UNIVERSAL
LIBRARY

OU_156515

UNIVERSAL
LIBRARY

OSMANIA UNIVERSITY LIBRARY

Call No. 338.5 P49p Accession No. P.G. 13701

Author Pettengill, R.B.

Title Price Economics

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

PRICE ECONOMICS

By

ROBERT B. PETTENGILL

DIRECTOR, TEACHING INSTITUTE OF ECONOMICS
UNIVERSITY OF SOUTHERN CALIFORNIA

THE RONALD PRESS COMPANY · NEW YORK

Copyright, 1948, by
THE RONALD PRESS COMPANY

Copyright, 1947, by
ROBERT B. PETTENGILL

All Rights Reserved

The text of this publication or any part thereof may not be reproduced in any manner whatsoever without permission in writing from the publisher.

PRINTED IN THE UNITED STATES OF AMERICA

PREFACE

In this college textbook in economics, attention is centered upon individual prices and quantities of goods and of services. Prices and quantities are considered both as causes and as effects of economic behavior. Because of this approach, the treatment of prices is broader and more thorough than that of books where price theory is only a subordinate theme.

The book deals with the fundamental mechanism of a free-enterprise system. It explains how prices affect demand and supply as well as how demand and supply affect prices. The author has tried to help the reader to understand that in economics there are always many causes and many effects, even though they must be studied one at a time. Each cause is the effect of some other cause, and each effect becomes a cause at the next step. Some economists would pursue the chain of causation further in certain directions than the author has done. Others would not go so far. The choice is a matter of personal preference, and instructors may add or subtract as they desire.

The author of a text in a well-worked field can make little claim to anything more than some originality of treatment. However, the present writer has tried to contribute something by his merger of the institutional and theoretical approaches to economics. Theories are developed from real life situations selected as typical, rather than from models which often give the impression of being unrealistic. Illustrations usually precede the statements of general principles rather than follow them. Most of the recent developments in price theory are used in one chapter or another, but with no pretense of omniscience. The only portions in which the author's approach may seem particularly novel are Chapters 11 and 12 with their functional approach to competition and monopoly, Chapter 22 on factor surpluses, and some sections of the chapters on interest and profits.

The treatment is nonmathematical except in a few geometrical notes which are not indispensable to the argument. Diagrams are included because of their convenience as a form of economic shorthand. In most cases they are used to illustrate, not to expound the argument, and they may be omitted if desired. The chapter on in-

difference curves may also be skipped by those who feel that its geometry is not essential to price theory.

This text will probably be found most useful in the second or intermediate course in economic theory. However, it is designed to be adaptable for use at different levels, depending upon the preference of the instructor. Some will want to employ the book in advanced courses for a review of price theory. Others who believe that a study of individual prices should precede a study of the economy as a whole may choose to use this volume in the principles course. The author has tried to define clearly every necessary concept. Students need not previously have acquired a vocabulary in the field. Wherever possible, words are used in their most popular connotations.

The author wishes to acknowledge the great help he has received from his students who have read different mimeographed versions of this book during its preparation. Their criticisms have prompted much rewriting to increase clarity and to improve exposition. The author is very much indebted to Dr. Anatol Murad of Rutgers University, who read and criticized the entire manuscript. The chapters on interest were particularly influenced by his arguments, although they remain the author's view, to which Dr. Murad would still take exception at several major points. Dr. W. H. Steiner of Brooklyn College and Dr. Kenneth Trefftzs of the University of Southern California also read these chapters and offered helpful comments. The chapters on wage theory have been improved by the criticisms of Dr. Spencer Pollard and Dr. Paul Prasow of the University of Southern California and Dr. Morrison Handsaker of Lafayette College. Helpful suggestions have been received also from Dr. Emmanuel Hacker of Brooklyn College, Dr. Armen Alchian of the University of California at Los Angeles, and Dr. Val Lehnberg of the University of Southern California.

Most of all the author is indebted to his wife, Margaret Miller Pettengill. Without her unfailing encouragement the author would still be trying to learn enough to write a definitive book. She insisted that he write what he knew and let others point its defects for subsequent revision. That is what he has done. He deeply regrets that she did not live to see in print the book which she did so much to bring into being.

ROBERT B. PETTENGILL

Los Angeles, Calif.
September, 1948

CONTENTS

PART I

The Prices of Commodities

CHAPTER PAGE

1 VIEWPOINTS IN ECONOMICS 3

1. Questions and Answers. 2. The Field of Economics. 3. Prices and Quantities from Various Viewpoints. 4. Problems in the Field of Causation. 5. The Vocabulary of Economics. 6. Purposes of Economists. 7. The Nature of Economic Principles. 8. Summary.

2 DEMAND AS A CAUSE OF PRICE 14

1. Fundamental Questions About Demand. 2. Two Ways of Looking at Demand: Market and Schedule. 3. Causes of Change in Schedule Demand. 4. Changes in Preferences Among Goods. 5. Cross-Elasticity of Demand. 6. Changes in Preference for Goods Instead of Currency. 7. Changes in Income. 8. Income-Elasticity of Demand. 9. Number of Buyers and Degree of Collusion Among Them. 10. Habit, Impulse, and Choice. 11. Summary of Principles Governing Demand as a Cause of Price.

3 TYPES OF DEMAND SCHEDULES 28

1. Demand in the Market Sense: The Law of Demand. 2. The Subjective Nature of Most Demand Schedules. 3. Statistical Demand Schedules. 4. The Relation Between Individual and Collective Demand Schedules. 5. Instantaneous Demand Schedules. 6. Successive Demand Schedules Should Indicate the Direction of Price Change, Its Starting Point, and Its Amplitude. 7. The Time Allowed for Buyer Response Should Be Indicated in Demand Schedules of the Successive Type. 8. A Classification of Demand Schedules into Short Run and Long Run. 9. Four Exceptions to the Law of Demand. 10. Summary.

4 PRICE-ELASTICITY OF DEMAND AND ITS MEASUREMENT . 39

1. The Price-Elasticity of Demand. 2. Ways of Presenting Demand Schedules: Emphasis on Total Revenue. 3. Demand-Elasticity, First Method: Change in Total Revenue. 4. Marginal Revenue as Change in Total Revenue. Note A: The Tangent Rule for Finding Marginal Revenue at Any Point on a Smooth Total Revenue Curve. 5. Average Revenue Is Another Way of Describing Unit Price. 6. Comparative Areas Under an Average Revenue Curve Also Reveal Changes in Total Revenue. Note B: Diagrammatic Comparison of Total, Marginal, and Average Revenue Curves. Note C: Point-Elasticity and Arc-Elasticity of Demand Shown by Use of Logarithmic Axes. 7. Demand-Elasticity, Second Method: Ratio between Per-

CHAPTER PAGE

centage Changes. Note D: Point-Elasticity When Demand Curves Are Plotted on Customary Arithmetical Axes. 8. Demand-Elasticity, Third Method: Ratio Between Segments of a Tangent to a Demand Curve. 9. Summary of Economic Principles Regarding the Effect of Price on Demand.

5 DEMAND ANALYSIS DEMONSTRATED BY INDIFFERENCE CURVES 57

1. Outlay Indifference Curves. 2. A Series of Outlay Indifference Curves. 3. Utility Indifference Curves. 4. A Series of Utility Indifference Curves. 5. Maximum Total Utility for a Given Total Outlay. 6. Income Consumption Curve: A Series of Tangency Points on Parallel Outlay Indifference Curves. 7. The Price Consumption Curve: A Series of Tangency Points on a Fan-Shaped Series of Outlay Indifference Curves. 8. Price Consumption Curves Become Demand Curves When Money Is Substituted for One Commodity. 9. Superior and Inferior Goods Revealed by One-Commodity Income-Consumption Curves. 10. Income Effect and Substitution Effect of Price Change. 11. Summary and Evaluation.

6 SUPPLY AND PRICE—GENERAL CONSIDERATIONS . . . 76

1. Statement of the Problem. 2. Supply as a Quantity Offered for Sale. 3. Supply as a Schedule or Series of Quantities. 4. Supply in the Schedule Sense and in the Market Sense. 5. Supply as a Price Asked by Sellers. 6. Supply Price as a Cause of Transaction Price. 7. Causes of Supply Prices. 8. Causes of Supply Quantities. 9. Summary.

7 INSTITUTIONAL DETERMINANTS OF SUPPLY PRICES . . 83

1. Supply Prices When Institutional Forces Dominate. 2. Motives Underlying the Determination of Supply Prices. 3. No Reservation Price. 4. Customary Price. 5. Prices Fixed by Law. 6. Resale Price Maintenance. 7. Imitative Price. 8. Markup Methods: (a) Markup for all Overhead and Profit Added to Direct Cost. 9. Markup Methods: (b) Markup for Administrative Overhead and Profit (after Cost Accounting). 10. Markup Methods: (c) When Producing to Order. 11. Summary of the Institutional Determinants of Supply Price.

8 SUPPLY PRICE: DERIVED FROM MAXIMUM PROFIT CALCULATIONS 93

1. Calculating the Maximum Profit Price Under Assumed Cost and Demand Conditions. 2. Seeking to Maximize Gross Revenue When Cost of Production Is Ignored. 3. Cost Schedules for the Individual Firm: (a) Total Cost. 4. Cost Schedules for the Individual Firm: (b) Average Cost. 5. Cost Schedules for the Individual Firm: (c) Marginal Cost. 6. Marginal Cost Usually Differs from Average Variable Cost. 7. Increasing and Decreasing Returns in Agriculture. 8. Transposing Diminishing Physical Returns into Increasing Physical Cost. 9. From Physical Cost to Money Cost. 10. Use of Cost and Revenue Curves in Setting Price. 11. Graphic Demonstrations of Profit-Maximizing Situations When Price Is Constant. 12. Graphic Demonstrations of Profit-Maximizing Situations When Price Varies Inversely with Quantity Demanded.

CHAPTER PAGE

13. Special Situations May Be Shown by the Shape of the Curves. 14. Some Observations on the Meaning of Cost Curves for the Firm. 15. Summary.

9 DETERMINANTS OF SELLING POLICIES 117

1. Statement of the Problem. 2. From Production Cost to Selling Cost. 3. Advertising and Other Forms of Selling Expense. 4. Product Differentiation Aids Price Control by Sellers. 5. Long-Run Perspectives in the Determination of Selling Variables. 6. Long-Run Effect of Price Changes upon Attitude of Buyers. 7. Effect of Price Changes upon Price Policies of Competitors. 8. Effect of Price Policies upon Entry or Exit of Competitors. 9. Effect of Price Policy upon Public Regulation. 10. Types of Price Discrimination. 11. The Basing Point System Reduces Location Advantages. 12. Other Aspects of the Basing Point System. 13. Requisites for Price Discrimination. 14. Limits of Price Discrimination. 15. Summary of Economic Principles Regarding Selling Policies.

10 PRODUCTION POLICIES INFLUENCE PRICES 140

1. Production Choices of Business Executives. 2. Deciding What to Produce and How Much. 3. Entrepreneurial Adaption to Shifts in Schedule Demand. 4. Deciding How to Produce: Choosing Proper Amounts of the Factors. 5. Deciding When to Produce: The Problem of Future Output. 6. Altering Plans for Future Production. 7. Equilibrium of the Firm, a Summary and Evaluation. 8. Normal Profit: Equilibrium of the Industry. 9. The Opportunity Cost Theory of Normal Profit. 10. The Bulk Line Cost Theory of the Normal Profit Price. 11. Other Uses of the Concept of Normal Profit. 12. Normal Price Causes Normal Profits and Holds Supply Constant. 13. Normal Price When Normal Profits Are Included in Cost. 14. Summary of Economic Principles Dealing with Production Policies Which Influence Prices.

11 A FUNCTIONAL APPROACH TO TYPES OF COMPETITION AND MONOPOLY 160

1. Difficulties with Substantive Classifications. 2. The Questions Asked in the Functional Approach. 3. Methods of Price Determination. 4. Ways of Competing Against Others. 5. Direct Methods of Competition. 6. Purchasing Competition. 7. Predatory Competition. 8. Indirect Methods of Competition. 9. Forms of Monopolistic Behavior. 10. Anti-Competitive Behavior. 11. Anti-Entry Behavior. 12. Anti-Buyer Behavior in General. 13. Four Types of Anti-Buyer Behavior by Monopolies. 14. Monopolies Charge "High" Prices and Make "Supernormal" Profits. 15. Summary.

12 CAUSES AND EFFECTS OF COMPETITIVE AND MONOPOLISTIC BEHAVIOR 176

1. Causes of Different Types of Market Behavior. 2. Causes of Independent Price-Setting by Sellers. 3. Causes of Collusion. 4. Causes of Price Acceptance. 5. A Diagrammatic Note on Traditional Demand and Supply Curves for Different Types of Competition. 6. Causes of Bargaining. 7. A Diagrammatic Note on Bargaining. 8. Bargaining and the Emergence of a Subjective "Surplus." 9. Parallel Classifications Are Possible on the Buying Side.

CHAPTER PAGE

10. Patterns of Price Behavior. 11. Causes of Price Stability. 12. A Functional View of Geographic Price Patterns. 13. A Social Appraisal of Competitive and Monopolistic Behavior. 14. Summary.

13 TIME FACTORS IN INDIVIDUAL PRICE TRENDS: COST CHANGES 192

1. Introduction. 2. Time-Elasticity of Supply. 3. Short-Run Supply Curves. 4. Moderately Long-Run Supply Curves. 5. Directional Supply-Elasticity. 6. Two Types of Moderately Long-Run Supply Curves: Industry Marginal Cost Curve or Successive Price Points. 7. Very Long-Run Supply Curves of Representative Firms. 8. Very Long-Run Supply Curves as Planning Curves. 9. The Apparent Inevitability of Monopoly Under Conditions of Decreasing Cost. 10. Summary of Possible Cost Reasons for the Downward Price Trend of a Commodity. 11. Summary of Possible Cost Reasons for the Upward Price Trend of a Commodity. 12. Summary of Economic Principles Governing Cost Changes Which Influence Price.

14 INDIVIDUAL PRICE FLUCTUATIONS 213

1. Introduction. 2. Effects of Falling Demand When Sellers Can Adjust Output but Must Accept Price. 3. Effects of Falling Demand When Sellers Can Adjust Both Output and Price. 4. Effects of Falling Demand on Price Through Firm Exit. 5. Upward Deviations from Normal Price. 6. Downward Deviations from Normal Price: The Influence of Durability. 7. Unexpected Supply Shifts May Cause Price Fluctuations. 8. A Summary of Supply-Side Causes of and Restraints on Individual Price Fluctuations. 9. Diagrammatics of the Time-Elasticity of Supply. 10. Summary of Economic Principles Dealing with Individual Price Fluctuations.

15 THE QUANTITATIVE APPROACH TO PRICE CHANGE . . . 229

1. Principles of Quantitative Economics. 2. The Prediction of Economic Effects or Changes Is Desirable, but It Is Difficult When There Are Many Variables. 3. Dynamic Determinants of the Price of Wool. 4. Quantitative Studies of the Elasticity of Demand. 5. Difficulties in the Quantitative Measurement of Economic Relationships. 6. Problems in the Application of Price Prediction Formulae. 7. Summary.

PART II

The Prices of Services

16 GENERAL PROBLEMS OF SERVICE PRICE THEORY . . . 241

1. The Prices of Services Resemble the Prices of Commodities. 2. There Are Many Subgroups in Each Factor Class. 3. The Factors of Production as Services and as Sources. 4. Use-Prices and Ownership-Prices. 5. Subjective Considerations in the Use-Prices of Factors. 6. Derived Demand for Factors. 7. General Aspects of Demand and Supply in Factor Markets. 8. Summary.

