suggesting an economically desirable development as a part of an integrated programme of international action.

We turn finally to the question of acceptance. The Soviets have most to gain and least to lose, because they do not enjoy monopolistic control of many such areas. They would doubtless examine such a proposal suspiciously, walking carefully around it to see if it was a capitalist trap, but unless they are really bent on world revolution and domination, they should accept it. It is so clearly in their interests to do so, and it is so obviously an alternative to, not a component of, a "stop Russia" policy. It must nevertheless be admitted that their present stubborn intransigence is such that their co-operation could not be counted on. The British would have the largest sacrifices to make. It would require real political sagacity and imagination. If the philosophy of imperial dominion prevails, they would, of course, reject any proposal of this sort. The British have, however, most to lose in another way; as a people they possess political wisdom, the capacity for the bold gesture. There are already intimations that the Labour Government is prepared to make just such concessions in the British colonial empire.

As for the United States, a lot would depend on luck, the chance of such a policy getting embroiled with domestic political issues, being presented as "socialistic", "appeasing Russia", "idealistic", or even as "paying out good money to save the British Empire". If, however, it could be presented as what it is, the alternative to domestic stagnation and depression, it might succeed in capturing that wonderful American enthusiasm, without which, in any case, no world programme today can succeed.

CHAPTER XVI

THE BUSINESS CYCLE AND ACCUMULATED SURPLUSES

In this work our interest in the special problems of the ten-year cycle has been occasional and only incidental. It has been possible to distinguish for analytical purposes purely cyclical phenomena from long-run trends. When we come, however, to the discussion of policy, that distinction can no longer be readily maintained. The very nature of analysis is to break down a complex whole into parts or aspects of itself, to hold these in isolation and study them by themselves. The construction of policy, on the other hand, must be a work of synthesis, the creation of an integrated programme which holds continuously in view all the variables, all the economic forces and processes which must be channelled and directed towards the achievement of broad objectives of welfare.

Thus the question of finding investment opportunities at home or abroad, of population movement, of overcoming secular stagnation from failure of population growth, all involve policy to control the ten-year cycle. Indeed the distinction in practical policy matters shows itself as a false one. For the control of cyclical unemployment is part of the job of offsetting long-run stagnation in the economy, and the discovery and use of remedial action against long-run stagnating forces is a contribution to cycle control. Similarly, just as the effects of innovations have a long-run aspect, they have, as we saw in Chapter VIII, a cyclical aspect as well. Policy here, too, must be conjoint in its direction and results. Thus a great deal of what we have had to say in the preceding parts, although distinct as analysis from the pure theory of the cycle, has direct bearing on questions of contra-cyclical control policy.

Implicit in what we have written is a qualified acceptance of the theory of the cycle associated with the name of the late Lord Keynes and his followers. We have assumed, and shall

continue to assume, that our readers are familiar with the general thesis of *The General Theory of Employment, Interest and Money.* Without pretending in any way to contribute to the impressive literature on this subject, we shall, however, indicate summarily the qualifications and footnotes which the general theory of long-run changes obliges us to append to pure cycle theory.

It will be remembered that the Keynesian theory of the cycle is a theory of cyclical movements in the marginal efficiency of capital—and consequently in the rate of investment—governed by fluctuations in liquidity, on the one hand, and the propensity to consume, on the other, both of which are dependent functions of the aggregates of investment and income. Thus the cycle is a self-generating one, and its explanation in terms of cause is a self-sufficient one. This means that, whatever the direction and nature of the long-run trend, and whatever the economic accidents of war or revolution, there would be a cyclical movement in the private capitalist economy.[1] Wise fiscal and monetary policies can be devised to reduce the amplitude of this cycle. Looked at empirically, however, the movement of this cycle may be in conjuncture with other processes or events, such as the upward or downward long-run swing in economic activity, generated by a linked advance or a declining population, or the economic adjustments attendant on war and its aftermath. It may be, to choose a special case, that the peak of a long-run linked advance will coincide with the peak of an investment cycle, and that the consequent decline into depression will be the more severe, especially when there is a failure to appear of a new innovation of wide adaptability. This is what happened, we believe, in 1929, with the added conjuncture of a sharp decline of post-war reconstruction investment opportunities, especially in Germany.

[1] My friend, Dr. H. G. Littler, has shown in a series of papers all but one of which are unpublished at the time of writing, that there is reason to believe that there is in all economies, in which ordinary economic motives may be assumed, a pure monetary cycle. He is inclined to believe that this monetary cycle is the true ten-year cycle, and that all other movements of real investment and consumption should be explained in terms of conjunctures and historically. Since, however, his thesis has not as yet been made generally available, I have confined my remarks in the text to the generally accepted theory of an investment cycle.

The footnotes we would add, therefore, to the pure theory of the investment cycle have to do with these conjuncture movements. They are as follows:

1. In a period of general expansion of population with attendant technical advance, the crises may be sharp, but will be followed by a less severe and relatively short-lived depression, while in a period when population is stable or declining, or when technical advance is proceeding at a slow rate, the cyclical upswing will be less pronounced and the periods of depression more prolonged.

2. The nature of innovations and the rate of technical advance have an important and two-sided effect on the cyclical movement of the economy. Innovations of limited adaptability tend to be bunched at the bottom of the cycle. This is because they are introduced to replace obsolete and worn-out capital instruments whose replacement cannot be longer postponed but which are possible only on the probability of reduction of operating costs. The fact that these innovations tend to be of the labour-saving sort, and to be of strictly limited adaptability, means that, in the absence of other events, they reduce the probability of full employment in the recovery period.

Linked advances, on the other hand, generate long-run upswings that, speaking generally, tend to reduce the amplitude and duration of cyclical downswings. As we have already seen, however, the conjuncture of a peak in the long-run technical advance with an investment cycle peak may, in the absence of new innovations of wide adaptability, lead to a serious and prolonged depression. More particularly, we have argued (Chapter VIII) that a linked advance will lead to increased investment and increased national income with increased concentration. There will be, we saw, a larger volume of saving and, for a time, a diminished preference for liquidity. The decline in anticipated profits with the eventual decline in marginal productivity (there being eventually a decline in the rate of adoption and there being, for a time, no further technical advance of any generality of adoption) will lead to a decline in the demand for loans at current interest rates. The probability, we showed, was for liquidity preference

to jump at the same time, leading to higher interest rates and a consequent sharp decline in the rate of investment. Entrepreneurs, especially those guiding the affairs of the great corporations, will prefer to hold company reserves, as well as private fortunes, in the form of "gilt-edged" securities rather than venture them in real capital expansion. It is this kind of causal conjuncture which leads to the severe and prolonged type of depression. The recovery will be brief and imperfect if the sort of investment on which it depends is in innovations of limited adaptability. It may be complete and lead to new levels of activity and real income, if new innovations of wide adaptability appear.

3. We should like to throw more emphasis than is usual on the existence and nature of rents and monopolistic surpluses. We have shown these differential gains to be unearned and unnecessary to cover real costs. Their retention by private corporations has therefore two effects. Directly, this retention affects welfare unfavourably by concentrating the national income in the hands of a few who have the least urgent need of increments of income. The real income, in terms of satisfactions, is less great than it would be in a community in which the same national income were more evenly distributed. In the second place the concentration of these surpluses leads to a lowered propensity to consume and is the basis of those very reserves which, unspent and uninvested, we saw in the last paragraph to have a destabilizing effect on the economy. Professor Schumpeter, however, appears to believe that these differential gains are necessary to the progress and healthy activity of the private capitalist economy. Here, surely, is the intrusion of doctrinaire thinking. These surpluses, with their destabilizing effects and their negative welfare effects are socially undesirable, and that is not in dispute. But economic progress in the trustified kind of capitalist economy we have is impossible without them. That fact, too, we should not, on our part, dispute. Professor Schumpeter concludes that these "windfall gains" are therefore necessary. That is logically a fallible conclusion. The proper conclusion is that these gains are necessary to the progress of an economy in which these great

monopolistic corporations are a characteristic institution. If we believe in the retention of these institutions as an article of faith beyond the reach of sceptic doubts, we must conclude that these dynamic differentials are a necessary evil. If, however, we bring the characteristic institutions of western capitalism down from the altar of idolatry and subject them to criticism, admitting the possibility and desirability of purposive change, we are logically permitted to conclude that these surpluses and their evil effects could be attacked by policy, realizing, however, that in this area of decision was involved a modification of one of the characteristic institutions of our economy. It is our contention that this is an area of free decision and that policy towards this end can at least be considered.

Now the policy usually advocated by Keynesian economists for control of the cycle is a compound of monetary and fiscal control. The monetary authority directs its attack against sharp jumps in liquidity by carefully timed purchases in the open market, and the fiscal authority, in close harmony with the monetary authority, budgets so as to over-balance the budget when the rate of private investment exceeds the rate of saving in order to to avoid inflation, and to under-balance when the level of investment is inadequate to maintain the national income requisite for full employment.[2] Recent Keynesians[3] have laid more emphasis on fiscal policy than on monetary policy as a technique of control, but it is doubtful if Lord Keynes himself would have put the same degree of confidence in purely fiscal methods.