CHAPTER PAGE

17 SPECIFIC WAGE RATES: AN INSTITUTIONAL APPROACH . 249

1. Problems in Wage Theory, a Preview. 2. Topics to Be Studied by the Institutional Approach. 3. Labor and the Worker; the Factor Service and Its Source. 4. Particular Money Wage Rates May Be Set: I. By an Employer. 5. Particular Money Wage Rates May Be Set: II. By Bargaining. 6. Particular Money Wage Rates May Be Set: III. By the (Organized) Workers. 7. Particular Money Wage Rates May Be Set: IV. By the Government. 8. Three Methods of Attracting Additional Labor. 9. Overtime Work as an Alternative to Hiring Additional Workers. 10. Wage Raises Are a Last Resort. 11. Wage Increases Vary with the Time-Elasticity of the Supply of Labor. 12. Wage Increases Are Limited by the Intensity and Elasticity of the Demand for the Labor Product. 13. Wage Increases Are Limited by the Production Function. 14. When Should Individual Workers Demand Higher Pay? 15. Organized Labor Should Also Count the Cost of Wage Demands. 16. Time-Elasticity of Demand Following a Forced Wage Increase. 17. Workers May Accept Improved Conditions of Employment in Lieu of Wage Increases. 18. When Should Employers Try to Cut Wages? And Workers Accept? 19. Elasticity of Labor Supply Following a Forced Wage Decrease. 20. Summary of Forces Determining Specific Wage Rates.

18 EQUILIBRIUM WAGE RATES: SOME THEORIES APPRAISED . 269

1. Theory of Equilibrium Wage Rates. 2. Demand Equilibrium for One Type of Labor Hired by One Firm. 3. The Usefulness of Labor Demand Equilibrium Formulae Is Limited. 4. Collective Demand Equilibrium for One Type of Labor. 5. General Demand Equilibrium Considers Roundabout Effects. 6. Specific Wage Rates Influence the Supply of Labor. 7. Labor Supply Equilibrium for the Individual Worker. 8. There Are Many "Typical" Supply Curves for Individual Workers' Labor. 9. Labor Supply Equilibrium for Labor Groups. 10. The Collective Labor Supply Varies Directly with the Wage. 11. A Section with Negative Slope Appears in Some Collective Labor Supply Curves. 12. Different Types of Labor Often Receive Different Wage Rates. 13. Reasons for Differences in the Demand for Different Types of Labor. 14. Reasons for Differences in the Supply of Labor in Different Groups. 15. The Cost of Production Theory of Differential Wage Rates. 16. The Marginal Productivity Theory of Wage Rates: a Critical Appraisal. 17. The Influence of Skill on Wages. 18. The Bargaining Theory of Wages: a Summary and Appraisal. 19. Unions May Strengthen the Demand for Their Services by Opposing Substitutes and by Aiding Employers. 20. Wage Objectives of Union Labor May Vary. 21. Wage Differences among Members of a Given Vocational Group. 22. Summary of Equilibrium Wage Theories.

19 COMPARATIVE REAL WAGES 294

1. Statement of the Problem. 2. Definition of Terms and Statement of the Basic Data Needed. 3. Per Capita Productivity Changes with the Quantities of the Factors. 4. The Law of Proportionality and Some Applications. 5. Capital Accumulation Increases per Capita Productivity. 6. The Effects of Inventions upon Real Wages May Be Good or Bad, Depending upon Circumstances. 7. Other

CHAPTER PAGE

Effects of Inventions. 8. The Effects of Unions upon Real Wages. 9. Indirect Effects of Union Activity. 10. An Increase in Labor Supply Usually Reduces Real Wages. 11. The Dynamics of Real Wages—Statistical Investigations. 12. Differences Between per Capita Real Incomes in Different Countries at the Same Time. 13. Intercountry Differences in Living Standards Affected by War. 14. "Economic Parasites" Reduce the Level of Living of Those Who Work. 15. Large "Unearned Incomes," a Problem of the Unequal Distribution of Wealth. 16. Roundabout Effects of Differences in Income Distribution. 17. Summary of Economic Principles Explaining Differences in Real Wages.

20 WAGE THEORY AND UNEMPLOYMENT 319

1. The Wage Theory Approach to the Problem of Unemployment. 2. Unemployment as an Individual Problem. 3. When Product Demand Declines, Labor Demand Declines Also. 4. Wage Cuts Will Not Cure a Depression, May Make It Worse. 5. Wage Rates Promote Unemployment When Other Factors Become Relatively Cheaper Than Labor. 6. Unemployment May Be Caused by Wage Demands Being Too High. 7. Unemployment May Be Caused by Decreases in the Relative Prices of the Other Factors of Production. 8. A Decrease in Labor Efficiency Reduces the Demand for Labor Like an Increase in Pay. 9. Unemployment May Be Caused by Changes in Technology. 10. Unemployment Caused by Appraisal Price Being or Becoming Too Low. 11. Cyclical Unemployment Is the Worst Type. 12. Summary.

21 RENT 330

1. The Problem to Be Considered. 2. Alternative Definitions of "Rent." 3. The Peculiar Nature of Rentable Goods. 4. The Demand for Rentable Goods; Problems of Measuring Incremental Revenue. 5. Demand Appraisals by the Income Surplus Approach. 6. Other Ways of Determining Demand Bids for Rentable Goods. 7. The Determination of Lessor's Asking Rentals. 8. Vacancies Compared with Unemployment. 9. Rents Differ with the Relative Intensity of Use: The Product Surplus above the Intensive Margin. 10. Rents Differ with the Quality or Productivity of the Rented Good. 11. Differential Costs May Be Used Instead of Differential Product Surpluses. 12. Normal Rents: Long Run, a Supply-Side Concept. 13. Normal Rents: Short Run, as Amounts and as Percentages. 14. Rents Which Deviate from Short-Run Normal Cause Changes in the Demand for Rented Goods: Capitalization. 15. Other Causes of Changes in the Prices of Durable Goods. 16. Fluctuations in Rents Tend to Be Wider Than Fluctuations in the Prices of Nonrented Commodities. 17. Substitution and Voluntary Vacancies Reduce the Amplitude of Rent Fluctuations. 18. Royalties Are a Special Form of Rents. 19. Summary of Economic Principles in the Analysis of Rent as a Price.

22 SURPLUS ELEMENTS IN RENTS AND IN OTHER FACTOR INCOMES 354

1. Statement of the Problem. 2. Four Types of Surpluses May Appear in Factor Incomes. 3. The Lessor's Surplus (I) in Rent Incomes. 4. Residual Surpluses (II) in Rent Incomes. 5. Economic

CHAPTER PAGE

Surpluses (III) in Rent Incomes. 6. Economic Surpluses in the Rent of Land and Other Nonreproducible Goods. 7. Opportunity-Cost Surpluses (IV) May Arise in Rent Incomes from Long Leases. 8. Extending the Surplus Concepts to Wage Incomes. 9. Surplus Elements in Interest Incomes. 10. Surplus Elements in Profits. 11. Some Tax Problems Involve Consideration of Factor Income Surpluses. 12. The Taxation of Land and Improvements. 13. Taxes on Interest and Profits. 14. The Personal Income Tax Involves Surplus Problems. 15. Summary of Rent and Surplus Concepts. 16. The Relation Between Normal Rent, Interest, and Profits. 17. In One Sense Commodity Prices Determine Rents. 18. In Another Sense Rents Determine Commodity Prices. 19. Summary of the Economic Principles and Concepts Regarding Surplus Elements in Factor Incomes.

23 SPECIFIC INTEREST RATES: AN INSTITUTIONAL APPROACH . 371

1. Problems Connected with Interest. 2. The Borrower Seeks Improved Purchasing Power. 3. Collateral as Well as Interest Often Must Be Pledged by Borrowers Seeking Loans. 4. Lenders Demand Payment for the Service of Providing Funds Which They Create or Accumulate. 5. There Are Various Meanings of Interest as a Price. 6. Comparisons Between Interest on Loans and Rent on Capital Goods. 7. Funds Are One Form of "Capital." 8. There Are Four Major Classes of Borrowers and Lenders. 9. Business Borrowing. 10. Estimating the Worth of Funds to the Firm. 11. Banks May Supply Funds to Business Borrowers. 12. Interest Rates on New Loans Are Influenced by the Going Rate on Old Loans as Determined on the Security Exchanges. 13. Business Firms May Borrow from Individual Lenders. 14. The Proximate Supply of Funds Available to Government Borrowers. 15. Interest and Non-Interest Considerations in Government Borrowing. 16. The Proximate Supply of Funds Available to Consumer Borrowers. 17. Non-Interest Determinants of the Amount of Consumer Borrowing. 18. The Proximate Supply of Funds Available to Banker Borrowers. 19. The Ultimate Supply of Funds Offered by Lenders. 20. The Nature of Credit and How It Is "Created." 21. Credit Instruments and Their Uses. 22. Determinants of Interest Charges for Commercial Bank Credits, Most of Which Are "Created Credits." 23. Non-Cost Factors Influence Many Interest Rates of Commercial Banks. 24. Limits to the Quantity Commercial Banks Can Lend. 25. Individual and Business Savers of Funds May Lend, Hoard or Invest. 26. Descriptive Summary of Loans and Interest Rates. 27. Arbitrary Determinants of Interest Rates: Political Motives of Governments. 28. Summary.

24 DIFFERENCES AMONG INTEREST RATES 401

1. The Problems of This Chapter. 2. Differences among Interest Rates at the Same Time: I. Risk. 3. Differences among Interest Rates at the Same Time: II. Cost. 4. Differences among Interest Rates at the Same Time: III. Competition. 5. Differences among Interest Rates at the Same Time: IV. Demand. 6. The Interdependence of Interest Rates. 7. The Average Level of Interdependent Rates Will Be Explained Later. 8. Differencs in Interest Rates at Different Times: Demand Side. 9. Special Prob-

CHAPTER PAGE

lems of Changes in the Demand for Funds. 10. Differences in Interest Rates at Different Times: Supply Side. 11. Differences in Interest Rates in Different Places: Demand Side. 12. Differences in Interest Rates in Different Places: Supply Side. 13. Some Cases of Interregional Interest Differences. 14. Summary.

25 SOME PROBLEMS OF INTEREST THEORY 415

1. The Theoretical Approach. 2. Minimum Interest Rates Must Pay for Marginal Risk and Trouble. 3. The "Cost of Production" Theory of Interest. 4. Is There a Minimum Rate of "Pure Interest"? 5. The Waiting and Time Preference Theories of "Pure Interest." 6. Weaknesses of the Time Preference Theory of Interest Rates. 7. Liquidity Preference Theories of Supply and Demand for Funds. 8. Weaknesses of the Liquidity Preference Explanation of Interest. 9. The Interest-Elasticity of Supply of Funds: General Aspects. 10. The Interest-Elasticity of Supply of Funds from Commercial Banks. 11. The Interest-Elasticity of Funds Supplied by Other Lending Institutions. 12. The Interest-Elasticity of Funds Loaned by Private Savers. 13. The Interest-Elasticity of the Amount of Current Saving. 14. Saving Motivated by a Desire to Increase Income or Wealth. 15. Saving to Meet Contingencies or to Reduce Effort. 16. Collective Saving by Governments of Planned Economies. 17. Use of Funds to Increase Consumption in Distant Future vs. in Near Future. 18. The Interest-Elasticity of Demand for Funds in the Short Run When Rates Fall. 19. The Interest-Elasticity of Demand for Funds in the Short Run When Rates Rise. 20. Interest Rates in the Long Run. 21. Summary.

26 PROFITS: AN INSTITUTIONAL APPROACH 438

1. Statement of the Problem. 2. "Accounting Profits" Are the Chief Concern of This Chapter. 3. Accounting "Profits" May Be Negative, i.e., May Be Losses. 4. Problems Involved in Determining the Amount of Accounting Profits. 5. Valuation Problems Connected with the Rate of Profits. 6. Profits May Include Work Income as Well as Property Income. 7. Differences in Profit Rates Among Firms in the Same Industry. 8. Efficiency of Management as a Source of Profit Differences. 9. Efficiency Differences in Buying and in Selling. 10. Efficiency Differences in Production Cost or Output and in Planning. 11. Underpayment of Factors as a Source of Profits. 12. Differences in Capital Structure Affect Equity Profits. 13. Profits May Result from Lucky Events. 14. Monopoly as a Cause of Profits. 15. Owner Profits Compared with Firm Profits. 16. Changes in Profit Rate of a Firm from Year to Year. 17. The Role of Inventions in Causing Changes in the Rate of Profits. 18. Differences in Average Profit Rate Among Industries. 19. Industry Profit Differences Are Affected by Firm Expansion Programs. 20. Differences in the Average Profit Rate in Different Regions. 21. Summary.

27 PROFITS: SOME THEORETICAL AND SOCIAL PROBLEMS . . 460

1. Problems of Profit Theory. 2. The Profit-Elasticity of Supply. 3. The Profit-Elasticity of Individual Firm Capacity. 4. Firm Capacity Expansion When Profits Rise. 5. Firm Capacity Contraction When Profits Fall. 6. The Profit-Elasticity of Other Deter-

CHAPTER PAGE

minants of the Profit of the Firm. 7. The Profit-Elasticity of Supply of Firms: The Problems of Entry and Exit. 8. Choosing What Industry to Enter. 9. The Profit Causes and Conditions of Exit. 10. The Profit-Elasticity of Supply of Individual Equity Owners. 11. Other Alternatives Exist for Potential Takers of the Equity Position. 12. The Concept of Normal Profits in Equilibrium Theory. 13. A Definition of Normal Profits for Short-Run Situations. 14. Pure Profits as the Excess Above Business Profits (Opportunity-Costs). 15. Profits as Payment for Taking Risk. 16. The Motives of Equity Entry: Risks Deter and Profits Attract. 17. Tax Policy in Relation to Profit Theories. 18. General Social Objectives in Relation to Profits Taxation. 19. Possible Effects of Profits Taxation upon Changes in the Business Cycle. 20. Possible Effects of Profits Taxation upon Economic Progress and Individual Freedom. 21. Summary.

INDEX 479

PART I

THE PRICES OF COMMODITIES

Chapter 1

VIEWPOINTS IN ECONOMICS

1. Questions and Answers.—Any expository book in the field of a science must be a book of answers. It must tell the how and why, the when, where, and how much of various problems. But before answers can be given, questions must be asked. And since no single volume can answer all the questions that might be asked about a subject, the questions that are chosen should be selected with care. They should, if possible, spring from a *central theme* or follow logically from an announced viewpoint. If this is done properly, the reader will realize that there is more than one possible approach to the field. He will not fall into the delusion that the particular exposition at hand is the only one he needs to study. Every choice involves rejection as well as selection. The viewpoints which are slighted or omitted by one writer may be very important to another who starts with a different premise.

The following outline indicates the questions to which brief answers are given in the numbered sections of this first chapter:

I. What is the field of economics as treated in this book? (2)
II. What are the prices and quantities which economists study? (3)
III. How should one approach the problem of explaining prices? (4)
IV. What terms should economists use? The problem of definitions. (5)
V. Why study economics? (6)
VI. How use economic principles? (7)

2. The Field of Economics.—Surely one of the first questions to be asked and answered is that about *the scope* of economics. What does it include and what does it omit? The viewpoint of this book is that economics may be considered one of the social sciences since it deals with certain actions of individuals in their relations with one another. It is particularly concerned with those actions which relate to the production, distribution, exchange, and consumption of goods and services, but not with all such activity. It excludes the physical problems of production, transportation, and marketing in order to concentrate upon two things: prices and quantities of goods and

services in the market on the one hand, and maximizing net satisfactions on the other.