There are certainly reasons to believe that fiscal policy, by itself, may prove disappointing as a method of cycle control, though we do not intend to mean that an appropriate fiscal policy should not form a part—and an important one—of general economic policy. If, however, reliance is placed on

[2] This is pretty cryptic and not too accurate. The clearest brief statement of Keynesian fiscal policy which I have seen is that of my friend and colleague, B. H. Higgins, "Post-War Tax Policy", *Canadian Journal of Economics and Political Science*, Vol. 9, Nos. 3 and 4. Dr. Higgins's treatment has the great merit that it applies the principles to the post-war Canadian budget and thus, by illustration, tests and illuminates the general principles.

[3] Cf. B. H. Higgins, op. cit.; Abba Lerner, *Economics of Control* (New York, 1942), and others.

a compensatory fiscal policy, and on fiscal policy alone, as a contra-cyclical measure, it is apt to prove disappointing and inadequate. This is for the following reasons.

1. The pure investment cycle by itself is, shall we say, comparatively innocuous. The really serious depressions are the results of conjunctures of the cycle with a downward movement of the long-term trend. Of course anything that successfully reduces the amplitude of the pure cycle is all to the good, but the fiscal remedies appropriate to that condition are, with respect to the long-run process, a treatment of symptoms, not of causes. Employment can be established by governmental spending, but it will not revive private industry in the face of basic causes of decline unless suitable investment opportunities are discovered, new innovations of wide adaptability introduced or new markets developed by bringing new populations into a purchasing position. As we have repeatedly urged, it is not simply a matter, as Professor Hansen seems to imply, of the rate of growth of population in a single country, it is a matter of bringing populations into a position to purchase. The trend in sugar consumption is down because sugar is consumed by the masses of the west whose diet is already advanced beyond the point of maximum per capita consumption of sugar. If the populations of the east could become sugar consumers the present sugar-producing establishments would be quite inadequate to meet the demand. The same thing is true of many commodities. Newsprint sales, to take a vivid example, have a high correlation with literacy and free compulsory schools. The demand of newsprint is exceedingly inelastic to changes in the price of newspapers. If the per capita consumption of newsprint in the rest of the world were just one-half as high as in Great Britain (before the second World War) present newsprint capacity of the world, at maximum physical production, would meet about one-third of the demand. Investment opportunities in Canada in this field alone would exceed the total capitalization of the present Canadian newsprint industry. If only Canada and the United States had a per capita newsprint consumption equal to that of Great Britain the Canadian industry could operate in a "normal" year at 100% capacity

instead of the 80% reckoned, on the basis of present markets, as "normal". We need not unduly flog this willing horse. Already, in Chapter XV, we have stressed the importance of an international programme for providing investment opportunities for western capital and for bringing new populations into a purchasing position in the market. The frontier awaits expansion, but it is not a geographic frontier. It is the frontier of our social thinking and our social conscience. The attack on poverty on the one hand and on inert or stagnating industrial economies on the other requires concerted international effort, in which the lead must be offered by the capital exporting countries of the west, but to make this lead effective they must, as we have argued, modify their economic institutions to the extent, at least, that they put developmental international investment on a non-profit making basis.

Internally, too, a proper population and immigration policy, governmental encouragement of research and industrial and technical development and the raising of the purchasing standards and purchasing capacity of the submerged classes is a necessary part of governmental policy. Without these, public spending is apt to be a mere boondoggling that may prevent or postpone the full impact of depression but which can do little to remove the basic causes.

2. There are political dangers in the compensatory fiscal programme. If it is undertaken by a government foresworn to the protection of "private enterprise", the authorities find themselves committed to public spending but limited to certain fields. They must not "compete" with "private enterprise". This means (a) that certain important areas of investment, such as working-class housing, are closed to government, areas in which spending will be socially beneficial as well as economically stimulating, (b) that if private industry chooses to resist the government programme the weapon of expropriation is denied the public authorities as a coercive instrument, (c) that, because public spending is confined to unprofitable fields and because government lacks coercive power, private business could discredit the public spending as wasteful "pump-priming" and could demand a return to "sound finance" and a balanced

budget and could coerce government simply by refusing to maintain private investment unless taxes were reduced and the budget balanced. Indeed it could plausibly be argued that private capital could not be ventured at the rate of return, net of taxes, allowed and that continuous deficit budgets would be necessary. This kind of capital strike seems to have been threatened in the United States and Canada in the post-war period. (d) Finally, there is the danger that public spending might be distributed regionally on a purely political basis without regard to regional needs, regional leakages in the multiplier and long-run considerations of industrial location.

We are inclined to favour the view that governmental spending be determined by the necessities of an adequate long-range attack on the basic economic problems, i.e., resource development, social amenities such as housing, governmental enterprises to compete with or replace monopolistic industries and so forth. Further spending on social security and the ordinary governmental and administrative needs should constitute a sum of sufficient magnitude for budgeting to be planned over the cycle.

3. Keynesian fiscal policy, of itself, contains no attack on the monopolistic nature of modern industry or on the existence of the monopolistic surpluses or dynamic rents which are harmful, we have argued, both to welfare and the stability of the economy. We believe that a definite policy towards monopolies, trusts and cartels must be worked out in the modern state and we are doubtful of the effectiveness of purely fiscal and tariff weapons for this purpose. Tariffs, as far as we can observe, are used by monopolies and cartels, not against them. Excess profits taxes and corporation taxes usually operate to the detriment of firms in competitive industries, and even tend to increase the degree of monopoly. That some success may be achieved by an ingeniously contrived tax, we do not deny. Dr. B. H. Higgins, in a recent article,[4] has suggested such an ingenious device to cajole monopolistic industries to operate at full capacity.

[4] B. H. Higgins, "Post-War Tax Policy", *Canadian Journal of Economics and Political Science*, Vol. 9, No. 3, pp. 325-426. Dr. Higgins attributes the genesis of his plan to a suggestion made in the Report of the Canadian Royal Commission on Price Spreads (Ottawa, 1937), for "the taxation of surplus profits beyond an allowed rate of return on the real investment in utilized capacity".

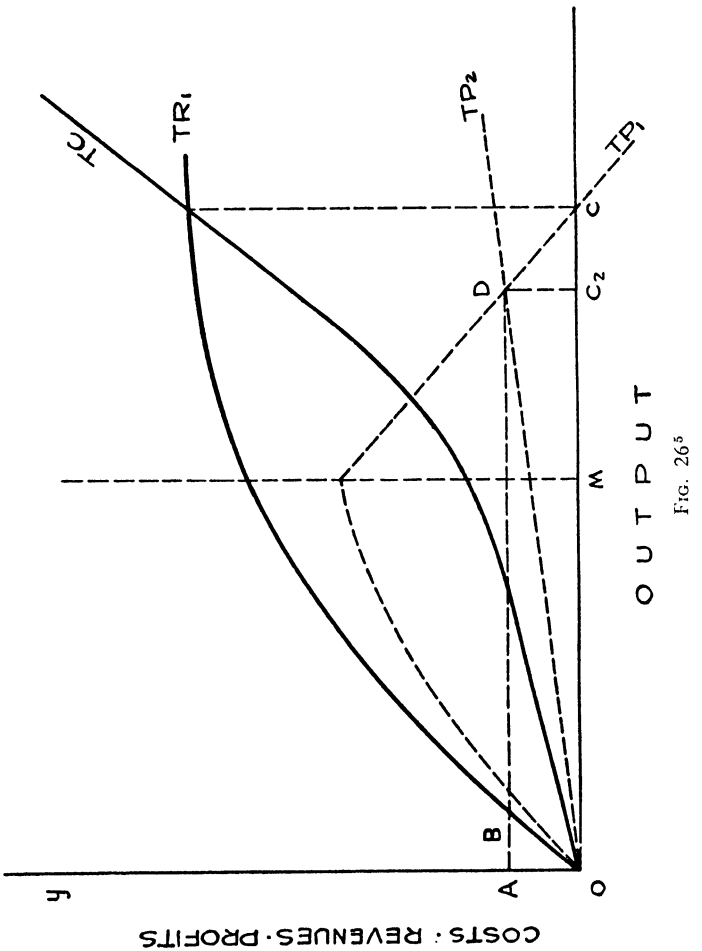


FIG. 26[5]

[5] Reproduced by kind permission of Dr. Higgins, and the editors of the *Canadian Journal of Economics and Political Science*. I note that the TP_1 curve has been drawn as though it were a straight line from the peak on. This is, of course, accidental, and has no significance.

In Figure 26 let TC be the total cost curve, TR_1 the total revenue curve of a monopolistic firm, so that TP_1 is the curve of the firm's total profits. Then C will be the point of "competitive" output, in the sense that it is the no-profit point, and M will be the theoretic monopoly output. The object is to cajole the firm to move from output OM to output OC, or an approximation to OC.

Now let a tax be imposed at the rate of 100% on all profits in excess of say 5% of real capital value, and let 5% of capital value be represented graphically by AO. Then the curve of total profit net of tax will be $OBDC$. It is a matter of indifference to the firm what output it produces between OB and OD. To induce the firm to move towards OD, Dr. Higgins suggests that the tax-free profits should be 5% of utilized capacity (instead of real capital investment as suggested by the Royal Commission). Capacity must be defined "to vary directly with output" (p. 426), not a proposition to appeal to the business community, chartered accountants or the tax administrator. However, in that case the curve of net profits (after tax) becomes a gradual positively sloped curve, lying for most of its length below ABD. This curve is shown as TP_2. (Its dotted extension beyond its intersection with TP_1 must be meaningless as long as lay-out is not increased.) Hence the now established output of maximum net profit is OC_2, an approximation to the desired OC. The lower the rate of tax-free profit and the lower the rate at which "utilization increased with profit" (p. 426), the closer will the approximation lie to the theoretic goal.

"The attraction of this scheme", Dr. Higgins writes, "is its simplicity and its virtually fool-proof character. It is unnecessary to know the exact shapes and positions of the cost and revenue curves, or even to have an accurate definition of capacity. It works equally well for monopsony and for monopoly, and for mixed cases. It is only necessary that profits after payment of taxes should rise continuously with output" (p. 426).

Dr. Higgins' suggestion is probably sound enough economics and it should have a place in a general control programme. There are certain industries, as we shall see in the next chapter,

where trustification cannot be successfully attacked by anti-trust laws and which cannot advisedly be expropriated by the state. To encourage such industries to operate to capacity a tax, such as Dr. Higgins advocates, would seem an appropriate remedy. The plan poses administrative questions of difficulty. The executive definition of capacity as a variable would surely raise accountancy problems involving questions of equity. Great resistance on the part of industry would be encountered. I am inclined to believe that the scheme would entail more "nuisance bureaucracy" and governmental inquisition and interference than would more direct and radical controls. The levy would have to violate a cardinal principle of orthodox tax policy, viz., a fixed schedule of rates for all taxpayers, as the actual levy would be determined according to percentage of capacity and not income. This might open the way for serious abuses. There is also the complicated question of the relation of this tax, levied against domestic trusts, to tariff and excise taxes on the produce of foreign competitors, though this problem could, I think, be worked out with equity.

I cannot help but feel, however, that the whole problem of monopolies, trusts and cartels requires a more direct approach. To this question we shall now turn.

CHAPTER XVII

MONOPOLIES AND CARTELS

It is commonly believed today that monopolies, trusts and cartels are socially undesirable. This belief rests on a sound social instinct, but unfortunately public understanding of the matter has been so confused by propaganda that people have come to mistake form for substance and to believe that, if they can only prevent the legal fact of a contractual combination and so "restore competition", all will then be well with the "system of private enterprise". Nothing could be more wrong-headed as a basis for policy than this legalistic legerdemain. It confuses the issue, misleads people as to the nature of the problem, and incorrectly defines the area of decision.

The object of economic policy is welfare. Perfect competition is not, and never has been an end in itself. When it has been defended by economists—when Adam Smith defended it—it has been defended because "the public would be the gainer, the work of all artificers coming in this way much cheaper to market".[1] We are all familiar with many instances of economic activity in which, for obvious welfare reasons, we should not find competitive enterprise tolerable. A town served by a dozen competing tram companies would not be a convenient place to live. I am, as I write, in an area served by competing telephone companies. As a consequence it is more difficult to speak to someone at Dalvey, some twenty miles away, than it is from Montreal to speak to someone in London or even Paris.[2] Society has long recognized the superior efficiency of monopoly in what are called "public services" and "public

[1] Adam Smith, *The Wealth of Nations* (J. M. Dent ed., London, 1910), Vol. I, p. 112.

[2] This is not an exaggerated and speculative illustration. A fellow-guest at the pension but yesterday attempted to make this call. It took over an hour to get the connection, and then he had to speak through a third party in Charlottetown. In the exercise of his profession this man commonly talks from Montreal to London and Paris. The connection is always quicker and sufficiently distinct as to obviate the necessity of intervention by a third party.

utilities", and has begun to recognize the need for social control or ownership of all such enterprises. There has been also a steady growth in the number of activities comprised by these concepts. At one time education, communications, the provision of roads and bridges, transportation and other services now recognized as public services or public utilities were regarded as proper to private, competitive enterprise. It will not be long before transportation, some housing, health services, the provision of power and probably of base metals will also be included in this group.

Not only, however, are public utilities more efficiently provided by single firms than by competitive firms, but also in many other lines of industry, the economies of scale are such that the public is better served by a few large firms in monopolistic competition than it could have been by the small-scale establishments of a perfectly competitive industry. This is not always true. The evidence we have previously referred to,[3] shows:

A. In some lines of industry, e.g., bespoke tailoring, brick-making, "custom-built" furniture, etc., where either the skilled nature of the work or the limited nature of the market, make for small-scale organization, no increase in efficiency is achieved by increasing scale.

B. In some lines of industry, e.g., shoemaking, ground-wood pulp manufacture, etc., the team or technical combination of labour and machine affording maximum technical efficiency is achieved by a fairly small-scale organization and no additional economies appear to be achieved by further increases in lay-out. Indeed increases in lay-out can only effectively be made by doubling, tripling and so on the basic team, and such small economies as are achieved are offset, frequently according to the empirical evidence more than offset,[4] by the rising marginal cost of management. In any case such economies as are achieved in this group are ordinarily external to the firm.

C. There is a great and growing number of industries in

[3] See Part V.

[4] See T.N.E.C. Monograph, *The Relative Efficiency of Large, Small and Medium-Sized Businesses.*

which the characteristic mass-production methods of our period are possible, and these industries evince economies of scale up to an optimum scale of great size.[5] The tendency is for the point of optimum scale to shift positively as technical advance enables the firms to use more efficient power units, to reduce

[5] The extent of this sector of our economy is not always realized. Taking the existence of monopolistic practices as the criterion (not always justifiable, be it admitted), I have prepared the following list of commodities and services which, in Canada, are affected by monopolistic practices. The list, which does not pretend to be complete, is based on evidence contained in L. Reynolds, *Price Control in Canada* (Cambridge, Mass., 1940); the *Report of the Royal Commission on Price Spreads* (Ottawa, 1935); and *Canada and International Cartels* (Ottawa, 1945). We must remember that in alleging monopolistic competition to exist, we are not alleging that a combine, illegal under Canadian law, exists. Nor are we alleging any moral depravity on the part of business men in the various industries. They are not usually responsible for a situation whose real cause was in basic technical and social forces. Pure monopoly touches only public utilities and a few other commodities. Aluminum, nickel, cotton thread, matches, bricks and building stone in some local markets, are examples of near monopoly.

The following commodities are affected by international cartels:

A. *Imports*

Potash	Certain Aircraft instruments
Nitrogen fertilizer	Beryllium
Phosphorous	Diamonds
Quebracho (essential for tanning)	Camphor
Flat glass	Optical goods
Sulphur	Magnesium
Some dyes	Some plastics
Tungsten Carbide	Pharmaceuticals
Magnesia Refractories	Diesel engines & others
Titanium pigments	

Many of these are key materials for industry or agriculture.

B. *Domestic manufactures connected with foreign cartels*

Some chemicals
Electric Lamps
Radio Tubes and Sets
Matches
Some types of paper and paper board

C. *Exports*

Copper
Acetic Acid
Nickel
Radium
Aluminum

Some of these cartels have broken down since the war. There is no suggestion that any of them were illegal under Canadian law.

The following commodities are produced under conditions of monopolistic competition, whether formalized or not, and are representative only of the more pronounced cases:

Coal	Farm Machinery
Oil	Automobile and Trucks
Electrical power	Newsprint
Railway transportation	Kraft Paper
Ocean shipping	Paper Board
Road Transport and Trucking	Bond Paper
Iron and Steel and their Products	Radios and Parts
Machine Tools	Electrical Instruments

more and more operations to the conveyor belt, to eliminate hand-work bottlenecks and to improve the techniques of management, administration and merchandizing. This process we have examined theoretically in Parts III and IV. We have seen how it is simply an aspect of technical advance, and it leads inevitably to ownership concentration. We do not believe this process could be reversed without grave sacrifice of industrial efficiency. To attempt to break up the Aluminum Company of America, or Aluminium of Canada, Limited, or the International Paper Corporation into two or three thousand independent plant units to produce the same total output would increase costs by an incalculable amount.

Indeed we should remember that perfect competition is an analytical concept, not an historical reality. As an analytical concept it has been, as any economist knows, of immense value. When, however, economic illiterates use the concept as a description of historical reality, something out of a golden age to which we should aim to return, they are committing not only a logical fallacy and historical solecism, they are directing policy towards a false aim, dangerous in the extreme to social

(*Fn. 5 continued*)

Refrigerators	Typewriters
Vacuum Cleaners	Books
Electric Washers	Fountain pens
Aircraft and engines	Alcoholic Beverages
Ships	Tobacco
All Chemical and many Pharmaceuticals	Some Types of Retail Trade
	Canned foods
Toilet Preparations, Soaps, Cosmetics and Lotions	Bread—in some communities
	Dairy Products—in some communities
Furniture	
Most Textiles	Packed Meats
Some textile products, e.g., mens' shirts	Bricks and Building Materials—in some communities
Women's Hosiery	Insurance and Banking Services.

There are probably important omissions.