These two central themes are closely related. If, for instance, we suggest that economics studies the manner in which scarce resources are divided among alternative uses by people seeking to satisfy most adequately their manifold desires, we must recognize that prices and quantities are major guides. In modern capitalistic society individuals compare prices in their efforts to maximize personal incomes. Business men compare costs and selling prices in order to discover the best profit opportunities. Consumers are influenced by relative prices in deciding which things to buy to make their spending yield as much satisfaction as possible. Workers, landlords, and investors have services to sell and often choose one selling opportunity rather than another because the price offer is better.

For all these individuals the problem is not merely one of maximizing *gross* satisfactions. It is also, and more significantly, one of maximizing *net* satisfactions. This adds the problem of minimizing cost. Where the cost is a problem of buying cheaply, market prices are again the central consideration, but when we consider what must be given to get purchasing power, we often encounter the quantity problem, too. Workers, for instance, often must decide what quantity of their leisure and energy to give up to get the income which will buy what they want. The problem of maximizing *net* satisfactions includes, therefore, the problem of minimizing outgo.

The two themes might perhaps be combined in a single definition of the scope of economics if its implications are made clear. For instance, economics *as defined in this book* deals largely with market prices and market quantities, their magnitudes, changes, causes, and effects.[1] This should be understood to include among the causes, the maximum-net-satisfaction motive described above. The prices and quantities studied include those of both commodities and services. The actions of individuals relative to the market may be considered both separately and collectively. In the latter approach there are many problems of group choice. Some economists give their chief attention to questions of how to maximize the welfare of the group as a whole. Only a few welfare topics will be treated in the present volume. Its chief emphasis is upon the way in which prices and related quantities are determined in modern capitalistic society, together with the effects of such prices and quantities.

[1] This is the "correct" definition for this volume. Other authors prefer other definitions. The present author might use a different concept in another volume. See Section 5.

3. Prices and Quantities from Various Viewpoints.—The prices which economists study are chiefly transaction prices, the actual prices paid by buyers and realized by sellers in market transactions. To explain these prices and their fluctuations, economists usually introduce also the subjective prices which express the maximum amounts that individual buyers would pay or the minimum amounts that individual sellers would accept at any given time. These subjective prices exist in the minds of the buyers and sellers. They often differ from the actual market prices (transaction prices) at which goods are bought and sold. Subjective prices may be combined into demand and supply schedules, as will be explained in later chapters.

Demand and supply quantities also may be conceived in various ways. They may be the amounts actually exchanged, or the subjectively determined maximum quantities that buyers would purchase or sellers would sell at given prices. Another distinction is that between quantities as stocks and quantities as flows. Thus, there may be a certain quantity of wheat in storage at any given time (a stock). There are also certain quantities sold each day, week, or month (a flow). Quantities as stocks are absolute amounts. Quantities as flows are amounts per unit of time. They are often called "rates" of sale, of purchase, of production, etc.

Any rate of purchase is also a rate of spending. There is a total amount of purchasing power that changes hands during the given interval. This introduces the concept of the velocity of circulation of currency and brings together price theory and monetary theory. The income level is as important as the desire pattern of individuals in determining purchases of many goods. Changes in this level must be explained before price theory is complete.

The transactions which give rise to prices and quantities are transactions between individuals as buyers and other individuals as sellers. This is true whether individuals act independently as buyers or as sellers, or whether they are the agents of others. Even when a group takes action or delegates some one to act for the group, the individual members of that group make decisions before they vote or otherwise agree upon the desired course of action. Economic forces are therefore merely the market expressions of individual decisions to buy or not to buy, to sell or not to sell, to pay or ask a certain price, or to yield or seek a certain quantity. These decisions will be influenced by changes in prices or quantities that occur within the market itself and also by changes that occur elsewhere. The latter include technical inventions, natural calamities, the destruction of war, birth and death

rates, education, and a host of other changes which influence human judgments.

In many cases the decisions of individuals cancel out. At other times, like periods of panic, they reinforce one another. In either situation the cumulative effect of individual decisions is generally more important in economic problems than the separate judgments of single persons. There are occasional unique sales and purchases where an individual buyer meets an individual seller and they agree on the price of a discrete good. But more often there are many individuals on either the buying side or the selling side, or on both, and their collective impact on the market is the important thing in price-quantity determination, not the viewpoint of any one of them. Economics studies individual motives chiefly to understand group activity and therefore should include elements of social psychology in its field as well as individual psychology. For prediction it must, of course, rely upon the law of large numbers just like insurance companies which pool risks no one of which is predictable but whose total outcome may be very closely determined in advance.

4. Problems in the Field of Causation.—When economists set out to examine the price-determining forces in a given institutional setting, there is always the question of how far to go in the explanation. It is easy to begin by saying that prices are determined by supply and demand, but what causes supply and demand to be what they are? And what are the causes of these causes? How far back should one go in examining links of the chain? Here again, as in the scope of economics and its goals, economists differ. Some want to go back through market demand to an explanation of incomes; others seek chiefly to explain human choices.

To explain incomes we must explain the prices paid for the services sold by individuals in their capacities as workers, landlords, or capitalists.[2] A circularity in causation quickly appears. The business firm usually pays its wages, rent, and interest out of the revenue derived from selling its products and therefore the amount paid out is partly dependent upon the prices of the goods sold. In similar fashion it can be shown that wages, rent, and interest represent costs which are considered by managers when they calculate the selling prices they must ask. Therefore, if one follows this line of causation far enough back, he must perceive the basic interdependence of all prices and all quantities which are in any way variable.

[2] Part II of this book treats the prices of these services.

Some economists insist that this fundamental truth must be firmly established before simple cause-effect problems are considered. They formulate an extensive series of equations which express the relationships which are most common in modern economic society and then demonstrate by simultaneous solution that all variable prices and quantities are so interdependent that none can be singled out as the cause of another. This approach, however, involves much simplification of the complexities of real life. Its generalizations are rarely useful in prediction or control unless greatly qualified. The exposition typically moves from the general to the particular, from the abstract to the concrete.

The alternative approach adopted in this volume begins with specific situations and moves toward the general. This requires choosing from the multitude of different market problems those which are most common, or which are most significant because of their widespread effects. Classifications must be established which are easily described and readily understood. Small sections of the causation chain may thus be chopped off for examination. This is not completely realistic but it is helpful. Eventually these sections may be linked together and the whole dynamic picture revealed.

Returning to the second branch of market demand study, that of choices as opposed to incomes, we may note that some economists follow this lead on out into the fields of psychology, history, and social institutions. Here too there is some circularity, but chiefly one encounters topics which do not impinge directly upon market prices and quantities. There are problems of the motives of men, how they choose, whence they derive their scales of preference, and the influence of impulsive, nonrational decisions. Institutional economists point out how our statutory laws and our court decisions erect a framework within which we develop both our ideas and our market practices. The present volume will consider the elementary foundations of desire and choice, but must leave to others the questions of institutional development and psychological drives.

A similar line of argument might trace through the supply side the influence of cost as a primary determinant of the prices of the factors, such as labor and land. The reasons for supply quantities being what they are would also be considered, and the origin and influence of the state of technology which determines the efficiency with which they are applied. Obviously there are here a whole host of sidelines which one might explore if time permitted, but which must be left to those who write monographs for advanced students.

5. The Vocabulary of Economics.—Every science has its own vocabulary of words which are the time-saving shorthand of its writers. These technical terms are abstract summaries of principles and concepts developed in the analytic process. They are used both to expedite thinking and to convey ideas. Many of the physical sciences use terms freshly coined to describe their new ideas, but social scientists find word coining much more difficult. Economics, for instance, deals with problems which are part of man's everyday experience. Both the amateur and the expert talk about rent, interest, profits, value, debt, inflation, etc. For use in careful analysis, these terms need to be shorn of their fuzzy edges and narrowed down to explicit definitions. Sometimes, however, in his efforts to sharpen these concepts into useful tools the economist has reshaped them so much that he cannot communicate his ideas to the people whom he would instruct. The inner circle can understand one another, but their language is often unintelligible to the outsider or the novice.

Therefore, a volume like this one, directed at those who have had only elementary instruction in economic theory, must be guided by two considerations in matters of terminology. It must define terms in such a manner that they can be understood by beginners and it must also bow in the direction of the esoteric so that students will be prepared for the jargon of the initiates when they reach higher levels.

Despite the dogmatic attitude of many writers, there is no single right meaning for terms used in economics nor for "economics" itself. There are only meanings which are more or less useful in analysis and communication. Some persons prefer one meaning, some another. A choice can be made by a third party on the grounds that to him a given meaning helps him in his objectives more than a different meaning. Take the term *capital* for instance. Some define it as goods used in producing other goods; others limit it to man-made production goods; while another group defines it as liquid purchasing power. Which of these concepts is "right"? None. Which then, should an author use? The one which serves his purpose best, regardless of whether this aligns him with the majority or the minority, or leaves him a maverick off by himself. If he does depart from customary meanings, however, it is incumbent upon him to make very clear both why he does it and exactly what the term means in his own vocabulary.

The present volume makes extensive use of adjectives and phrases in an effort to descend the abstraction ladder far enough to distinguish between the various viewpoints. Often all of them are useful and should be incorporated into the body of economic thought.

This is generally preferable to isolating one viewpoint or meaning and proclaiming loudly that this is *the* one which all persons should adopt. One of the important contributions of economics is its explanation and clarification of the relationships between the various aspects, viewpoints, or meanings of general terms which are often misused by people just because they are too abstract.

On the other hand, some concepts or definitions are clearly harmful. They either obstruct thought, preventing it from being as fruitful as it should be, or they mislead the thinker toward false and dangerous conclusions. Consider, for instance, the idea that the term *inflation* should always be associated with the printing of government notes in large quantities. This concept may obscure other possible causes of rising prices, such as an increase in the volume of bank credit, a rise in the velocity of circulation of currency, or even a decrease in the volume of production. It may also lead to the conclusion that governments should under no circumstances resort to the printing press to meet their needs for purchasing power, but should always tax or sell bonds. Although this latter argument is probably valid in most cases, exceptions are conceivable and must not be overlooked because one has accepted as final a given definition of the term.

6. Purposes of Economists.—This leads to the fundamental question of the goal of economists. Why study the science? "The proper study of mankind is man." Agreed, but for what reason or objective? Several goals have been suggested such as to understand the past including how we reached our present state, or to predict the future and how to adapt ourselves to it or to control it.

First, we may seek to understand what happened in the past or what is happening now. This involves learning facts about past events, arranging them into what seem to be logical cause-and-effect sequences, developing hypotheses, and thus proceeding by the scientific method to tested generalizations, principles, or laws. Such are the methods of the analyst who tries to answer the great group of "Why?" questions. The pure scientist seeks truth for its own sake. If past explanations seem inadequate, he tries to make them more complete; if they are contradictory, he tries to reconcile or choose between them; if entirely lacking, he poses new questions and seeks answers.

A second objective of economists is to learn enough about the workings of our economic system so as to be able to predict future developments. This problem may be approached with the detachment of the pure scientist who may, for instance, predict the date and

extent of the next eclipse of the moon. But more often the individual who is concerned about the future wants to know what is going to happen so that he can adjust his actions accordingly. He may want to do that which will prevent a future event, like a depression, from doing him harm. Or he may anticipate a future change by action which makes it bring him more benefit than would otherwise accrue. Such individuals are concerned with the usefulness of knowledge. They tend to narrow the scope of their questioning to "practical" topics, often to the extent of ignoring fundamental problems of very great, but unperceived, relevance. They are often impatient with attempts to discover elementary principles when applications are not immediately apparent. This book is not for the shortsighted.

A third possible objective of those who study economics is the control of future events. This is based upon the second objective and through it upon the first. One must analyze before he can predict, and he must predict before he can plan how to control. Control differs from prediction and adaptation in that it assumes the course of events may be changed. Implicit in this objective is the desire, of course, to make those events turn out "better" than they would otherwise do. The focus of this desire need not be limited to the individual, but may be extended to the group. This is the field of the economic planner. He is concerned with the welfare of his group and seeks to promote it by applying pressures from outside the economic system to make it yield those price and quantity changes which he believes beneficial to his group. At times this means advancing one group at the expense of others, as through raids on the public treasury, striking for higher wages, or even international war. But it may also involve such goals as the welfare of the nation as a whole or even the entire world, unsemantic though these objectives may be.

In short the purposes of economists differ according to their scales of values, their vocational problems, their interests of the moment, or their endocrine balances. The present volume will concentrate upon the first of the three objectives described above, not because the author is uninterested in group welfare, but because analysis must precede attempts at synthesis. The person who attempts to say what laws should be passed, what ideas transmitted through education, or what emotions aroused should be thoroughly grounded in the principles of economic, political, and social analysis if he is to be more than a mere demagogue. Analysis therefore is given more emphasis in this volume than synthesis, although it seems neither possible nor desirable to exclude entirely the fascinating problems of applied economics.

7. The Nature of Economic Principles.—The principles of economics derived from the analytical methods of pure science are generalizations which purport to explain the effects which follow from given causes under certain conditions. Since it is often impracticable to describe all of the essential conditions when stating a principle, a generalization is usually intended to express a tendency that will work out if nothing interposes to prevent it. For instance, the first law of price is often stated as follows: "The price of a good varies directly with the demand for that good." This does not mean that in every case when the demand for a good rises the price of that good will also rise, but that the price will *tend* to rise if an increase in supply or some other change does not offset the effect of the increased demand. The principle merely expresses a cause-effect relationship, or in mathematical terminology, a function. If the relationship is very common, and particularly if obstacles to the indicated result rarely appear at the same time as the determining cause, then the principle is often called a theory or a law.

When individuals seek to apply economic principles to explain, to predict, or to control, they should at the outset recognize that the principles are definitely generalizations, not absolutes. That is, principles state what may be expected to follow a given cause most of the time, not all the time. The most useful principles are those which are true more often than, say, 90 per cent of the time. However, a 60 to 70 per cent generalization may sometimes be raised to the 90 to 95 per cent level by stating carefully the conditions under which it operates. These include particularly an enumeration of the other causes or variables which might influence the effect of the given cause and stipulating that they must not change. This is the familiar *ceteris paribus,* or "other things equal," qualification which hedges the application of many scientific laws. But one should remember that the law is valid even if other things do change to resist its effect. Thus the force of gravity operates upon an object even though impeded by a plane which prevents the object from falling, by friction which slows its rate of drop, or by an air blast or magnetic field which makes the object move in an upward direction.

In addition to a stated or implied *ceteris paribus* qualification, the statement of conditions should include several other things which will guide the person trying to apply an economic principle. In the first place, the institutional framework should be described. The first law of price described above, for instance, will hold only when sales are made under conditions where the supply quantity varies directly with the price. If there were individual, collusive, or gov-

ernmental price-fixing, the results would not be as indicated. An increase in schedule demand would bring a change in the volume sold, but no change in the price.

In the second place, the result may depend upon the length of time allowed for a given cause to have its possible effects. Thus, an increased demand may raise the price in the short run, but reduce it in the long run when sellers realize lower costs through economies of large-scale production. Therefore, the time condition should be made explicit in most cases.

In the third place, the direction and the rate of change of the independent variable (the cause) will often be important in determining the change in the dependent variable (the effect). A sudden and sharp rise in demand will ordinarily raise the price further than the same increase occurring over a period of months. Or a 2 per cent per month rise in demand may change the price more than a 2 per cent per month drop.