Even this does not really tell us how far these practices affect us. Consider shoes. In most lines of shoes there is pretty full competition both in distributing and in final manufacture. But if we look further back we find the shoes are made on the machines of a monopoly which leases them to the shoemakers. Moreover, the leather is bought from tanners who are definitely monopolistic in their practices and the tanners themselves have to buy quebracho from a cartel which holds a complete monopoly over this material. Thus a considerable element of monopolistic profit enters indirectly into the price the consumer pays for shoes, though the Canadian shoe manufacturer himself gets none of it.

welfare. There never was such "perfect competition". Business enterprisers have always attempted to escape the discipline of the too competitive market. "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".[6] The decline in the number of firms has simply made the conspiracy easier. When the situation reaches the point of oligopoly some sort of agreement, tacit or open, becomes necessary. Otherwise, as we have seen, the "firms will chase one another down their cost curves into bankruptcy."[7] If anti-trust laws are enacted to prevent explicit agreement under these conditions, they succeed only in making it tacit. They cannot create the conditions of perject competition or even an approximation to those conditions. The economic process forbids it. If "competition is restored" and the trust-busters happy, the competition that exists is monopolistic competition and the trust-buster's happiness is of a fool's paradise. "It is impossible to prevent such meetings (i.e., of people of the same trade) by any law which either could be executed, or would be consistent with liberty and justice."[8]

We are not, of course, trying to say that unregulated monopoly, monopolistic competition, trusts or cartels are in the public interest. They are not. But the evil is not in the legal fact of combination. It is in the economic condition of monopolistic competition. This evil cannot be remedied by anti-trust laws. They are irrelevant to the facts of the economic process if directed against the legal fact of agreement. If one could conceive of anti-trust laws directed against the physical fact of large-scale organization, and if one can conceive of such laws actually being enforced and administered, they would still form no remedy in the public interest, for they could be enforced only at the sacrifice of the economies of scale that, in the first instance, led to the situation of the industry.

---

[6] Adam Smith, op. cit., p. 117.

[7] I attribute this quotation to Mr. Harrod, but I am unable to find where he uses it. If my memory misleads me and I do some other economist phrase-maker an injustice, I give him my apologies.

[8] Adam Smith, op. cit., p. 117.

Some people, having reasoned so far, and having reacted violently against the propaganda techniques of such organizations as the American Department of Justice and the shoddy arguments of the trust-busters, have concluded that, though there are abuses in monopolistic industry, it is better to suffer those abuses than to sacrifice the advantages of modern productive techniques and to bring the law into contempt and disrepute by attempting statutory prohibitions where these are clearly inappropriate and ineffective. This is a comfortable and Panglossian philosophy, but a false one.

The evils of the trust, the dangers of ownership concentration cannot be exaggerated. They consist in part of conspiracy against the public, of price-rigging, of misallocation of resources, of interference with the free flow of trade, of the waste of excess capacity, of excessive profits, and of consequent economic instability, indeed of all these economic evils which economists since Adam Smith have recited.[9] But I doubt if these are, even in sum, the chief part of the indictment to be made against them. The gravamen of the charge to be preferred is the awful social and political effects of the concentration of economic power.

Political power follows economic power. The Athenian democracy which the aristocrat Plato criticized was the crea-

[9] I omit the charges made before the Kilgore Committee of the United States Senate. Evidence there adduced showed suppression of inventions, deliberate evasion of law, price discrimination in the case of needed medical materials of a shocking nature and even activity of an almost traitorous sort. I omit these charges because a Senatorial investigation is not a Court trial, and the charges cannot be regarded as proven. Again it does not seem fair to regard them as typical. Department of Justice officials giving evidence were anxious to make the participants in trust and cartel agreements seem as black as possible. They selected forty or fifty cases as bad as could be. These cases cannot be regarded as typical. We could doubtless find some pretty bad cases of business skulduggery if we examined enough competitive firms, but we should hardly regard them as typical of business behaviour under competitive conditions. Some of the biggest monopolists and cartelists are among our most enlightened employers and in their labour and community programmes set an example we could wish others would emulate. In Canada, Bell Telephone, Canadian Industries Limited, and most of the newsprint paper firms are progressive, liberal and enlightened far beyond the average Canadian business concern. The social evil of monopoly or cartel has its source in a situation or condition, not in the wickedness of individual business men. Once more it is a case where hatred and emotional attitudes are not helpful counsellors of social action, and where only analysis and reason can lead to remedies that really attack the cause of our discontents.

tion of the traders of an expanding mercantile economy. The republic on the Tiber arose after an industrial revolution which destroyed the economic position of the old feudal or patrician class, declined and fell when a new patrician order came to control the vast estates of conquered lands and with them the food supply of the imperial city and its armies. European feudalism as a political and social system was based on the ownership of land, the principal, and for long, the only productive capital of society. As a new merchant class slowly emerged the sovereign power of feudal princes was seized by corporations of city merchants, at first in the chief centres of Mediterranean trade, later as the mercantile economy spread along the water routes of trade, in the towns of Flanders and the Rhine, the Baltic ports and the Channel towns of England. The constitutional history of England reflects these shifts in the reality of political power to conform to shifts in the possession of economic power. King, Church and Barony owned the land and governed. The history of the conflict over investiture, ecclesiastical courts and taxation, of the civil wars of Stephen and Mathilda, of John's reign, of Magna Carta, the establishment of the Circuit Courts, and the war of Simon de Montfort, is the record of the struggle between Crown and Church and Crown and Barons for ultimate sovereignty. The alliance between Crown and merchant was everywhere an uneasy one. When the Crown found in the new merchants and in the woollen trade a source of economic power and of cash for the maintenance of armed force independent of the untrustworthy feudal levy, it was able to emerge triumphant in this three-cornered struggle, to subject the Church, destroy the feudal order and to create from among its bourgeoise supporters a new aristocracy supported with Church lands, and the wealth of trade and the treasure of the Spanish Indies.

The constitutional struggle of the seventeenth century was the culmination of this battle. The party of the Crown supported the attempt of the Stuart line to free the King from dependence on his erstwhile allies against the Church and the old feudal order. The "King's writ ran in all the land", he, through parliament, was supreme. Now he wished, following the ex-

ample of monarchy in France, to abolish this qualification on his power. In his party were those of the old order, the squirearchy for whom land and husbandry remained the basis of power and privilege, who saw in the merchants and the masters of enclosed land their mortal enemies, whose every encroachment marked a step in the steady invasion of tillage land. Against them was the wool trade, the cementing interest of the strange alliance of urban traders and great aristocrats that was to become the Whig Party. The great landowners were enclosing and using land for pasture. The trade of England depended on the staple export of wool. Here was the new wealth, the new concentration of economic power and it was sufficient to carry the Whig party to a power which was unbroken and nearly unchallenged for a century.

The new Tories of Pitt and George III succeeded in catching power on the strength of reaction following the French Revolution and by basing their strength on the corrupt borough system. They found the economic support necessary for bribery and the purchase of rotten boroughs in the spoils of the East India system. In a word, their economic basis was Indian exploitation, not the sound and expanding foreign commerce of Britain. Their economic basis was impermanent and unsound and in 1832 the Whigs overthrew them. The Tories missed their opportunity of founding themselves solidly on the new industrial classes the Industrial Revolution had created, which they might have led to victory over the older mercantile and commercial interest led by the Whigs. As it was they were allowed to catch up the new industry and so to strengthen the foundations of political power. Under Peel and later Disraeli they tried belatedly to win the support of the industrial masses, but their successes were impermanent; they sprang from one opportunist platform to another; they were unable to find either a consistent system of principles or a firm economic basis.

This endured until in the twentieth century the continued play of economic circumstances undermined the whole Whig system. The great corporation replaced the independent entrepreneur, ownership was concentrated, restriction and protection from competition was demanded in place of the

free market, labour became organized also, and came to possess great economic power. The Liberal spirit had its final flowering in the magnificent achievements of the Campbell-Bannerman and Asquith governments, but even then it was apparent that on the one hand the new industrial giants were moving into opposition, disgusted by the Lloyd George budget, despairing of protection from a government bound by the tradition of free trade and frightened of the concessions given labour. On the other hand the reforms were insufficient to meet the demands of labour, so that this support, too, dwindled away. The great Whig compromise was ended, and the Labour Party emerged to represent the new economic power of organized labour and the Conservative Party with its traditional strength of Church, reaction and agriculture augmented by the real economic power of corporation capital came once again to taste the reality of power. Its record of subservience to the Federation of British Industries proves that it had learned where its strength derived, though that seems the only lesson it had learned since it toasted the King and Bonnie Prince "across the water", for surely not even the Tory governments of North or Pitt have rivalled the blundering and pusillanimous betrayals of the late Earl Baldwin and the late Mr. Chamberlain.

In Germany the great military and political power of the Empire and the third Reich were based on the fostered growth of great monopolies. The Nazi regime was the undisguised exercise of absolute political power by those possessed of concentrated economic power. True, in the later stages of the Nazi rule, the Party and the Army freed themselves of industrial control. This is to be understood as the natural result of the mistake the German industrialists made. Political power can be safely wielded by those possessed of economic power through the corruption and manipulation of the institutions of sovereignty as in Britain in the inter-war period and as in the United States. An open fascist regime, i.e., the brutal exercise of direct and absolute power, can only be achieved when some sort of mass support can be organized and won, as Hitler succeeded in doing among the lesser bourgeoisie and the unskilled and unorganized and bewildered mass of unemployed

workers. The trades unions, of course, he never captured. The Communist Party greatly assisted him by preparing the unorganized workers (a) to hate the Social Democrats of the Trades Unions during their struggle against the Socialists and (b) to accept the methods of violence and "the street" as a substitute for ballots and parliamentary procedure. Thus the Nazi Party, with its mass organization, found itself finally in possession of the great police powers of the modern state and able consequently to free itself, at least in part, from those who had created it.

However this may be, the records of constitutional struggle in all countries teach the same lesson. Concentrated economic power is a constant threat to any sort of broad-based or democratic political sovereignty. The process we have observed which is occurring in our economy is a danger to the state. If it is allowed to go on unchecked, it becomes questionable if parliamentary government can continue, or, if it does continue in outward form, if it can for long express in any approximate way, popular sovereignty.

There is a further political danger to be observed. This is the threat to the concept of a community of interest. The notion of a national interest is a condition of parliamentary democracy. That national interest is not absolute and unchanging, it must constantly be reformulated, but at any time it is there, and different parties represent alternative formulations of it and alternative methods of attaining it. There is a division in interpretation, a division over means, but there is agreement that a true community exists on the strength of community of interest and purpose. In the society of the nineteenth century, such a community of interest did, in fact, exist. Though there were many divisions, many conflicts of group interest, there was an over-riding, if imperfectly realized, community interest to be achieved. A free market with an expanding economy was in the equal interest of producer and consumer, capitalist and wage-earner. The profit motive in a reasonably competitive society made for an optimum allocation of resources and the maximization of income. Today this is not so; the profit motive leads to restriction of output, wasteful excess

capacity, misallocation of resources, a reduction in national income. There is a real conflict between producer interests and consumer interests. Similarly in the labour market, when there is buyers' competition, the worker gets a wage commensurate with his productive contribution. When there is monopsony, he is exploited in the sense that his wage is less than his contribution to produced value. It may well be that, because of improved technology, his wage is considerably greater than it would have been had this improvement not occurred and the factor market remained competitive, but, though this fact is important as an economic welfare consideration, it is perhaps insignificant in relation to the political problem we are now considering. This is because social strife is occasioned more by the sense of injustice or exploitation or comparative poverty, by the feeling of being used and taken advantage of, by the impression of comparative want and great insecurity of tenure in the face of accumulated wealth, power and security, than by a lesser degree of welfare in a society which, however, is progressing and is conscious of a mutual sharing of the benefits of progress.

The failure to define or formulate a concept of the national interest, to associate all classes in the sense of national achievement, the consequent growing recognition of conflict and the organization to exert political pressure for the victory of group or sectional interests, is fundamental to the disintegration of modern capitalist society. If that disintegration is to be stopped and a healthy parliamentary democracy re-established, the great concentration of economic power must be brought under social control. It cannot be left to possess itself of the reality of political sovereignty and to rot and corrupt our political system and our sense of community.

We have shown that the concentration of economic power occurs in modern economic society. We have analysed how and why it occurs and why it must continue to occur. We have shown that legal action of the anti-trust type is irrelevant, ineffective and undesirable as remedial policy. We have shown, nevertheless, that the process of concentration has most undesirable economic consequences and is fraught with the most

serious menace to our political institutions and social stability and harmony. We must now ask along what possible lines effective policy might be directed.

We shall consider first domestic monopolies and combines. In our opinion the policy that is practicable is not the didactic application of any single formula but rather, in the words of the Report of the Canadian Commissioner under the Combines Investigation Act, "an integrated programme to ensure that the possibilities in modern industrial efficiency are developed and stimulated to serve the public interest".[10] This integrated programme should combine various remedies, each to be applied in the instances to which it is appropriate.

The provision of certain essential services, such as power, transportation, communications, banking services, public health, low-cost housing and, probably, the provision of fuels and base metals, can most efficiently be supplied by a monopoly, and, by common consent, such monopolies cannot be allowed to function free of public control. In many instances the services are supplied by publicly owned corporations, either municipal, provincial or state, or national. In other instances privately owned corporations supply the services, but do so under the supervision of a Public Utilities Board or Commission. Public utility commissions usually regard their chief function to be regulation of rates to prevent monopolistic exploitation of the public and general supervision of the quality of service provided. In three ways the method of public utility supervision is inferior to public ownership. It fails to prevent the concentration of economic power in private hands, to which we have referred, it leaves giant corporations with this power to influence public opinion and even to corrupt, sometimes, the supervisory officials, and, finally, the experience in controlling rates (or prices) has been that most rates or prices can be justified by corporation accountants on the basis of existing costs, and it is impossible to show that the existing cost structure is an unreasonable one. Nevertheless we all know of instances where we should hesitate to recommend the municipal or provincial ownership

[10] *Canada and International Cartels* (King's Printer, Ottawa, 1945), p. 68.

and operation of public utilities. Where the record of political corruption is bad, one would hesitate to offer further rich opportunities for graft to the politicians. In general, however, we should admit that the principle of public ownership is the appropriate one for this group of services.[11]

At the opposite end of the scale from these "natural" monopolies, as someone has called them, are a few instances where small-scale firms, who would normally compete, are organized in a trust, pool, or combine. In such cases, where the combine is a legal act and a conspiracy against the public conforming to no economic or technical necessity, anti-trust legislation, adminstered, as is the Canadian Act, with great powers of discretion allowed the Commissioner, may well be effective. We should emphasize, however, that, in our opinion, the number of cases in which such action would be appropriate is small and is becoming smaller. There are also more frequent cases in local markets where the chief function of a combine is to prevent the entrance of new capital and enterprise. There, too, anti-trust legislation may be appropriate and effective.

Between these two extremes of monopoly and near competition, in which public ownership and anti-trust legislation are appropriate remedies, will be the majority of domestic combines and trusts, the great sector of monopolistic competition. I should not, myself, exclude public ownership as one remedy appropriate in this sector. I should think, however, that any democratic state that began, like the United Kingdom under its present (1946) government, with the nationalization of central banking services, coal, steel, transportation and communications and health services, would have its hands pretty full. Administrative experience would have to be acquired. The principles of parliamentary control over broad issues of policy and corporation autonomy in purely business affairs would have to be worked out. Industries where monopolistic competition obtained, could be nationalized only slowly and as

[11] I do not deal with the vulgar notion that public ownership is inefficient. There are some instances of corruption, just as there are instances of corruption of public utility commissions. But the general standard of publicly owned corporations is good, and, indeed, is so much better for the public interest, so much more really efficient, that no one who knows the facts requires further argument.

they "became ripe" for expropriation, i.e., as it became apparent that other controls in the public interest were ineffective or inadequate. Over a considerable period of time, therefore, firms in monopolistic competition would continue to be privately owned and would constitute, as we have argued, a danger to public order and a detriment to economic welfare. It is in this sector that Dr. Higgins's tax against excess capacity might on experiment prove one appropriate and effective remedy. Publicly owned yardstick plants are another remedy that in certain instances have been effective. The consumers' cooperative movement has had beneficial effects in forcing firms in monopolistic competition to reduce their prices, and has the further advantage of developing democratically organized consumer bargaining power as a counterweight to the power of the combines. In many countries, too, such as Canada, such combines and price agreements are able to exist only by reason of the national tariff. The general reduction of national tariffs and the abolition of the tariff on all articles marketed under pricing agreements, whether tacit or otherwise, would have most healthy results.

One further control will be in public hands. A compensatory fiscal policy and the fact that government is already "in business in a big way" will combine to make publicly owned corporations and government departments very large purchasers. The power of monopolistic sellers will be diminished in all markets where they are confronted with one such large buyer. Monopolistic competition among sellers will be faced by monopsonistic competition among buyers, the chief of whom will be the institution, charged above all others with the protection of the public interest. Wartime experience has shown how powerful may be the buying power of government as an instrument of economic control.

Eventually, I strongly suspect, all large-scale industry will become state-owned and the private sector of a mixed economy will be confined to small-scale industry and services. I do not, however, think that such socialization could be successfully attempted in a single and abrupt step. It is doubtful if either the techniques of political control or of administration are

sufficiently worked out, or could be sufficiently worked out in advance, without the experience of a gradual policy, to permit this kind of socialization to occur without serious harm to parliamentary democracy and to the administrative competence of the expropriated firms. In the meantime, therefore, alternative effective controls must be devised to prevent the abuses of economic power possessed by the great private corporations which will continue for some not inconsiderable time to exist.

In the field of international cartels the problems are similar. It is doubtful here if there are any instances where parallel anti-trust legislation by the various co-operating states would be an appropriate remedy. Indeed many cartels have the function of stabilizing supplies and prices over the period of the cycle. To break them up, and to leave it at that, would be to create a chaotic condition. Before this could advisedly be attempted international action to ameliorate cyclical fluctuations and an all-round tariff reduction would be necessary. In place of the cartel, however, the technique of inter-governmental commodity agreements could be substituted. It is true that the history of such agreements is not a happy or promising one. If they are to be freed of the cartel-like character they have previously had, two conditions must be fulfilled. The first is that the member states must become masters in their own houses, must achieve discipline and control of their great business corporations and so free themselves of their influence on public policy. The other condition is that consumer as well as producer states must be represented on the boards of control. There is no reason why the inter-governmental commodity agreements should not play a useful and stabilizing role in the modern world economy, and so replace the private cartel with an institution in which public rather than private interest would be the operative principle.