This leads to a final comment about the applicability of economic principles. In their simplest form they usually express merely the fact of a cause-effect relationship and whether the changes in each occur in the same direction or in opposite directions. For instance, if demand goes up prices tend to go up; if supply rises, prices fall. In general terms, prices are a direct function of demand and an inverse function of supply. These principles become even more useful if the amplitude of the change is indicated in addition to its direction. This involves much more careful study of past records than that required to develop simple principles. Statistical materials are often inadequate, and many subjects are not adapted to statistical treatment. Often the most that economists can say is that in certain described cases the results are more pronounced than in others. For instance, the demand for bananas usually declines more after a 10 per cent price rise than does the demand for salt under similar circumstances. This is one of the elasticity problems which will be examined for both demand and supply in later chapters.

8. Summary.—The viewpoint of this volume is basically that of one living in a modified capitalistic economy like that of the United States after the second world war. In such an institutional setting, the writer seeks answers to questions about the causes and effects of price relationships and price changes. This involves also a consideration of related quantities. Proximate causes and near-by effects are given more attention than more distant ones, although the problems of interdependence are not overlooked. Individuals are seen making

the choices which determine the trends and magnitudes of economic quantities. They also determine whether economic principles shall be valued merely for their own sake or whether they shall be used in prediction and control. Analysis is the main theme, but synthesis and applications are not overlooked.

Part I deals chiefly with the prices of tangible goods. Part II concentrates on the price problems connected with the services rendered by the factors of production.

Chapter 2

DEMAND AS A CAUSE OF PRICE

1. Fundamental Questions About Demand.—Since Part I of this book endeavors to explain chiefly the prices of goods and related quantities, it is appropriate to begin with a study of the basic market forces. These are popularly described as "demand and supply," but they take many forms and their analysis from various viewpoints will occupy most of the present volume. Demand will be treated first, somewhat by arbitrary choice, but also because most production takes place to satisfy known or expected demand.

There are two fundamental questions about demand from which others spring. The first is, "How does price affect demand?" This will be the topic of the next chapter. The second is, "How does demand affect price?" This leads to questions about what causes the changes in demand which in turn cause the changes in price. It is axiomatic that an increase in demand tends to cause an increase in price, but the causes of changes in demand are not so easily understood. The remainder of this chapter will be devoted to classifying and explaining these causes.

When demand changes are considered to be causes, price changes are the effects. There are several other ways of expressing the same idea which are used from time to time as they seem more appropriate. For instance, one may say that price is a function of demand. Or demand may be called the independent variable and price the dependent variable. One might say that price is determined by demand if this did not seem to make demand the sole determinant. Supply is always a second determinant and there are others such as government price-fixing and custom.

The alternative approach to demand is to consider it as an effect of price. Here price is the independent variable and demand the dependent variable. The other phrases may be reversed in similar fashion. The problem is sometimes said to be one of the elasticity of demand, as explained in detail in the next chapter.

The specific questions to be answered in this chapter may be outlined as follows:

I. How may demand as a cause of price be distinguished from demand as an effect of price?
II. What are the causes of change in schedule demand?
  A. Changes in buyer preference among goods; advertising, technology, and cross-elasticity of demand
  B. Changes in buyer preference for goods instead of currency; liquidity preference and propensity to consume
  C. Changes in income; income-elasticity of demand
  D. Changes in the number of buyers and in the degree of collusion among them
III. What part do habit, impulse, and choice play in determining schedule demand?

2. Two Ways of Looking at Demand: Market and Schedule.—When we say that a change in demand changes price, we are speaking of a change in "schedule demand." On the other hand, when we speak of a price change causing a demand change, we refer to "market demand." The difference between these two concepts is fundamental. In the first, "demand" is a cause. In the second, "demand" is an effect. Many problems of price analysis are simplified by remembering the differences between these two ways of looking at demand.

The term *demand* itself, when used without any qualifying adjective, usually refers to a quantity, but may refer to a price. The viewpoint is that of the buyer. When adjectives are used, demand quantities become either "market demand" or "schedule demand." Demand prices in similar fashion are called either "market prices" or "schedule prices." The following outline shows the various viewpoints:

1. Demand as a quantity
  A. The quantity *taken* by buyer(s) in *a given transaction* ..... "market demand"
  B. The quantity that *would be taken* by buyer(s) at *any one price* ..... "demand quantity"
  C. The quantities that *would be taken* by buyer(s) at *a series of prices* ..... "schedule demand"
2. Demand as a price
  A. The *price paid* by buyer(s) in *a given transaction* ..... "market price"
  B. The price that *would be paid* by buyer(s) for *any given quantity* ..... "demand price"
  C. The prices that *would be paid* by buyer(s) for *a series of quantities* ..... "schedule prices"
3. Demand as a schedule
  A. A series of price-quantity pairs showing both schedule demand and schedule prices "demand schedule"

The different ways of looking at demand may be illustrated further by the use of numbers and diagrams. For instance, robust John Smith's demand schedule for hamburgers to eat for his noon lunch on a given day may be as follows:

If the price is	The quantity purchased will be
10¢	3 hamburgers
15¢	2 hamburgers
20¢	1 hamburger

Smith's actual market demand cannot be stated until we know the price he has to pay for his hamburgers. At 10 cents, Smith will buy 3, at 15 cents, he will buy 2, etc. In other words, *changes* in the market demand quantity may be considered an *effect* of price change. They will be discussed in the next chapter.

On the other hand, *changes* in the quantity demanded *at any one price* cannot be due to price change since the price is assumed to be held constant. Instead they are due to changed desires, changed incomes, etc., as explained more fully below. And when there does occur a change in the quantity demanded at any one price, it is likely to occur at other prices too. The whole schedule may change as illustrated by the following demand schedule which may reveal the effects of an increase in Smith's wages:

JOHN SMITH'S DEMAND FOR HAMBURGERS AFTER HIS WAGES ARE RAISED

Price	*Quantity*	
10¢	4 hamburgers	(Each quantity is
15¢	3 hamburgers	larger than before
20¢	2 hamburgers	at that price.)

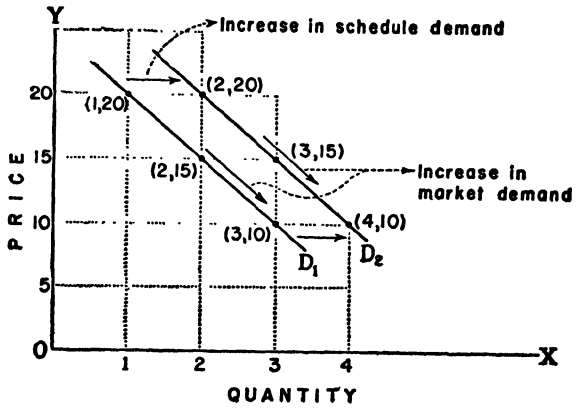
These two ideas of a change in demand in the market sense and a change in demand in the schedule sense may be illustrated by a diagram. In Figure 1 the changes in market demand are shown by moving up and down either curve. An increase in demand in the schedule sense may be seen in the fact that the second demand curve, D_2, is further to the right than the first, D_1. Each quantity that D_2 represents is greater than the quantity at the same price in D_1. (The paired figures in parentheses give the quantity-price amounts which appear in each schedule and when plotted on quantity-price axes give the points of the two demand curves.[1]

[1] The phrase "price-quantity" is more common than "quantity-price" but mathematical usage requires that quantity precede price in graphs.

The contrast between the two ways of looking at demand may be summarized once more. Changes in market demand are caused by price changes. Changes in schedule demand are a cause of price changes. When referring to demand changes in the market sense, price change is a cause; when using the schedule sense, price change

FIGURE 1

SCHEDULE DEMAND AND MARKET DEMAND



is an effect. The former topic will be examined in the next chapter dealing with the problems of the price-elasticity of demand. The present chapter will explore some of the reasons for the changes in schedule demand which tend to cause changes in price. At the same time it will cast light upon the absolute size of the quantities that exist in any given demand schedule. This is a different question from that of the changes in the quantities that occur from time to time.

3. Causes of Change in Schedule Demand.—It is an axiom of elementary price theory that an increase in the intensity of demand for a good tends to bring an increase in its price. More formally, the First Law of Price states that price (or quantity exchanged) varies directly with the demand. The amount of the variation, and whether it is price that changes, or quantity, or both, depends also upon the conditions surrounding the supply of the commodity, a topic the discussion of which must be deferred to a later chapter. At the present juncture, the question to be answered is, "What causes lie behind demand as a cause?" We cannot explore the causation chain very many links backward from the market situation, but some at least should be examined.

There are three main causes of changes in demand intensity: changes (1) in preference among goods, (2) in preference for goods versus currency, and (3) in income. Each of these causes may be approached from the viewpoint of a single individual, but they are most significant when they hold for a group, since changing attitudes of individuals often cancel out against opposite changes in other individuals.

4. Changes in Preferences Among Goods.—The intensity of demand for a given good, X, will depend in the first place upon how people value it for personal consumption or use in business as compared to other goods, Y, Z, etc. The general appraisal patterns of the members of a group are derived from the culture of that group. If they live in a highly industrialized civilization, they may prefer factory products to those that are hand-made, and speedy methods of transportation to slow ones. If it is a capitalistic economy with stress upon individual competition and the acquisition of wealth and income, people will have a scale of values in which high rank is given to consumption goods which help one to "keep up with the Joneses," or go them one better, such as fine clothing, new automobiles, large homes, etc. Income-yielding securities will rank high in the preference scale of things to buy. Or, to consider more specific desires, if smoking and drinking are prevalent culture traits, people in general will demand tobacco, liquor, and all the paraphernalia that go with these habits.

Changes in the demand for particular products within a given culture depend upon changes in the relative desires of people. Consumer demand, for instance, may be changed by advertising which makes people believe that they will be better satisfied, happier, more envied, or more successful in love, if they buy more of X than they have in the past. The slogans are varied, but they all seek to raise a given good or type of product in popular estimation so that the intensity of demand for it will increase. Ideas of desirability will also change because of education, the effect of motion pictures, the impact of religion, or the fervor of war. Certain seasonal changes in relative preferences are obviously due to external changes in the weather. Other illustrations might be found in legislation which forbids a given industry, like the Prohibition Amendment, or subsidizes it, like aids to cotton growers or shipbuilders.

In the demand for products that comes from business managers who want the goods not for personal consumption but for use in the production process, there are certain additional factors. If consumer

fancy happens to turn toward a particular fabric, garment, hat, or form of entertainment, the demand for goods needed to supply these whims will rise under the impulse of the profit motive. More important in scope and in human history, however, has been technological change. The effects are both direct and indirect. When a new process is invented or a new product devised to meet a long-felt need, businessmen increase their demand for the things needed to produce the product or install the new process. At the same time they may decrease their demand for other things no longer so profitable to use as before. Thus the demand for capital goods may rise at the expense of labor in cases where labor-saving machinery is being introduced, internal combustion engines may replace steam engines, structural steel may take the place of stone, and in a host of other ways technological change may modify producer demand schedules.

The indirect effects of changing technology upon total demand are no less important than the direct, though they may take longer to become apparent. The industrial revolution accelerated the development of a capitalist class on the one hand and a laboring class on the other. This brought increasing demand for luxury goods for leisure-class living. As the ownership of tools shifted from the artisans to the business unit and as the scale of operations increased, the type of tools and machines demanded changed also. The growth of factory and commercial cities brought new ways of living and new desires. Laws were changed to aid the enterprise at first and later to protect the proletariat. In short, the entire culture pattern of modern industrialized countries has been largely shaped by a series of innovations in various fields of production, including agriculture, mining, transportation, etc., in addition to manufacturing. And this culture pattern is the basic cause of the scale of values which most people adopt relative to things to be purchased on the market. It influences prevailing demand intensities, and as it changes, they change. Through these links in the causation chain, technological changes influence the prices and quantities that prevail in the marketplace.

5. Cross-Elasticity of Demand.—The final point to be discussed under the general heading of things causing changes in relative preferences among goods is the prices of other goods. This is sometimes called the *cross-elasticity* of demand and will receive further elaboration in the next chapter on price-elasticity of demand. But at this point it is desirable to point out that a rise in the price of one good may cause a decrease in the demand for it and drive buyers to purchase instead some substitute good. Thus if the price of oranges goes

up, some consumers may turn to grapefruit for their vitamins. But this increase in the demand for a substitute good will tend to increase its price, too. On the other hand, if two goods are complements instead of substitutes, the influence on schedule demand will be reversed. Thus if the price of gasoline is driven upward by the imposition of a large tax, people will tend to buy less of it and therefore will demand fewer automobile tires. The appropriate generalization is as follows: changes in the price of A tend to have a *direct* effect upon the schedule demand for B (and thus upon B's price) if B is a substitute for A. If B is a complement, the effect is *inverse.*

The two types of cross-elasticity are illustrated by the following hypothetical case:

CROSS-ELASTICITY OF DEMAND FOR SUBSTITUTES AND COMPLEMENTS

Price of A (the independent variable)	Demand for B, a SUBSTITUTE for A (price of B unchanged)	Demand for C, a COMPLEMENT of A (price of C unchanged)	*For reference only:* Demand for A as Its Price Rises
6¢	1 unit	8 units	4 units
↓			
7¢	2 "	6 "	3 "
↓			
8¢	3 "	4 "	2 "

6. Changes in Preference for Goods Instead of Currency.—The demand for consumption goods will diminish if people choose to save more out of current income. People who saved only $10 out of a monthly income of $300, for instance, when business conditions were improving, may later become apprehensive about the future. They decide to save more seriously for a possible "rainy day" ahead. If monthly savings are then increased to $30 or $50, spending is reduced accordingly. These people are said to reveal a decreased *propensity to consume.* Their scales of preference have changed so that cash in hand or in bank accounts seems preferable to certain goods. Demand slackens first for luxury-type products and for durable goods whose useful life may be prolonged by repair. On the other hand, if people get the idea that prices are going to rise, they may spend a higher percentage of their incomes and save less. They have an increased propensity to consume. This change in consumer preferences is most clearly revealed during those periods of rapidly rising prices popularly called inflation.

The demand for income-producing goods, including securities, may also change because of a change in the relative desirability of

goods as compared with currency. This is usually called a change in *liquidity preference.* The effect is more upon the disposition of past savings than upon the amount of saving out of current income. Thus, if a man has saved $10,000 and invested it in stock, he may get the idea that the stock market is going to fall and that he should protect himself against that loss. He sells his stock and puts the proceeds in the bank, while he waits for the day when he can reverse the process and buy again with profit to himself. At that time his attitude changes a second time. His liquidity preference decreases, and he chooses to hold his savings in the form of stock instead of a bank deposit.[2]

Occasionally there are people who invest regularly the amount saved each month out of income. For these persons, a decision to spend less and save more may influence two markets at the same time. The decreased propensity to consume may diminish the demand for consumer goods. The increased savings will raise the current demand for income-producing goods. But if at the same time that these people decide to save more they also get more cautious in their investing, they may reveal an increased liquidity preference by holding current savings for a longer time before investing. This will diminish the current demand for income-producing goods at the same time that the demand for consumption goods is falling. If the demand for both types of goods changes in the *same* direction, that means liquidity preference and propensity to consume have changed in *opposite* directions. This is a frequent but not inevitable relationship. During most of the downswing of the business cycle, both demands are decreasing, and during most of the upswing both demands are increasing, but at the turning points one change precedes the other.

7. Changes in Income.—One of the most important causes of shifts in the demand curve is changes in income. This was illustrated above in the effect of a wage increase upon John Smith's demand for hamburgers. Rising incomes, however, do not affect all rates of purchase to the same degree. Smith may buy more hamburgers for lunch, but no more coffee for breakfast. Because he eats a bigger lunch, he may not want as much for supper or may want a different type of food. If, when his wages were low, his family had to eat chiefly the cheaper foods such as beans, his wage increase may cause the family to decrease their buying of beans as they turn to meat

---

[2] Currency hoards could be substituted for bank deposits in this illustration, and real estate or bonds could be substituted for stock.

for nourishment. The following table illustrates some of the different reactions which may occur when incomes increase:

SMITH'S DEMAND AT TWO INCOME LEVELS
WHEN PRICES REMAIN UNCHANGED

Weekly Income	Luncheon Hamburgers at 15¢ Each	Breakfast Coffee at 10¢ Cup	Weekly Beans at 18¢ Pound	Weekly Roast at 50¢ Pound
$30	2	1	2	0
40	3	1	1	3

The typical reactions are sometimes classified. Goods, the demand for which usually rises as incomes rise, are called *superior goods.* Goods the demand for which usually falls under these circumstances are called *inferior goods.* If all buyers responded like Smith, hamburgers and roasts would be superior goods and beans would be an inferior good. But buyers differ. Andy McTavish, for instance, might not increase his buying of hamburgers and roasts at all and might purchase the same quantity of beans as before. When his income increased, he might choose to increase his savings instead of pampering his appetite. And when incomes in general rise, some people's incomes rise more than others. Nevertheless, experience shows that most goods are "superior." There are certain exceptions which are generally admitted to be "inferior." Poor people have to buy them to satisfy their vital physical needs. For instance, they generally buy margarine instead of butter, flour instead of bakery bread, beans instead of meat, street-car transportation instead of automobiles and gasoline, and tickets to third-run movie houses instead of first. When their incomes increase, they begin to satisfy taste as well as hunger, desire for comfort as well as shelter. Conspicuous consumption begins to take precedence over sustenance.

For certain individuals and perhaps for the market as a whole, some goods are "superior" for a rise in income up to a certain level and then become "inferior" as the individual becomes still more affluent. A poor woman may use a washboard for the family laundry. If her financial status improves, she may enter the market for a washing machine. But if she becomes very well to do, she may send her laundry out and have no use for a machine in the house. The same story might be told for a given grade of shoes or meat or housing.

8. Income-Elasticity of Demand.—The whole concept of the response of demand to income changes may be summarized by the phrase *income-elasticity of demand.* This may be defined as the degree to which the demand for a good changes with a given small change in income. Income-elasticity of demand will vary with the person, the good, the income level at which the income change occurs, the direction of that change in income, etc. Using this terminology one may say that superior goods have positive, and inferior goods have negative, income-elasticity of demand. Luxury goods will generally have high positive income-elasticity, while staple foodstuffs will have low, or negative.

Certain applications of the concept may now be considered. Economists point out that the income-elasticity of demand for savings rises as incomes rise. This has an important bearing upon the end of boom periods and the arrival of the crisis which precedes a recession. It also affects the distribution of incomes among people by increasing the property incomes of those already in the high income brackets. Businessmen are often interested in what a rise or fall in the general income level of their customers will do to the demand for their particular products. If managers find that consumers' incomes are falling, they may strive to prevent demand from declining by using more advertising, or by adopting more liberal credit terms. If they cannot maintain schedule demand by such devices, they may seek a larger market demand by cutting prices.

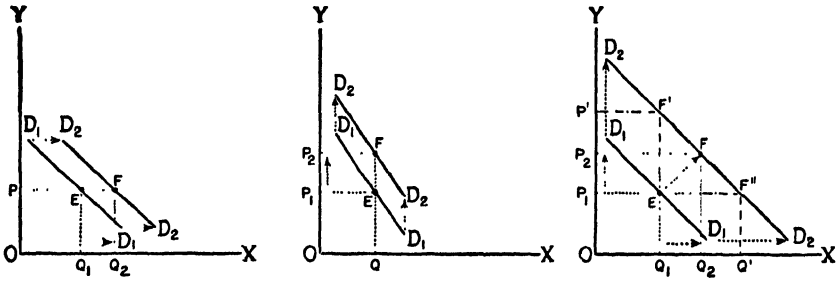
This leads to a final comment on the importance of understanding the two ways in which a change in income may affect market transactions. To this point the implication has been that there will be a change in the quantity purchased. The other possibility is a change in the price. Or both may change. For instance, a rise in income may not lead a person to buy two automobiles, but merely to buy a better one. He may also rent a better home, pay more for his clothes, or go to more expensive places of entertainment. Or the increased income may make him willing to pay more for such items as oranges and steaks. When these rise in price at the same time that his income rises, he continues to buy, instead of dropping out of the market. In short, an increase in schedule demand, however caused, may be described as a rise in the demand quantities, with prices constant; a rise in demand prices, with quantities constant; or a rise in both. The first possibility was portrayed diagrammatically in Figure 1. It is repeated below in the more complete presentation of Figure 2, which shows also the second and third possibilities.

An increase in demand quantity means a shift of the demand curve to the right as in Figure 2A. An increase in demand price shifts the demand curve upward as in Figure 2B. When both price and quantity change, the movement is both horizontal and vertical, but is most easily conceived as moving diagonally to the "northeast," as from E to F in Figure 2C.

FIGURE 2

THREE WAYS OF SHOWING AN INCREASE IN SCHEDULE DEMAND

A. Increase in Quantity B. Increase in Price C. Increase in Both Price and Quantity



9. Number of Buyers and Degree of Collusion Among Them.— The arguments of the three preceding sections have been based upon individual reactions, although the importance of the group aggregate has also been stressed. We turn now to note the importance of the number of people in the buying group. Whenever the number of buyers changes, the demand curve shifts in one direction or the other with resultant changes in prices or quantities. Changes in the population through migration or a balance of births versus deaths have been of some historical importance. The composition of the population is also significant, as when a declining birth rate decreases the percentage of children to be provided for at any given time and increases the percentage of oldsters. Immigration may change the number of people wanting particular types of goods much more than it changes the total number of buyers.

In the short run there is also the possibility that in some markets buyers may get together and agree not to engage in price bidding against one another. When this happens, as in agricultural buying cooperatives, or joint army-navy purchase commissions, there will be a decrease in schedule demand below what it was before the collusion occurred or what it might have become in the absence of such

inter-buyer agreements. In a later chapter this problem will be discussed under the heading of monopsony.[3]

10. Habit, Impulse, and Choice.—Most of our buying is routine. We have formed habits of buying things which have satisfied us sufficiently in the past and therefore we continue to do the same in the present. This holds for most of our purchases of food and shelter, and for some of our clothing and other needs. Nearly all goods or services which we consume regularly and have to purchase frequently are bought without much conscious choice. However, the goods which are bought only once or twice a year, or even less frequently, like certain articles of clothing or other durable consumption goods, vacation trips, etc., often involve a weighing of the relative merits of alternatives before deciding upon market action. Other purchases are impulsive, made upon the decision of the moment without reflective thought, and are outside the orbit of routine buying habits. A certain food or dress may catch a housewife's fancy, or a youngster may be inveigled by a sidewalk barker into taking a dollar diamond to his girl friend.

The generalization to be made with regard to all buying of the habitual type, most of that where choice enters, and some of that which is described as impulsive, is that in its broad outlines it is largely determined by the culture pattern of the group in which the person has been brought up or in which he currently lives. This is true even where the buyer thinks he is exercising reason and trying to get the most for his money. His appraisal of what is good is chiefly a result of his social conditioning and his knowledge of materials. Impulsive buying is more apt to reveal individual differences, but much of it also reflects group-fostered desires and scales of value even if these are below the level of the buyer's consciousness.

On the other hand, nearly all market action to purchase goods is a choice between alternatives, even though the element of deliberation is not present. This follows because a person can rarely do more than one thing at a time; if he is buying a coat, he cannot at that moment be buying another coat, nor a hat, nor an order of fried chicken. If we rule out the narrow time element which makes this generalization hold for even habitual choices, we encounter a few cases where people's incomes are so large relative to their desires that they can buy whatever they want in the course of a year or so

[3] Monopsony is to the buying side of the market what monopoly is to the selling side. Briefly stated, it indicates that there is only one buyer, or that buyers act in collusion to reduce prices to the detriment of the seller.

and need not deny themselves anything purchasable. But even the wealthy tend to live in a *milieu* which limits their range of choices. Though freed from the trials of thrifty shopping, they are often bound more tightly by convention than poorer people to whom social approval is not so important.

Only a few enterprising and fearless individuals break from the cultural mold which governs their choices among categories and rates of consumption. At the consumer level such conduct is considered queer, penny-pinching, or irrational by the people who conform. At the producer level it may be considered a sign of greatness, particularly if the innovation succeeds in raising profits far above what the conformists make. If the enterpriser fails, he may be dubbed reckless, visionary, or a gambler. The words which the conformist uses to appraise the nonconformist are matched by those of the radical expressing his disapproval of the conservative. They represent value judgments, not careful analysis.

The significance of the foregoing discussion for the marketplace is twofold: buyers change the pattern of their purchases very slowly, but they do change, especially in particulars. Hence the seller who wants to increase his sales may benefit from spending money to try to make buyers believe his product offers greater value than another which it resembles. This is most likely to be successful with articles of infrequent purchase, but it also holds for other goods where the difference between rival commodities is not much greater than the trademark which distinguishes them. Consumers may be stimulated, educated, misinformed, or otherwise influenced in their buying of particular articles, but major changes in the categories of consumption are not easy to achieve and take a long time. Considerable attention to these problems as seen from the seller's viewpoint will be given in a later chapter.

11. Summary of Principles Governing Demand as a Cause of Price.—Prices vary directly with changes in schedule demand. These changes themselves depend chiefly upon four causes: (1) changes in the *relative desirability* of various goods; (2) changes in preference for *goods as compared with currency;* (3) changes in *current income;* and (4) changes in the *number of buyers* in the purchasing group. The first gives rise to the useful concept of cross-elasticity of demand. The third suggests the term, income-elasticity of demand. The second is the source of concepts of liquidity preference and propensity to consume. The fourth refers to problems of collusion, monopoly, migration, and birth rates. The causes of these causes may be

examined through many links of the causation chain. Of major importance is a study of the influence of habit and impulse upon the choices which buyers make. All of these topics will receive further attention in later chapters as they are applied and developed. This chapter merely seeks to lay the foundation by defining certain important concepts in the field of demand. The related topic of price-elasticity of demand is so important that all of Chapter 4 is devoted to it.

Chapter 3

TYPES OF DEMAND SCHEDULES

1. Demand in the Market Sense: the Law of Demand.—There are two economic laws about demand which have become axiomatic to the man in the street, but he often confuses them. The first states that a rising demand tends to raise prices; the second says that rising prices tend to reduce demand. The economist sharpens and generalizes these two axioms. The first becomes *"Price varies directly with demand,"* and is called the *First Law of Price.* The second is made to read *"Demand varies inversely with price,"* and is called the *Law of Demand.* Chapter 2 dealt with the first of these economic laws and suggested the use of the phrase *schedule demand* in connection with it and *market demand* in connection with the second law. The present chapter will explore general aspects of the relation between changes in prices as causes and changes in quantities as results. This requires an analysis of the basic nature of demand schedules and their various types. It will be followed in Chapter 4 by an analysis of "demand-elasticity," the mathematical relation between successive quantity-price pairs in demand schedules.

The questions to which this chapter gives answers include the following:

A. How are demand schedules derived?
   1. What are subjective demand schedules?
   2. What are statistical demand schedules?

B. What are the major types of demand schedules?
   1. What is the difference between individual and collective schedules?
   2. How does a successive demand schedule differ from an instantaneous demand schedule?
      (a) How allow for the direction of price change? its starting point? its amplitude?
   3. How may long-run demand schedules be distinguished from short-run?

C. What exceptions may be found to the Law of Demand?

2. The Subjective Nature of Most Demand Schedules.—As explained in the preceding chapter, a demand schedule may be considered as a series of quantities that some one *thinks* a buyer or group of buyers would purchase at a series of prices. The difference between the several quantities is assumed to be due solely to differences in the price. Demand variations in this sense (market demand) are an effect of price, not a cause. Any person may describe a demand schedule. He may be a buyer telling what he thinks he would buy at various prices, a seller estimating what he thinks other people will buy if the selling price is at various levels, or a neutral third party, such as an economist who is appraising probable market activity in a given situation. These three persons, viewing the same set of conditions, would probably suggest similar but slightly different demand schedules.

Since demand schedules are subjective, they are also transitory. They change as rapidly as individual attitudes change. Unpredictable and trivial developments may be important in deciding how much an individual thinks he will buy, such as changes in the weather, an attack of indigestion, or the arrival of a bill collector. Sellers' forecasts of the demand for their product are apt to change with rumors of strikes, inflationary legislation, or war. Economists often seek objectivity by examining the records of the past and the probable conditions of the future, but prejudices and apprehensions influence their judgment, too.

3. Statistical Demand Schedules.—For certain staple commodities like wheat, sugar, potatoes, and butter it is possible to accumulate statistics showing the amounts which have been purchased in the past at different market prices. These prices did not occur successively through time in any regular sequence from low to high or vice versa, but they may be rearranged into such order by rearranging their dates. The price-quantity pairs thus obtained are then usually plotted on a graph and a trend line calculated which represents the best compromises between similar yet discrepant points. From this calculated line, price-quantity pairs may now be read at regular intervals and arranged in schedule form.

Statistical demand schedules thus derived may have some usefulness in economic prognosis, but they must be interpreted with care and applied with caution. Extending over many years as the basic records do, they show the effect of many other forces upon quantity demanded in addition to price change. Hypothetical or subjective schedules imply that nothing changes except the stated variables, but

the statistical schedules cannot. However, this may be turned into an advantage for the latter type *if* deviations from the trend line are examined and explained in terms of the other independent variables. For instance, a year in which the supply of potatoes is exceptionally large may occur either during a depression or during boom times. The size of the national income will surely cause the resulting price to be higher in the latter case than in the former. For *any given price,* the quantity taken from the market will be greater. In other words, supplementary generalizations about the changes in income and other forces described in the preceding chapter will help the forecaster who knows probable changes in these other variables in addition to price.

Statistical demand schedules are retrospective or *ex post* and may be terminologically contrasted with subjective demand schedules by calling the latter prospective or *ex ante. Ex post* demand schedules of the price-quantity type may be computed from data as simple as the sales records of any given firm which changed its selling price during a given time period. There will be as many price-quantity pairs as there were prices established—sometimes only two. In such cases, the analyst must be careful not to fill in other figures by interpolation since alteration in the amount and direction of the price change would have changed the quantity responses of buyers. Statistical demand schedules are sometimes interpreted to show what the approximate demand for a given product will be in the future, judging from what it was in the past. This procedure is more likely to yield more useful results for agricultural price and production planning where there are many buyers and sellers than for single firms or for industries where either buyers or sellers are few. In the latter instances changes in the attitudes or tactics of competitors may quickly and sharply alter the response of buyers to the original price changes of a given firm or commodity.

4. The Relation Between Individual and Collective Demand Schedules.—Statistical demand schedules are almost always of the collective type. That is they describe the demand of a *group* of individuals at various different periods of time. Subjective demand schedules may also be collective, but they are often constructed for separate individuals.