We have already suggested (Chapter XVI) the device of the inter-governmental public corporation. This device could be extended, as experience was gained, to operate in cartel areas, i.e., to take over the supply of certain key materials now privately supplied on a cartel basis. Such materials might include heavy chemicals, oil, steel and aluminium, and,

it is early yet to say, the whole organization might be brought under the atomic energy control commission as part of the international direction of the supply of key power and materials.

When a country which is itself trying to prevent the exploitation of its nationals by its own corporations finds itself the victim of an international cartel to which none of its nationals is a party, it can respond by punitive tariff action, by itself developing a source of supply under public ownership or by subsidy or by other retaliatory action against the cartel, such as, for example, prohibiting the sale to any member of the cartel of any raw materials or other commodities by any of its nationals.

There is finally the question of patent reform and compulsory licensing. As things stand, international patent agreements constitute an instrument of the cartels. Cartels assign patent rights as one way of dividing the world market and establishing a monopoly position for each member in the share of the market so assigned. Direct and cross-licensing of patents is one of the bargaining weapons in forcing agreements. Infringement suits and threats of infringement suits are used to intimidate possible competitors. Exclusive licensing and patent pools form a part of the armament of nearly every cartel. Patent control for the suppression of innovations dangerous to the cartel's monopoly is not unknown.[12]

The reform of patent law by international convention is perfectly possible and desirable, though it is not easy. It is not for the economist to enter lightly here, where competent lawyers fear to tread. There are certainly severe limitations on the power of any single country to effect reform on its own account. What may be done by international convention remains to be established.

[12] See *Canada and International Cartels*, Ch. III.

CHAPTER XVIII

CONCLUDING REMARKS

An historian friend who has read this manuscript says that all the important questions are neither discussed nor explored. He thinks that the whole matter, touched on in Part I, of the freedom of a society to decide and to act, the possibility of a community of interest emerging and being clearly formulated within the institutional framework of modern society, the kind of political institution that can reconcile efficient administration with parliamentary responsibility and civil freedom, the relation of international economic policy with established international political institutions—all these, he says, are the really important things touched on in this book, and none of them is given significant treatment. Perhaps he is right, and perhaps, some time, I should see if I can make clear the political assumptions and implications of what is intended to be, after all, a study in theoretic economics.

I do think, however, that exploratory work in the theory of economic change is one of the conditions of successful political adaptation and of economic decisions at the policy level. I do not expect a great deal of the theoretic structure of this work will stand, after my colleagues have brought it fully under the fire of criticism and empirical testing. All I hope is that it will serve to interest still more economists in these problems which lie so much closer to the level of policy decisions than do those deep strata that compose the branches of static theory. If these questions can be more adequately dealt with, the areas of decision more clearly defined, and the probable course of events flowing on alternative decisions more definitely known, the problems of policy and of appropriate political institutional adaptation will become that much easier. I have never doubted the importance of the work of the historian and the political scientist in analysing the working of political institutions, but I must insist that on the level of policy, economics

and politics merge, and the decision of action must be illuminated equally by both kinds of analysis.

Some time ago I read a novel of the thriller type entitled *When Worlds Collide.* (I am sorry to have forgotten the names of the authors.) It describes an imaginary cosmic catastrophe in which a planet from some other system wanders into the gravitational pull of our own sun, and, in the process of taking up its new orbit, crosses the orbit of the earth. The two planetary bodies collide, but fortunately the new planet is attended by a moon, of approximately the same size, specific density and atmosphere of the earth. When the two great planets collide and are mutually destroyed, the moon of the new planet is thrown off from its mother and takes up the orbit of the old earth around the sun. The astro-physicists of the world, who have long seen the disaster preparing, succeed in developing rocket ships, by means of which, at the critical moment, they transfer themselves and a few scientific friends, carefully selected, from the old, doomed world, to the moon, destined to succeed it in the economy of the firmament. They neglected, with the usual arrogance of natural scientists, to take any social scientists with them, and in consequence, rather bungled the economic and social organization of life on the new planet, but all that is not to the present point, and, in any case, provided the authors with abundant material for a sequel, called *After Worlds Collide.*

When the little group of scientists, having brought their rocket ship safely across the ether and down to their new earth, opened the doors and stepped out onto the land of the new home of our race, they witnessed in the heavens the final awful destruction of old Mother Earth and of all human life but them. They realized then that they stood alone in the universe. They, with their frail hands, the few tools and materials they had brought with them, their little knowledge and training, represented humanity, its destiny, its scheme of values. So far as they could know their survival meant the survival of all that was good in the universe, their extinction, the extinction of all that had meaning and significance. Man may infer from his scanty knowledge the existence of a higher power than

himself, of an immanent purpose and plan, of a conscious mind which conceives a value system of perfection. If he does, he creates It in his own image, for all he knows of good is human good, of purpose, human purpose, and he must, therefore, conclude that the survival of man is necessary to the fulfilment of the divine purpose and the realization of good. There is no evidence, known to man, to justify the belief that there is any moral significance apart from himself in the universe. If one believes in a system of values, one must will the survival and progress of the race. If one does not, neither survival nor progress matters, and only the blind instinct to live can prevent self-destruction.

Our scientists on the new world had none of these doubts. They believed that the race must survive and they had to set to work to make the most of their slim resources, to obtain immediate food and shelter, to provide for a continuous supply of food, to find mineral resources which would give them heat, power, and the materials for machinery. Each of the slender pool of productive resources had to be carefully allocated so as to obtain the maximum of production. Their first problem was, in crude and startlingly simple form, the problem of economics. Its successful solution was a condition of their survival.

Now, of course, except for the matter of numbers, their position was identical with ours. We are a group of newcomers to this universe, standing, as far as we know, alone on a little planet in space. We believe, or most of us believe, that our survival as a group is somehow important, for we possess the unique faculty of conscious choice, of creating good or evil. Our capacity for good is conditioned by our survival and our manner of survival. Today we possess, or shall very soon possess, for the first time, the power to destroy ourselves, to obliterate life altogether, and there are signs, too, of a great weakening in the conscious will to survive. We are tired, and tortured with self-disgust, and many must doubt if our so-called values have any significance apart from a kind of justification of creature appetites. So deep does the instinct for death run today, the "suicide will", that in many it has overcome the

instinct to live. Had Nazi Germany possessed in 1945 the secret of atomic power, but only in a form which, if once released would have set up uncontrollable chain reactions fatal to all life on the planet, who can doubt but that, in the awful moment when self-murder was decided upon, the high command would have given the order of universal destruction? Who can doubt, too, that within a short time that power of obliteration may be held by those who direct the affairs of the great nation states? In some future war, the possibility that a nation facing defeat may make the fatal decision is a real and terrible one.

It is easy to say that we need a revival of the liberal faith in humanistic values. But this is one of those facile, if true propositions, that burke the real issue. The liberal faith in humanity exists and can exist only in a certain social milieu, and our problem is really how this milieu is to be re-established. One condition of it seems to be the sense of economic security possessed by people, both as individuals within a community and as members of a national community in the greater community of the world as a whole, when they have a sense of sharing in a common progress, sharing in the work and responsibility the common effort requires, and sharing in the benefits of achievement.

I should not like to say that our survival is impossible without economic progress in this sense, nor that economic progress can ensure it. However, that progress I consider as contributing to survival, and it is for that reason I believe that inquiries into the nature and conditions of economic progress are important.

INDEX

INDEX

A

Abuse of rights, in English law, 35fn.
Africa, 343-344
"Agglomerating factors" 270-278
Albert, Ruth, 68, 76, 76fn., 89fn.
Alexandria, 327
Alice in Wonderland, 186
Allen, G. C., 332fn.
Allen, R. G. D., 55fn., 183fn., 210fn.
Aluminum Co. of America, Aluminium of Canada, Ltd., 362
America, United States of, 6, 13, 72, 143, 151fn., 154, 162, 165, 263fn., 274, 277, 281, 298fn., 333, 336, 337, 340-341, 343, 344, 347, 353, 355
American Department of Justice, 364, 364fn.
American Federation of Labour, 155
American Revolution, 4fn.
Amoroso, L., 183fn.
Angus, H. F., 276
Area of decision, 33, 34, 315-374
Aristotle, 16, 51fn.
Asia, 343-344
Australia, 336
Austrian school, 33, 34, 68, 94, 227fn.
Ayres, C. E., 11fn.