Collective demand schedules may be constructed from individual demand schedules when needed for economic analysis. The method followed is to set up individual demand schedules with the same series of prices. The quantities which are paired with each price in

the several schedules are then added together to obtain the total, or collective demand at that price. The following table shows hypothetical individual schedules for Adams, Brown, and Cross, together with a collective demand schedule according to the rule just given.

Adams		Brown		Cross		Collective	
Price	Quantity	Price	Quantity	Price	Quantity	Price	Quantity
8¢	2 units	8¢	0 units	8¢	0 units	8¢	2 units
7¢	3 "	7¢	1 "	7¢	0 "	7¢	4 "
6¢	4 "	6¢	2 "	6¢	0 "	6¢	6 "
5¢	5 "	5¢	3 "	5¢	1 "	5¢	9 "

Another type of collective demand schedule is of importance because of certain market situations. It is based upon the principle that a person who is willing to pay a given price for one unit of a good will be willing to buy that good at any lower price. Thus we might have a collective demand schedule for apartments, shall we say, made up of individuals Dean, East, and Foss as shown in the following schedule:

Dean		East		Foss		Collective	
Price	Quantity	Price	Quantity	Price	Quantity	Price	Quantity
$80	0 units	$80	0 units	$80	1 unit	$80	1 unit
70	0 "	70	0 "	70	1 "	70	1 "
60	0 "	60	1 "	60	1 "	60	2 "
50	1 "	50	1 "	50	1 "	50	3 "

Dean is willing to rent a four-room apartment if he can get one for $50 but he is not willing to pay more. East is willing to pay as much as $60, and Foss will go as high as $80. This means that there is only one apartment demanded at $80. The same is true if apartments rent for $70. At $60, however, East enters the market and his demand is added to that of Foss, who of course remains willing to buy at the lower price. The demand in the market sense rises to two at $60. At $50 East and Foss remain in the market and Dean enters so that there are three apartments demanded. This type of collective demand schedule is said to be cumulative downward. As the price falls, the demand present at the preceding higher price is added to the demand which enters the market at the next lower price so that a

larger total accumulates for the group. A similar type of collective schedule will be described later for supply as being cumulative upward. The difference between demand and supply schedules lies in the fact that the quantity demanded usually varies inversely with the price, whereas the quantity supplied varies directly.

5. Instantaneous Demand Schedules.—Up to this point a demand schedule has been defined as showing the quantities that would be demanded at each of a series of prices. Such demand schedules may be described as being of the instantaneous type. They implicitly assume that each price originates suddenly from nowhere. The quantity paired with the prices is the amount that would be demanded if that price existed. Such might be the calculation of a person who was about to market a new fountain pen and had to decide what would be the best price to ask. Should it be priced to sell at $4, the quantity demanded would be a certain amount; at $5, another amount; at $6, still another amount. The manufacturer would have to decide between the various possible prices upon the basis of the quantity that he thinks can be sold at those prices, the cost of production, and other variables.

6. Successive Demand Schedules Should Indicate the Direction of Price Change, Its Starting Point, and Its Amplitude.—For all old products already on the market the seller's problem is one of price *change*. He wants to know what the demand would be at a price higher or lower than his present price when buyers will be influenced by memory of what the price was previously. His demand schedule will, therefore, be a *successive* demand schedule since it will indicate the quantities that he thinks people will buy if his price is changed in succession from one figure to another.

In a successive demand schedule the *direction* of price change must be shown. A price increase will affect demand in a different way from a price decrease. Therefore, if three prices such as $5, $6, and $7 occur in the demand schedule for hats, one must state whether the quantity shown at each figure represents what could be bought if price succession runs from $5 to $6 to $7, falls from $7 to $6 to $5, or starts at $6 and moves either $1 up or $1 down. Each of these three situations would be likely to show different quantities paired with the given price.

Most merchants know from experience that a 10 per cent price reduction will bring a different percentage change in demand than a 10 per cent price increase will bring. A decrease usually does not

attract nearly so many buyers as an equal increase repels. Old customers find out about price boosts more quickly than noncustomers find out about price cuts. That is why advertising is generally essential in price-cutting campaigns. In most cases when a price is reduced and later returned to its original figure the demand is likely to be different at the end of the down-up change than it was at the beginning. This fact is sometimes generalized as a statement that (successive) demand schedules and curves are not reversible.

In addition to the direction of price change, it is also essential to know the *starting point.* The demand for hats this year at $5 will probably be different if the price has been reduced from $6 than if the reduction started from a previous price of $7. Similarly, a 10 per cent price increase will have a different effect on the demand for butter if it starts from 40 cents than if it starts from 80 cents.

This leads to the question of the effect of the *amplitude* of price change upon the change in the quantity demanded. Sellers sometimes find that their sales decline less if their prices are raised one cent at a time in five successive jumps than if there is a single 5-cent raise abruptly imposed. Therefore, if 50 units are being purchased at 20 cents, it is impossible to make a good guess about the amount that would be demanded at 25 cents without stating how rapidly the increase is to be made.

7. The Time Allowed for Buyer Response Should Be Indicated in Demand Schedules of the Successive Type.—When prices change, they affect the attitudes of buyers in different ways depending upon how long the price change remains in effect. For instance, when a price is increased, people at first may be repelled by the higher price and may refrain from buying. When they find that the price does not come down and they can do nothing about it, they may accept it as inevitable and resume purchases. When prices fall, it takes time for people to learn about the price cuts, and therefore the full market response is not revealed until sometime afterward. Therefore, in addition to showing the other things described in Section 6, a demand schedule of the successive type should tell the length of time allowed for consumer reaction to each price increase. This is shown in the schedule at top of page 34.

The difficulties involved in trying to make such a schedule realistic are obviously increased by making it longer. In actual practice, most schedules calculated by sellers rarely contain more than three or four prices: the going price, and one or two higher or lower alternative

A Successive Price-Quantity Demand Schedule
(Assume price rises one cent each week)

Price	Quantity Purchased per Week (After one week is allowed for consumer reaction to each price increase; demand not shown for other reaction times)
Beginning: 5¢	100
6¢	90
7¢	80
8¢	60
Final: 9¢	30

prices. For instance, the following schedule might be made to read: Beginning at 5 cents, an abrupt change to *either* 6 cents or 7 cents would cause demand quantities to fall as indicated; i.e., the jump to 7 cents would not be through an intermediate 6-cent stage, but all at once. The seller would also be wise to add a second column of quantities to indicate what the demand is likely to be at either the 6- or 7-cent figure after a longer time, say five weeks, is allowed for buyer adjustment to the higher price.

A Seller's View of Potential Demand Responses to Price Change

Price	Quantity Currently Sold per Week	Quantity He Expects to Sell in First Week after Change	Quantity He Expects to Sell in Fifth Week after Change
Present price: 5¢	100		
First alternative: 6¢	—	90	92
Second alternative: 7¢	—	80	85

8. A Classification of Demand Schedules into Short Run and Long Run.—Economists often describe things as happening in either the *short run* or the *long run.* There is obviously a time difference implied between them. For analysis, however, it is more important to concentrate upon the number of independent variables which operate in the hypothetical or real situation being described. The *short run* is best defined as a period of time in which, actually or hypothetically, only one independent variable is present. The *long run* should then be defined as a period of time in which two or more independent variables, such as price change and desire change, operate to affect the dependent variable. The difficulty with the latter definition is that we often want to distinguish between the operation of two, of

three, or of a larger number of independent variables. The phrase *long run* includes too much. We are forced to use qualifying adjectives such as the "moderately" long run or the "very" long run. This is cumbersome, but it is sanctioned by usage.

The various types of demand schedules suggested in the preceding paragraphs may be classified to illustrate this point.

I. Short-run schedules: one independent variable, the price of A, and one dependent variable, the market demand for A
   A. Instantaneous (included for completeness although the price of A really does not *change* in the usual sense of the word)
   B. Successive
      1. Reversible (probably the most common form in elementary analysis; though reversibility is often by implication only)
      2. Directional (with starting point and amplitude of changes also indicated)

II. Long-run schedules: the price of A changes, and one or more other changes occur to influence the schedule demand for A
   A. Successive, hypothetical
      1. *Two* independent variables: the change in the price of A and changes which it stimulates in buyer attitudes (many schedules possible depending upon the length of time allowed for attitudes to change)
      2. *Three* independent variables: *add* to the above a change in the price of other goods stimulated by the change in the price of A (many schedules possible depending upon time and the number of these other price changes)
      3. *Many* independent variables: *add* any other variables which are dependent as related to the price of A, but are independent as related to the schedule demand for A (such as the supply of a substitute commodity B)
   B. Statistical, historical
      1. *Many* independent variables; some of their changes may have been stimulated by changes in the price of A, but others are likely to have been caused by different forces. (Pairing the prices of A with quantities demanded often obscures these other variables and the time required for them to change and to influence demand. For instance, buyer incomes may change as suggested in Section 6 of the preceding chapter dealing with income-elasticity of demand.)

Much price analysis involves the simultaneous use of demand and supply curves. The analyst must be careful to keep their presuppositions in mind. For logical reasoning or accurate prediction it is often wise to specify each of the variables the force of whose change is being considered. Usually this brings to mind the *excluded* variables whose influence is not being taken into account, but should be.

9. Four Exceptions to the Law of Demand.—In most situations it is true that the demand varies inversely with the price, but there are at least four exceptions which deserve attention. These are the cases of discontinuous demand, speculative demand, style or prestige goods, and inferior goods. Much of the foregoing material sheds light upon these exceptions, which will now be examined.

The first case is that of a price change which causes no change whatever in the quantity demanded. The demand schedule or curve is said to be *discontinuous* for that interval. For instance, a decrease in the price of salt from 4 cents to 3 cents per pound may not increase John Smith's demand for salt at all. In fact the price might even fall to zero without increasing his demand. If the price should rise to 8 cents, a similar inelasticity of demand might be apparent. Somewhere above that figure, maybe at 50 cents, maybe at a dollar, Smith's demand for salt might decline a little as he chose to economize upon its use or to employ substitutes. Individual demand schedules for a great many commodities contain discontinuous segments, but few of them have as wide a gap between changes as salt. Collective demand schedules, on the other hand, usually reveal a continuous change in demand throughout the whole range of price change.

A second exception deals with the *speculative* type of market response. A rise in price stimulates demand to increase, and a falling price depresses demand. This reaction is the opposite of that described in the Law of Demand. The stock market offers an obvious illustration. A bull market attracts buyers whose added demand sends prices still higher. A bear market drives buyers away and prices fall still lower. Speculative reactions may be interpreted in the language of Chapter 2 as a change in investor preference for particular securities or for securities as compared with cash. A price increase, for instance, stimulates an increase in schedule demand which more than offsets the decrease in market demand that would otherwise be expected. In the language of Sections 7 and 8 of this chapter the speculative response might be described as a type of long-run demand schedule in which the buyer's attitude is changed. It is

different, however, from the reactions previously portrayed. In Section 7 the buyer was pictured as changing his attitudes in a way that offset somewhat the adverse effects of a price rise. In neither case did the reaction partake of the "bandwagon" type of response. Speculation falls in this latter category. This exception to the Law of Demand is quite common in security markets and on the organized commodity exchanges. It is much less common in business or consumer buying, but in periods of rapid price change it may appear quite often.

Style and prestige buying create the third exception. Some wealthy buyers of style goods show increased desire for a new item because the rising price puts the item clearly in a prestige class open only to those with large incomes. On the other hand, consumers who are in the habit of thinking that price measures quality, are likely to decrease their purchases of items whose price is cut below that of accepted substitutes.

The fourth possible exception to the Law of Demand is quite rare. It is found in connection with certain commodities and services which were described as "inferior goods" in Chapter 2, Section 7. The distinction between inferior and superior goods is most clearly seen when incomes change. In a very few cases it may appear when the price of a commodity itself changes. For extremely poor people, a fall in the price of a staple food such as flour or beans may enable them to buy other things for taste as well as nourishment. They do not feel the need for so much of the staple and therefore may reduce their purchases of it. The argument may also be applied to price increases, but with greater difficulty.

10. Summary.—Most demand schedules obey the simple Law of Demand in that demand in the market sense varies inversely with the price. This is clearly true for most instantaneous demand schedules and holds for all successive demand schedules except the speculative type and a few others. When the price of a good changes, it has a direct influence upon the quantity demanded.

The price change may also have an indirect influence through its effect upon other things that may alter demand in the schedule sense. These include particularly changes in buyer attitudes and changes in the prices of other commodities. Short-run demand schedules eliminate these secondary effects. Long-run demand schedules allow one or more of these roundabout influences to intrude as additional independent variables determining the quantity demanded. Statistical demand schedules reveal the effect also of many completely inde-

pendent variables, such as changes in income which are not caused by the change in price of the given good.

For maximum accuracy in interpretation and analysis, each demand schedule should state the variables assumed to be operating in addition to price change. It should indicate also the direction of price change, its starting point, and its amplitude.

Chapter 4

PRICE-ELASTICITY OF DEMAND AND ITS MEASUREMENT

1. The Price-Elasticity of Demand.—Despite obvious difficulties in constructing and interpreting demand schedules, the fact remains that the demand for A is a function of the price of A. In most cases it is true that a rise in market price reduces demand and a fall in price increases demand. This is the familiar Law of Demand defined in the first section of Chapter 3. But this law does not tell us enough. It merely says that the *direction* of demand change is usually the opposite of the *direction* of price change. Businessmen and economists want to know the *degree* to which the demand will change following a given change in price. This is the problem of the *price-elasticity of demand.* The phrase is commonly shortened to *elasticity of demand,* but the longer phrase is more precise and distinguishes the concept from the ideas of income-elasticity of demand and cross-elasticity of demand which were explained in Chapter 2.

The demand for some goods is much less responsive to price change than is the demand for others. Salt, for instance, is a traditional example of a good the demand for which does not change much with a change in price. The problem of demand-elasticity is of chief interest to growers, manufacturers, merchants, and other sellers of goods. They want to know what will happen to the demand for their products if the price is changed either by them or by persons or forces over which they have no control. Economists have made the problem one for much formal analysis, and the terms and logic employed must be mastered in order to facilitate the solution of more advanced problems.

The most important questions regarding price-elasticity of demand which are answered in this chapter are given below. The less important answers are printed in reduced type. They need be studied only by those who wish to become skilled in the use of diagrams which constitute the "shorthand" of the language of economics.

I. What are four ways of presenting the data of a demand schedule?

II. What are the meanings of these terms:
   A. Elastic demand

B. Inelastic demand
C. Demand-elasticity of unity

III. How may demand-elasticity be measured?
A. By changes in total revenue
1. Directly
2. Indirectly
(a) Through marginal revenue
(b) Through average revenue
B. By relative percentage changes in quantity and price

IV. What are the different concepts of marginal revenue?
A. Average marginal revenue
B. Marginal revenue at a point

V. In Notes:
A. How may the marginal revenue be found at any point on a smooth total revenue curve (the tangent rule)?
B. How are total, average, and marginal revenue curves related diagrammatically?
C. How may point-elasticity and arc-elasticity be demonstrated by plotting demand schedules on logarithmic axes?
D. How may point-elasticity be shown on demand curves using the customary arithmetical axes?

2. Ways of Presenting Demand Schedules: Emphasis on Total Revenue.—The effect of a change in prices upon the quantity purchased may be described by comparing the total amounts of money which pass from buyer to seller in completing the transactions at the different prices. The amount of money changing hands when goods are bought and sold at any given price is obviously the number of units exchanged multiplied by the price per unit, assuming that the price is quoted in unit terms. To the buyer this amount of money is his *total outlay* (TO); to the seller it is his total receipts or *total revenue* (TR). A neutral term which stresses neither the buyer's nor the seller's viewpoint might be *transaction money amount* or *total transaction money*. Despite the advantages of such a neutral phrase, for the present we shall use the term *total outlay* whenever we look at demand from the buyer's viewpoint and the more conventional term *total revenue* only when the seller's viewpoint is involved. The reader must remember that total outlay and total revenue are identical quantities.

The conventional demand schedule which shows a series of price-quantity pairs may be expanded into more complete form by computing a third column to show total outlay (total revenue) according to the equations: $TO = TR = P \times Q$, where Q is the number of

units changing hands at price P. From this three-column demand schedule it then becomes apparent that there are also two other ways of describing the demand situation defined by the conventional price-quantity demand schedule. We may set up a price and total outlay schedule or a quantity and total outlay schedule. From any one of these four schedules the other three may be obtained. The following table shows the relationship of the conventional demand schedule to these three others.

FOUR WAYS OF PRESENTING THE DATA OF A GIVEN DEMAND SCHEDULE

"Conventional" (Price and Quantity)		"Complete" (All Three)			"Price and Total Outlay"		"Quantity and Total Outlay"	
P	Q	P	Q	$TO(TR)$	P	$TO(TR)$	Q	$TO(TR)$
11¢	1	11¢	1	11¢	11¢	11¢	1	11¢
10¢	2	10¢	2	20¢	10¢	20¢	2	20¢
9¢	3	9¢	3	27¢	9¢	27¢	3	27¢

3. Demand-Elasticity, First Method: Change in Total Revenue. —The simplest approach to the measurement of demand-elasticity makes use of the total revenue column in a "complete" demand schedule of the type described above. For instance, a businessman who is contemplating a cut in his selling price is very much interested in how much that reduction will change his total revenue. There are three possibilities: total revenue may increase, decrease, or remain the same. Economists describe the first result as revealing an "elastic demand." The second result is called "inelastic demand"; and the third is called "demand-elasticity of unity." These three possibilities are shown by the three different demand schedules of the following table:

DEMAND SCHEDULES OF THREE DIFFERENT ELASTICITIES

Demand 1 ("Elastic")			Demand 2 ("Inelastic")			Demand 3 ("Elasticity of Unity")		
P	Q	TR	P	Q	TR	P	Q	TR
11¢	1	11¢	11¢	22	242¢	30¢	10	300¢
10¢	2	20¢	10¢	23	230¢	25¢	12	300¢
9¢	3	27¢	9¢	24	216¢	20¢	15	300¢

In Schedule 1 the total amount of money passing from buyer to seller increases as the price falls; in Schedule 2 total revenue decreases.

In the third schedule the total revenue remains unchanged for each price drop. In order to lend greater precision to the terms, economists sometimes describe "elastic demand" as "demand-elasticity greater than unity." The example has been given in terms of a price decrease. It should be obvious that if we were to consider a price increase instead, the relationships would be reversed. For instance, demand-elasticity greater than unity ("elastic demand") would occur when total revenue declined as the price rose.

4. Marginal Revenue as Change in Total Revenue.—The change in total revenue that results from each change in the independent variable is sometimes called *marginal revenue.* In the "elastic" demand schedule given above, the price decline from 11 cents to 10 cents causes an increase of the total revenue from 11 cents to 20 cents, a difference of 9 cents. Similarly, the next price change from 10 cents to 9 cents causes a 7 cent increase in total revenue. The 9 cent and 7 cent changes in total revenue may be called marginal revenues. The formula for calculating MR when the price decreases is as follows:[1]

$$MR_n = TR_n - TR_{n+1}$$

This holds when n represents the price at the end of a change interval of one. That is, when the price changes from 11 cents to 10 cents, the marginal revenue is said to be 9 cents *at 10 cents,* not at 11 cents, nor for the interval between 11 cents and 10 cents. Applying the formula:

$$MR_{\text{at } 10¢} = TR_{\text{at } 10¢} - TR_{\text{at } 11¢}$$
$$MR_{\text{at } 10¢} = 20¢ - 11¢ = 9¢$$

A rule may now be advanced for determining the elasticity of demand from marginal revenue instead of from total revenue. Demand is elastic when MR is positive. In the schedule on page 44 the rule applies from a price of 10 cents to a price of 6 cents. Demand is inelastic when MR is negative, as from 5 cents to 3 cents. Between a 6-cent price and a 5-cent price MR passes through zero. When $MR = 0$ this means that there is no change in TR, and we have demand-elasticity equal to unity.

The same three rules also may be expressed diagrammatically. When MR lies above the OX axis, demand is elastic. Where MR crosses OX, demand-elasticity is unity. And where MR lies below OX, demand is inelastic.

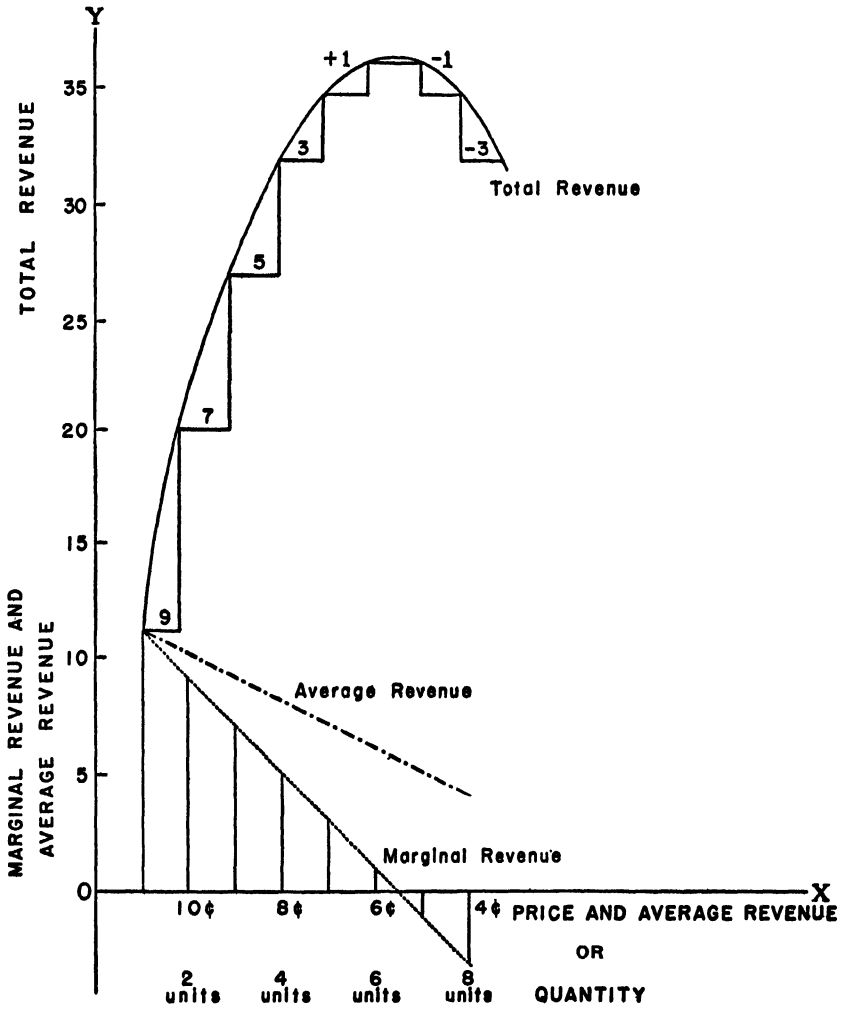
The diagram in Figure 3 is drawn in steps to show clearly the

[1] For a price *increase* the formula is $MR_{n+1} = TR_n - TR_{n+1}$, and $n + 1$ is the end price. Cf. Note A.

relation between *TR* and *MR*. If, now, the changes in price marked off on the *OX* axis are made very small, the steps become narrower. If one could conceive logically of price changes as small as 1/1000 of a cent or less, the corresponding changes in total revenue and marginal revenue would become so small as to make steps invisible to the naked eye. The curves would become "smooth." This, in fact, is the way they are usually drawn.

FIGURE 3

DIAGRAMMATIC RELATION BETWEEN TOTAL REVENUE AND MARGINAL REVENUE



DEMAND SCHEDULE CALCULATION OF TOTAL, MARGINAL, AND AVERAGE REVENUE

Price	Quantity	Total Revenue		Marginal Revenue	Average Revenue
12¢	0	0			
11¢	1	11¢		11¢	11¢
10¢	2	20¢	Demand	9¢	10¢
9¢	3	27¢	elastic	7¢	9¢
8¢	4	32¢		5¢	8¢
7¢	5	35¢		3¢	7¢
6¢	6	36¢		1¢	6¢
5¢	7	35¢	Demand	−1¢	5¢
4¢	8	32¢	inelastic	−3¢	4¢

When a smooth marginal revenue curve is used, the formula given above in terms of one unit changes in price must be modified to show infinitesimal changes in price. This is done by substituting m for 1 and stating that m approaches zero as a limit. That is, the horizontal width of the steps is not one cent, but is assumed to be such a small fraction of one cent as to be virtually zero. The formula is written:

$$MR_n = \frac{TR_n - TR_{n+m}}{m} \text{ when } m \to 0 \text{ as a limit.}$$

The division by m indicates that the change is really an *average* rate of change between the two points $n + m$ and n, even though it is customary to say that the change takes place *at* n. The need for an average rate may be seen clearly if m is taken to be a large number, not a very small one. For instance, if a demand schedule is written in terms of 5-cent drops in price, not one cent, we would have to compute the *average* change in total revenue between each two given prices in order to find the marginal revenue per one-cent change. This is shown in the following schedule.

Average MARGINAL REVENUE WHEN PRICE INTERVALS ARE LARGE

A Price	B Quantity	C Total Revenue	D Total Change in TR When $m = 5¢$	E *Average* Change in TR: "Marginal Revenue" (D/5)
$1.00	200	$200.00		
.95	230	218.50	$18.50	$3.70
.90	250	225.00	6.50	1.30

When $m = 1$, the divisor is rarely mentioned, since anything divided by one equals itself, unchanged. The total change in TR becomes equal to the average change in TR, and MR can be read directly merely by subtraction. The concept of demand-elasticity at a point is developed further in Note A.

NOTE A.—THE TANGENT RULE FOR FINDING MARGINAL REVENUE AT ANY POINT ON A SMOOTH TOTAL REVENUE CURVE. When smooth curves are used which have no "steps," a method must be derived for calculating MR at any point. The step-type curves use the convenient fiction that MR exists "at" the end of each step. Thus the step formed by reducing the price from 11 cents to 10 cents was 1 cent wide and 9 cents high. The MR was therefore declared to be 9 cents "at 10 cents." Later price reductions created other steps of diminishing height until the 6-cent price was reached, after which the "curve" turned downward. Steps can be measured easily, but on smooth curves there is no visible step at all. Instead of talking about the height of steps, we speak therefore of the "rate of change" in the dependent variable at any given point. For instance, a smooth TR curve corresponding to the step curve of Figure 3 might be said to rise at a diminishing rate which becomes zero just beyond the 6-cent price.

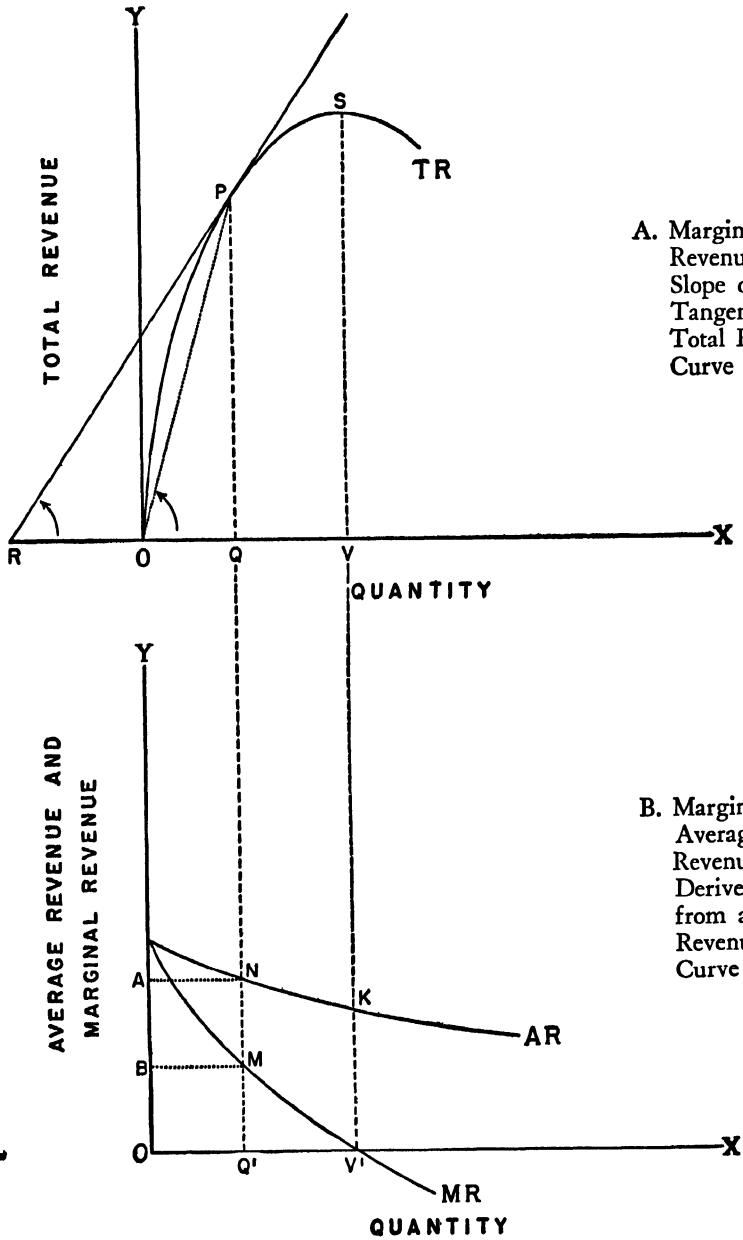
In still more general terms, we are seeking a rule for the diagrammatic measurement of the rate of change in the dependent variable at a given size of the independent variable. The technique is as follows. Choose any point P on TR which corresponds to a given price, and draw a straight line which just touches the TR curve at that point (see Figure 4). Such a line is said to be "tangent" to TR at P. The line itself is also called "the tangent." This tangent line makes an angle PRQ with OX at R and forms part of a triangle PQR. Two sides of this triangle represent an enlarged picture of the step at P, which is invisible because it is so small. RQ represents the width of the step and QP the height (by the geometric rule of similar triangles). In our original illustration, the large step was 1 cent wide and 9 cents high "at 10 cents." This indicated a "rate of change" of 9 cents up for 1 out, or a ratio of 9:1. We may perform the same sort of division with the sides of the larger but similar triangle PQR. Thus the marginal revenue, or rate of change in total revenue at 10 cents on the smooth curve is PQ (the height of this similar step) divided by RQ (the width).[2] The 10-cent *point* (P) on the smooth curve lies *between* the 9-cent and the 7-cent *steps*. Therefore, the marginal revenue *at that point* will be something less than 9 and greater than 7.

The marginal revenue rule may now be stated in the abstract terms of mathematics. It is merely one case of the general rule for determin-

[2] The diagrammatic distances do not precisely conform to the figures given here. $PQ:RQ$ is less than 9:1.