B

Bacon, F., 201
Baroque building, 133fn.
"Baubles", 255-257
Bell, E. A., 165, 165fn.
Bell, Spurgeon, 154fn., 155, 292fn., 298fn.
Bentham, J., 189fn.
Bergson, H. L., 53-55
Biscuit and confectionery industry, 291fn., 293-310
Bladen, V. W., 337fn.
Board of Inquiry into the Cost of Living (Canada), 165fn.
Bolton, 332
Boot and Shoe industry, 291fn., 293-310
Boulding, K. E., 45fn., 209fn., 250
Brady, A., 331, 331fn.
Bretton Woods Agreement, 346
British West Indies, 171, 277-278
"Bunching", of innovations, 137
Burns, A. R., 262, 264fn.
Bury, J. B., 4fn., 78fn.
Business cycle, and innovations, 144, 156-160; control of, 348-358

C

Calvin, 3
Cambridge, 94
Cameron, J. C., 334fn.
Canada, 150, 151fn., 162, 172fn., 243fn., 255, 260, 264fn., 333, 353, 355, 372; population increase of, 164-166, 333-335; Maritime provinces of, and location of industry in, 267-312; Quebec and Ontario industries compared with Maritime, 291-312; comparative early development of Ontario and Quebec, 277-290
Capital (*see also* Innovations) and innovations, 70-71, 145, 178-179; Smith's theory of, 70-75; Ricardo's theory of, 82-85; Marx's theory of, 86-91; Schumpeter's theory of, 95; deepening of, 92, 124, 135-136; widening of, 124fn., 135-136; thinning of, 124fn., 135-136; lay-out of, 231; international movements of, 144, 339-347
Carlson, Sune, 48, 48fn., 175fn., 210, 219fn., 221fn., 223fn., 224fn., 226, 226fn., 227, 227fn., 228
Cartels, policy, 359-374
Castings and Forgings industry, 291fn., 293-310
Causality, 11-26; and freedom, 27-37
Cause, concept of, 17-21; of economic change, 37; given, 63-68; endogenous, exogenous, external, internal, 63-66; selection of, 99-104
Census, of Canada, of Industry, *see* Dominion Bureau of Statistics
Chamberlin, E. H., 29fn., 39fn., 241fn.
Change, economic; nature of, 63-67; theories of, 68-104; criteria of, 99-104; models of, 105-264; and political consequences, 320-331; and the international order, 339-347
Charlottetown, 359fn.
Clark, J. B., 39, 42fn.
Commissioner of Combines (Canada), 241, 241fn., 264fn., 337fn., 361fn., 370, 370fn., 371, 374fn.
Communism, 321-322; communist party, 368

Comparative statics, as a method, 7
Competition, assumption of, 100-102, 176; degree of, 126, 126fn.; modifications by innovations, 240-253, 363-364; welfare implications of, 359-364. *See also* Monopoly
Condorcet, 77fn.
Confederation, economic effects of, 277-290
Coore, D. H., 176fn.
Corn, demand for, 199
Corn Laws, 78-79
Costs, decreasing, 28, 151; and innovations, 134
Cotton, demand for, 199
Cotton yarn and cloth industry, 291fn., 293-310
Cournot, A., 55fn., 192fn.
Coventry, 332
Creamer, D. B., 333fn.
Croce, B., 4fn.

D

Dalhousie University, 267
Davidson, J., 87, 87fn., 88
Day, defined, 114
Deepening, *see* Capital
Deglomerating factors, 270-278
Demand function, 166-174, 183-206; of sugar, 197-199; of corn, 199; of cotton, 199; of hay, 199-200; of potatoes, 200; elasticity of and measurement of, 196-200, 201
Descartes, R., 16fn.
Detroit, 333
Dickinson, L., 236fn.
Diminishing returns, 125, 125fn.
Dingwall, J., 47-49, 226-229, 229fn., 230fn., 234, 238fn.
Dobb, M., 182, 182fn.
Dominion Bureau of Statistics, 270, 277fn., 291fn., 292fn., 334fn.,
Douglas, Paul, 153fn.
Dunne, J. W., 53, 53fn.
Durham, Earl of, 277, 277fn.

E

Economies of scale, 236, 242-243, 274-276. *See also* Firm
Eddington, Sir Arthur, 52fn., 53, 54fn.
Edgeworth, F. Y., 40fn., 183fn.
Efficiency, and size, 292-293, 301-305. *See also* Firm
Elasticity, *see* Demand, Expectations, Income, Production
Enamel and Heating Products, Ltd., 288
Endogenous factors, 63-66
Engels, F., 3, 32
England, *see* United Kingdom
Envelope curve, 229-238
Equilibrium, concept of, 38-49; stationary, 94; limitations of in theory of the firm, 238; potential, 42, 250; of location, 269-290
Excess capacity, 126
Exogenous factors, 63-66
Expansion path, 48-49, 210-223
Expectations, elasticity of, 118-122, 127; and forecasting, 247, 253-254, and horizon, 256-258
Euler's theorem, 13, 174
Europe, 162, 182

F

Firm, theory of, 209-264; process analysis of, 210-228; criticisms of process analysis, 226-229; long-run costs of, and envelope curve, 229-238; temporal process of, 239-264; senility of, 261-264; effects of innovation on monopoly firm, 258-261; scale and efficiency, 274; size and efficiency in Canada, 292-293, 301-305; comparative growth of, in Canada, 293-310
Fiscal policy, 354-358
Fisher, I., 188
Florence, Sargant, 150fn., 242fn., 274, 274fn.
Flux, A. W., 174fn.
France, 133fn., 287
Fredericton, 278
Freight rates, in Canada, 288
Friedrich, C. J., 270fn.
Frisch, K., 48, 210, 219fn., 228

G

Gains, from a technical advance, 148-156
General Motors, 333
Germany, 13, 133fn., 274, 349, 367-368, 378
Gibson, 31fn.
Giffen paradox, 190
Godwin, 77fn.
Grand, A. T. K., 332fn.
Greer, Guy, 142fn.
Guthrie, J. A., 152fn.

H

Halifax, 278, 288
Hansen, A. H., 6, 33, 58, 71fn., 74fn., 78, 80, 90, 94, 97, 97fn., 100, 101, 129, 138fn., 140, 141fn.-142fn., 143, 144, 145, 161-166, 170, 171, 173fn., 178, 180, 181, 341, 342, 353
Harrod, R. F., 39fn., 57, 57fn., 134fn., 230fn., 231, 231fn., 232fn., 233fn., 236fn., 237fn., 363fn.
Hartman, N., 1fn.
Hawtrey, R. G., 124fn., 135
Hay, demand for, 199-200
Hayek, F. A., 330fn.
Heaton, G., 76
Hegel, F., 3, 85
Henderson, H. D., 144, 144fn.
Hicks, J. R., 55fn., 109fn., 110fn., 111fn., 121fn., 127, 127fn., 134fn., 167, 168, 170, 183, 183fn., 184, 185, 186fn., 189fn., 190, 198, 205
Higgins, B. H., 30fn., 343fn., 352fn., 355, 355fn., 356fn., 357, 358, 372
Hillman, H. C., 150fn., 274fn.
Historical view of time, 53-54
Historicism, 5
Hitler, A., 367
Hobbes, T., 340
Hobhouse, L. T., 31fn.
Hog cycle, 121fn.
Holland, 13, 73
Hook, S., 33fn.
Horizon, 256
Hume, D., 15, 18, 19, 21, 22
Hurd, W. B., 334fn.

I

Income, as a primitive function, 118-121; effect, 189-190; income changes and effect on demand, 204-205
Indeterminacy, 41, 43-44, 226-229, 251
Inertia, 115, 273
Inferior goods, 189-190
Innis, H. A., 3fn., 4fn., 267
Innovations, 70-71, 82-84, 90-92, 96-97, 132-160; as a cause of change, 99-104; capital-saving, 126, 134-138, 145, 178-179; strategic, 135fn.; adaptability of, 135-138; bunching of, 137; and business cycle, 144, 156-160, 350-351; and population, 161-182; effect on competition, 240-253, 363-364; adoption by monopolies, 258-261; and location of industry, 276
Instantaneousness, 48
Institutions, as a causal factor, 101, 178; affected by innovations, 148-156
International Paper Co., 362
Investment, intensive and extensive, 144; and cycle, 353. *See also* Capital
Irak, 344-345
Iran, 344-345
Iron and Steel industry, 291fn., 293-310
Isocost, 217-218
Isoproduct, 212-218

J

Jarrow, 332
Jeremiah, 28fn.
Jevons, W. S., 33, 34, 36, 55, 55fn., 67fn., 189fn.
Johnson, A. H., 76fn.
Johnson, W. E., 175fn.

K

Kaldor, N., 41, 41fn., 42, 209fn.
Kalecki, M., 154fn.
Kant, I., 13, 14, 14fn., 18, 19, 20, 52, 52fn.
Kaufman, F., 5fn.
Keirstead, B. S., 42fn., 150fn., 176fn., 279fn., 291, 311fn., 333fn., 336fn., 337fn., 338fn.
Keynes, J. M., 2, 6, 14, 15, 33, 57, 57fn., 58, 59, 82, 90, 98, 100, 110fn., 114, 118, 118fn.-121fn., 183, 184, 184fn., 317, 342, 348, 349, 352, 355
Kilgore Committee of the U.S. Senate, 364fn.
Kinked revenue curves, 251
Knight, F. H., 2, 2fn., 6fn., 33fn., 69fn., 183fn., 188, 188fn., 190, 190fn., 205, 205fn.
Kolthammer, F. W., 84fn.
Kondratieff, 137fn.

L

Länge, O., 134fn., 141fn., 184fn., 256fn., 260fn.
Lauderdale, 110fn.
Lausanne school, 167, 227fn.
Leibniz, 4fn., 20
Leontief, W., 191, 191fn., 192, 192fn.
Lerner, A. P., 209fn., 230fn., 236fn., 352fn.
Liberal Party, 366-367
Liberalism, basic concepts of, 31-32; modern position, 321-323
Linked advance, 137, 140-143

Lipson, E., 76fn.
Liquidity preference, 156-159
Littler, H. G., 58, 58fn., 118fn.-121fn., 157fn., 184fn., 349fn.
Location of Industry, 269-290
Locke, J., 31fn.
London, 332, 359
Long, C. D., 137fn.-138fn.
Lovejoy, A. O., 78fn.

M

Macdonald, Hon. A. L., 335, 335fn.
Macgregor, F. A., 241fn.
Mach, 21
MacIver, R. M., 3fn., 5fn., 16fn., 23fn.
MacKenzie, W. C., 279fn., 283fn.
Magna Carta, 365
Malaya, 345
Malthus, T., 32, 36, 77, 77fn., 78fn., 82, 84, 103, 110fn.
Manchuria, 345
Manifesto, communist, 85
Mann, Sir Horace, 75fn.
Mantoux, P., 76fn.
Marginal propensity to consume, 58, 118fn.-121fn.
Maritime Provinces of Canada, 267-312; economic development of, 277-290; location of industry in, 291-309; population transfers from, 334-335
Marshall, A., 38, 39, 39fn., 40, 41, 41fn., 42, 45, 46, 46fn., 55, 56, 56fn., 57, 65fn., 67fn., 166, 167, 168, 170, 176, 183, 184, 185, 187, 187fn., 188, 189, 192, 192fn., 196, 196fn., 199, 201, 209, 209fn., 230fn., 234, 235, 236fn., 249, 250, 253, 262
Marx, K., 2, 3-4, 32, 33, 66, 66fn., 70-71, 78; his theory of economic change, 85-94; 97, 100, 102, 180, 182
Measurement of demand, 191-206
Mill, J. S., 32, 83fn., 201, 201fn.
Mode, real, aggregative, defined, 109-111; in population change, 113-131; in technical change, 139-160; in population and technical change, 170-173, 177-181
Models, nature of, 107-112; population, 113-131; innovation, 139-160; mixed, 161-182; of the firm, 239-264
Monopoly, 126fn., monopolistic competition, 126fn., monopolistic surpluses, 98 (*see also* Rents); process of monopolistic firms, 241, with temporal maximization, 254-264; senility of, 261-264; innovations adopted by, 258-261; control of, 359-374; extent in Canada of, 360-362
Montreal, 282, 359
Moore, H. L., 191fn., 195, 195fn., 201
Moulton, H. G., 172, 172fn.
Myrdal, G., 6, 58, 58fn.

N

National Policy of 1879, 278-280
Nazis, rise of, 367-368
Nef, J. U., 4fn., 136fn.
New Brunswick, *see* Maritime Provinces
New Statesman and Nation, 344
Newcastle, 332
Newton, I., 136
Nitrogen cartel, 337fn.
Norway, 337fn.
Nova Scotia, *see* Maritime Provinces

O

Ontario, *see* Canada
Opie, R., 56fn.
Output curve in time, 248-249
Oxford, 332

P

Panama Canal, effect of on Maritime lumber trade, 282
Pantaleoni, M., 64fn.
Parameter, defined, 185
Pareto, W., 167, 168, 183, 183fn.
Paris, 359
Particular factors, 270-278
Patent law, and cartels, 374
Patents, life of and horizon of expectations, 256
Path, expansion, 210-228
Pearson, K., 15
Petty, W., 95
Phelps-Brown, E., 183fn.
Physics, notion of time in, 51-52
Pigou, A. C., 39fn., 43fn., 46, 46fn., 47, 65fn., 191fn., 201fn., 230fn., 317fn.
Planck, M., 21
Plato, 28fn.
Plummer, A., 332fn.
Popper, K., 5fn.
Population, treatment of by Smith, 72-75; by Ricardo and Malthus, 77-80; as a cause of change, 99-104; changes in, 113-131; and innovations, 161-182; transfer of, 332-339; international movement of, 339-340

Potash cartel, 263fn.
Potatoes, demand for, 200
Power, economic and political, 364-370
Prichard, H. A., 18fn., 20fn.
Prince Edward Island, *see* Maritime Provinces
Private Enterprise, 326-328, 354
Process, and equilibrium, 38-40; in time, 47. *See also* Firm
Production, function, 174-179; linearity of, 174-177; elasticity of, 175fn. *See also* Firm
Productivity, and wages, 300-310
Profits, 74-76, 80-81, 90-92, 93, 97-98; super, 92; temporal maximization of, 254-264
Progress, 74-75
Pulp and Paper industry, 150-154, 291fn., 293-310
Ptolemy, 136

Q

Quebec, *see* Canada
Queen Elizabeth, 13

R

Reciprocity Treaty of 1854, 277-282
Relations of causes, 101-102
Rents, dynamic, 176, 177fn.; and the business cycle, 348-358
Returns to scale, 125-126, 175-177
Reversibility, 250-251
Reynolds, L., 29fn., 264fn., 361fn.
Ricardo, D., 32, 66; his theory of economic change, 77-85; 86, 87, 88, 91, 100, 102, 103, 110fn., 125, 128, 129, 130, 180
Robbins, L., 39, 39fn., 42, 69fn., 209fn., 229fn.
Robertson, D. H., 73, 184fn., 209fn.
Robinson, E. A. G., 150fn., 231, 274
Robinson, Joan, 39, 39fn., 47fn., 88, 88fn., 134fn., 209fn., 227fn., 229fn., 230fn., 232fn., 235, 236fn., 251fn.
Rogers, N. McL., 279fn., 335fn.
Rosenstein-Rodan, P. N., 56fn.
Royal Commission on Price Spreads (Canada), 298fn., 355fn., 361fn.
Ruhr, 345
Russia, *see* U.S.S.R.

S

Saint John, 278, 288
Samuelson, P., 170
Say, J. B., 72, 81, 82, 84
Scale, *see* Economies, Size, Firm
Schultz, H., 42, 167, 168, 170, 183, 183fn., 187fn., 191, 191fn., 191-206
Schumpeter, J. A., 36, 40fn., 41fn., 44, 44fn., 46, 46fn., 63fn., 66, 66fn., 67, 71, 73, 75, 77, 77fn., 86; his theory of economic change, 94-98; 99, 100, 102, 103, 132, 133fn., 137, 162fn., 181
Scott, F. R., 331fn.
Secular stagnation, 142, 161-166, 341-347
Senility of firms, 141, 254, 261-264
Shipbuilding industry, in Maritime Provinces, 269, 277-278
Shove, G., 230fn.
Simultaneity, 48
Size of plant, measurement of, 292fn.; and efficiency, 292-293, 301-305; changes in Canada, 293-300. *See also* Firm
Slichter, S., 209fn., 250
Slutzky, E., 183fn., 190
Smith, A., 7, 12, 32; his theory of economic change, 69-77; 83, 84, 86, 94, 100, 101, 102, 142, 142fn., 143, 317, 327, 359, 359fn., 363fn.
Sorokin, P. A., 5fn., 25, 54fn., 60fn.
Soviet Union, *see* U.S.S.R.
Speenhamland, 32
Spengler, J. J., 78fn., 79fn., 82fn.
Spiethoff, 36, 137fn.
Spinoza, 109fn.
Sraffa, P., 230fn.
Stark, W., 31fn.
Steinbeck, J., 31fn.
Stocks, J. L., 21fn., 53fn.
Substitutability of factors, 211-215
Sugar, demand for, 197-199
Suggett, S. V., 191fn.
Surpluses, *see* Profits, Rents
Sweden, 172fn.
Sweezy, P., 86, 88

T

Technical advance, *see* Innovations
Temporary National Economic Committee of the 76th Congress (U.S.A.), 152fn., 242fn., 275fn., 301fn., 303, 360fn.
Thinning, *see* Capital
Tilley, Sir Leonard, 279fn.
Time, concept of in economics, 49-60
Timlin, M., 251fn.
Tinbergen, J., 256fn.

Tories, 78, 366-367
Toronto, 282
Townsend, Lord, 76
Toynbee, A., 18, 24fn.
"Trend", in demand, 171, 202-204
Triffin, R., 217fn.
Trust-busting, 364

U

United Kingdom (includes all reterences to England, Scotland, and Great Britain), 13, 23, 24, 70, 143, 274, 277-280, 281, 287, 312, 328-329, 331, 332, 337, 343, 344, 347, 353, 365-367, 371
United Nations, 344, 345, 346
United States, *see* America
Usher, A. P., 133fn., 135fn., 136fn., 137fn., 162fn.
U.S.S.R., 340, 341, 343, 344, 347
Utility, measurement of in theory of demand, 186-190

V

Vaihinger, H., 20fn., 52fn., 53fn.
Value productivity of wage dollar, 308
Veblen, T., 133fn.
Viner, J., 209fn., 230fn., 231fn., 237fn., 242fn.

W

Wages, and productivity, 300-310
Wages-fund, 83
Walker, E. R., 65fn., 94fn., 317fn.
Wallas, G., 31fn.
Walpole, H., 75fn.
Walras, L., 39, 44, 95, 97, 98, 167, 168
War of 1914-1918, effects of on location of Canadian industry, 280-285
War of 1939-1945, effects of on location of Canadian industry, 310-312
Wealth, progress in and innovations, 70
Welfare, affected by innovations, 148-156; economics of, 322, 323, 323fn.
Whigs, 78
Whitehead, A. N., 48, 52-55, 60fn.
Wicksteed, P. H., 3fn., 174fn.
Widening, *see* Capital
Wilkinson, Ellen, 337fn.
Woollen Yarn and Cloth industry, 291fn., 293-310
Working, E. J., 193fn.

Y

Young, Allyn, 191, 191fn.