FIGURE 4

GEOMETRIC RELATION BETWEEN THREE DEMAND CURVES



ing the *rate of change*: the change in the dependent variable divided by the change in the independent variable when the latter change becomes so small that it approaches zero as a limit. In symbols it is $\Delta TR/\Delta P$ when Δ signifies an infinitesimal change in the number which follows. The Δ symbol stands for the Greek letter delta, and the ratio may be read "delta TR divided by delta P." Diagrammatically a rate of change is the side of the triangle measuring the dependent variable divided by the side drawn on the axis which measures the independent variable. The triangle is formed by drawing a tangent to a given point on the curve which expresses the relationship between the given variables.[3]

As the point P moves upward along the TR curve, the tangent line drawn to it will become less steep. At the top of the TR curves, a tangent would become parallel to OX and the base RQ of the triangle would become infinite. Marginal revenue would become zero, as may be seen by observing (1) that the angle PRQ would shrink to nothingness and (2) that the ratio $QP:RQ$ would become zero.

5. Average Revenue Is Another Way of Describing Unit Price.—The total revenue in any demand schedule is equal to the price times the paired quantity ($TR = P \times Q$). If one takes the total revenue at each price and divides it by its demand quantity, the quotient will be the revenue per quantity unit, i.e., the average revenue ($TR/Q = AR$). From the formula, $TR/Q = P$, the quotient is clearly identifiable also as the price ($P = AR$).

Whenever comparisons are being made with marginal revenue, the most appropriate term for the quotient is "average revenue." In diagrams, AR always lies above MR, with one exception. When the selling price is constant regardless of quantity offered, AR will be horizontal and identical with MR. In all other cases, MR declines more rapidly than AR, as shown in Figure 3. When only one unit is sold, MR, AR, and TR are equal and meet at the same point.

6. Comparative Areas Under an Average Revenue Curve Also Reveal Changes in Total Revenue.—The conventional demand curve plots price on the vertical axis and quantity on the horizontal axis. Since the price is presumed to be the same for all units sold at a given time, total revenue equals price times quantity ($TR = P \times Q$). By the rules of algebra, price then becomes the total revenue divided

[3] In trigonometry the rule becomes: the tangent of the angle PRQ (formed by the intersection of the tangent line and the axis of the independent variable) is measured by the ratio between the opposite and adjacent sides of that angle. Readers unfamiliar with mathematics may be cautioned to remember that the word *tangent* may be used in three different ways: (1) as a line, (2) as a formula for measuring the size of the angle made by the tangent line, and (3) as the quotient or result obtained by applying that formula.

by the quantity ($P = TR/Q$) which is the same thing as the average revenue per quantity unit. Therefore, it is customary to refer to the price-quantity demand curve as an average revenue curve. The diagrammatic relation between total, average, and marginal revenue curves is demonstrated in Note B.

NOTE B.—Diagrammatic Comparison of Total, Marginal, and Average Revenue Curves. If a total revenue curve is drawn immediately above its corresponding average revenue and marginal revenue curves as in Figure 4, some interesting relationships may be observed. In the first place, either of these three revenue quantities may be measured directly by the height of its own curve at any given quantity. In the second place, any one of the curves may be derived from either one of the others. For instance, when marginal revenue is zero, total revenue cannot be increasing. This may be seen at quantity OV in the upper diagram where TR reaches its peak, S, and in the lower diagram where MR cuts the quantity axis at V'.

At any intermediate quantity, Q, the rules for computing marginal revenue and average revenue from total revenue are as follows. First, draw a perpendicular to OX to intersect the total revenue curve at P. From this total revenue point P draw a straight line to the origin, O, and a tangent which intersects the quantity axis at R. The angle PRQ shows marginal revenue by dividing RQ into QP according to the tangent rule. Similarly, the angle POQ shows AR by dividing OQ into QP. These quotients may then be plotted below as $Q'M$ (QP/RQ) for marginal revenue and $Q'N$ (QP/OQ) for average revenue.

One should note the change in the independent variable from that given in the body of the chapter. TR was there shown to vary with the price; it is shown here as varying with the quantity. Since the quantity varies also with the price, and since a complete schedule shows all three: P, Q, and TR, it is permissible to use whichever of the two pairs is the most useful (cf. Chapter 4, Section 2).

When comparing income with cost, marginal curves and average curves are usually used (MR, MC, AR, and AC). When considering the total amount of funds spent in the market, total outlay (same as TR) curves are sometimes helpful, but they do not appear very often in economic literature. All three revenue curves may be called demand curves. Some authors drop the term *demand curve* when talking about the situation faced by the individual firm planning its future sales policy and speak instead about a "sales curve." This term is not without merit, but tends to be confused with a curve showing the trend of sales quantities over a period of time.

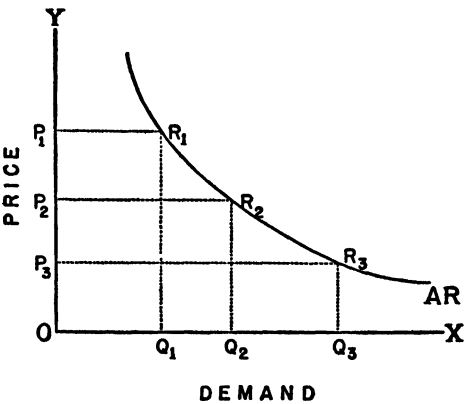
For simplicity, most TR, AR, and MR curves are assumed to be of the reversible type and no question is raised about their being instantaneous or successive (cf. Chapter 3, Sections 5 to 7).

The total revenue may be determined from an average revenue curve by selecting any point, R_p, on that curve and drawing lines perpendicular to OY and OX. This gives a rectangle of which one side (OP_1) measures price and the other side (OQ_1) measures quantity. Therefore, the area of the rectangle ($P_1R_1Q_1O$) represents price times quantity, or total revenue. By moving point R up and down the curve a series of such rectangles may be obtained as in Figure 5, and their areas may be compared. If the rectangle defined by R_2 is larger than the rectangle defined by R_1, this indicates that TR is rising. Demand, therefore, would have an elasticity greater than unity when the price fell from P_1 to P_2. If the rectangle defined by R_2 were to be equal in area to the rectangle defined by R_1, demand would have an elasticity of unity between these two prices. If the second rectangle should be the smaller of the two, then demand-elasticity would be less than unity. As a general rule, the steeper the curve at point R, the less elastic the demand is likely to be for a price change from that point. A nearly horizontal curve would indicate a very elastic demand. This rule, however, must be applied with great caution because it does not always hold. The reasons for the exceptions and two ways of resolving the difficulties are given in Note C. In the demand and supply diagrams most commonly used in price analysis, the comparative-areas approach to demand-elasticity is generally the easiest and best.

FIGURE 5

DEMAND-ELASTICITY SHOWN BY COMPARING AREAS UNDER A DEMAND CURVE

$$P_1R_1Q_1O \gtreqless P_2R_2Q_2O \lesseqgtr P_3R_3Q_3O$$



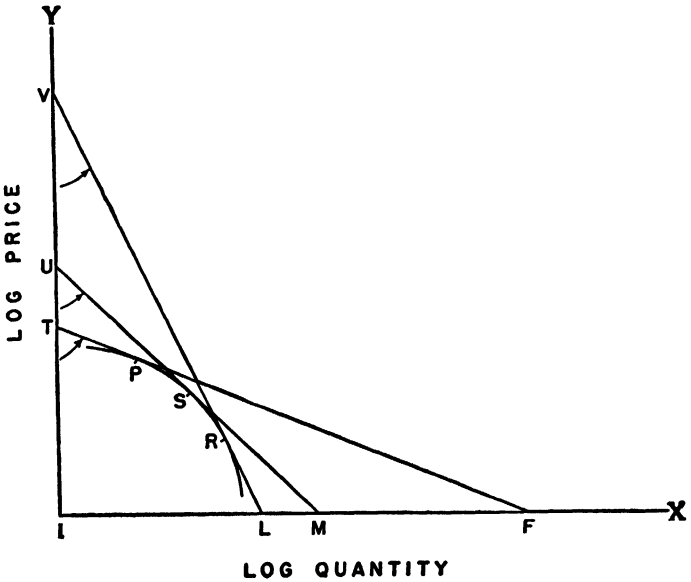
NOTE C.—POINT-ELASTICITY AND ARC-ELASTICITY OF DEMAND SHOWN BY USE OF LOGARITHMIC AXES. (1) *Average Revenue Curve Plotted on Logarithmic Axes.* If the price-quantity pairs of the *average* revenue demand schedule are plotted on *logarithmic* ruled axes, the slope of the AR curve at any point reveals the price-elasticity of demand. (The slope of the *total* revenue demand curve on *arithmetic* ruled axes reveals the marginal revenue.) Arithmetic axes start at zero and mark equal intervals for *equal additions,* such as marking points 0, 2, 4, 6, 8, 10, etc., for additions of 2 each time. Logarithmic axes start at one and mark

equal intervals for *equal multiplications,* such as marking points 1, 2, 4, 8, 16, 32, etc., for multiplications by 2 each time. Thus if each interval is one centimeter long, the point five centimeters out would be marked 10 units on an arithmetic axis, but 32 on a logarithmic axis.

Figure 6 shows the way in which the elasticity of demand may be demonstrated on logarithmic axes. Since price is the independent vari-

FIGURE 6

DEMAND-ELASTICITY REVEALED BY SLOPE OF AVERAGE REVENUE DEMAND CURVE ON LOGARITHMIC AXES



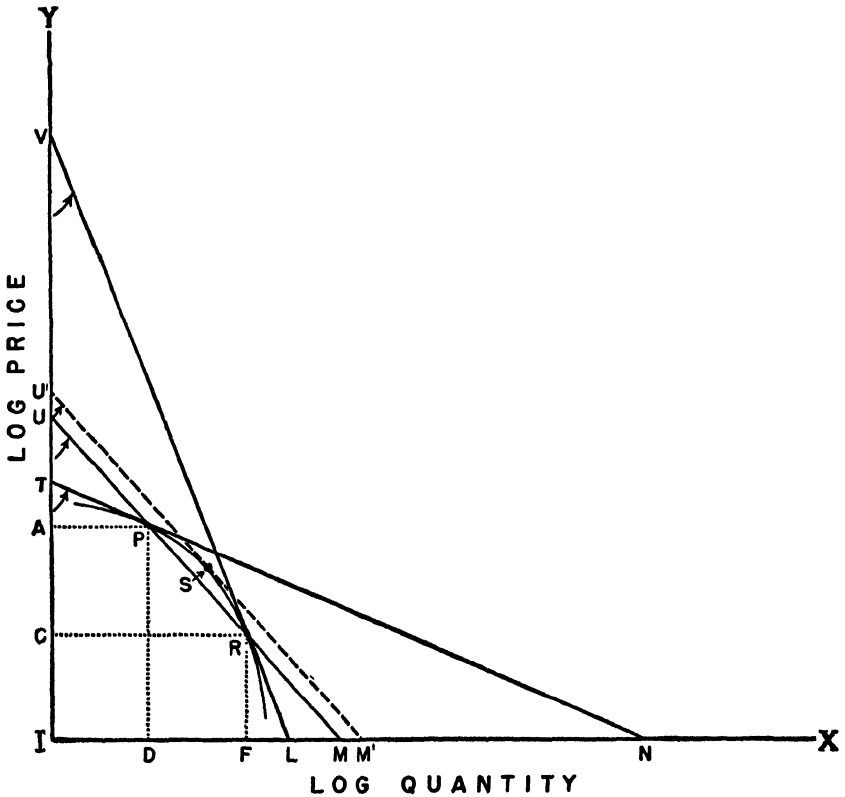
able, the angle which shows the slope of the tangent line is measured where that line intersects the *price* axis. If that angle is equal to 45 degrees (*SUI*), the demand is equal to unity. If the angle (*RVI*) is less than 45 degrees, the demand is inelastic, or less than unity. If the angle (*PTI*) is greater than 45 degrees the demand is elastic, or greater than unity. It is customary, however, to consider the "steepness" of the tangent lines and that means viewing the angles they make with the *quantity* axis. Here a small angle means great elasticity and a large angle means small elasticity.

The demand curve in Figure 6 is drawn for realism so as to bend away from the origin (the meeting point of the log X and log Y axes) instead of toward the origin. Most demand schedules are more elastic for very high prices than they are for very low prices.

(2) *Point-Elasticity Contrasted with Arc-Elasticity.* The use of logarithmic axes makes it easy to demonstrate why arc-elasticity must be considered the average elasticity between two points. On the demand curve of Figure 7 movement from point P to point R represents a

FIGURE 7

ARC-ELASTICITY IS AN AVERAGE ELASTICITY BETWEEN POINTS



decrease in price from IA to IC which causes an increase in demand from ID to IF. The *arc*-elasticity between P and R may be determined by drawing a straight line or chord between these two points and measuring the angle, PUA, which it makes with the price axis. This angle is smaller than the *point*-elasticity angle, PTA, for point P and larger than the point-elasticity angle, RVA, for point R. If a comparison between tangents is desired, a point S may be found on the arc of the curve between P and R such that its tangent SU' is parallel to the chord RPU and the angle $SU'A$ is equal to angle PUA.

7. Demand-Elasticity. Second Method: Ratio Between Percentage Changes.—A second method of measuring demand-elasticity gives numerical answers by the application of a simple formula to the changes in a conventional demand schedule. These results may then be compared with unity (1.00) to determine the degree of demand-elasticity for a given price change, the demand is said to be "elastic" if the formula quotient is greater than 1.00, and "inelastic" if less than 1.00. If price change is considered to be the cause and quantity change the effect, then demand-elasticity may be described in general as the degree to which the latter is affected by the former. The logical way to make this comparison is in terms of the percentage changes on each side of the schedule. The elasticity of demand may be expressed by the formula (in words):

$$\text{Elasticity} = \frac{\text{Percentage change in quantity}}{\text{Percentage change in price}}$$

There are two difficulties involved in the application of this formula. The first deals with the appropriate divisor to use in computing the percentages. Suppose the price falls from 11 cents to 10 cents and the quantity demanded rises from 1 unit to 2 units as in one of the schedules given above (page 41). The amount of change in price is 1 cent. This may be expressed as a percentage in any one of three ways: 1/11, 1/10, or 1/10.5. The latter possibility is the best because it recognizes that the change occurs *between* 11 cents and 10 cents and therefore the midpoint or average figure should be the divisor: $(11 + 10)/2 = 10.5$. Similarly, the quantity percentage is best expressed, not as 1/1 or 1/2, but as 1/1.5.

The second difficulty lies in the fact that one of the numerators in these percentage fractions will have a negative sign. At least this will be true in every demand schedule following the Law of Demand. In the above illustration the quantity changes from 1 to 2, and therefore subtracting the second number from the first gives a minus 1, not a plus 1. The price change is $11 - 10 = +1$, so there is no trouble there. But if the schedule were being read in the other direction because we were interested in the effect of a price increase, the plus and minus signs would be reversed. The subtractions would be: $2 - 1 = +1$, and $10 - 11 = -1$. Because of the appearance of a negative sign in either the numerator or the denominator, the quotient would be negative and would, therefore, have to be compared with -1.00, not with $+1.00$. This is illustrated by the following example based upon the figures given above:

$$E = \frac{-1}{1.5} \div \frac{1}{10.5} = \frac{-1}{1.5} \times \frac{10.5}{1} = \frac{10.5}{-1.5} = -7.00$$

The elasticity of demand is -7.00, which is said to be greater than -1.00 and therefore the demand is very "elastic." The next change in the schedule is from 10 cents to 9 cents, and from 2 units to 3 units. Applying the formula to test elasticity for this change, we find that:

$$E = \frac{-1}{2.5} \div \frac{1}{9.5} = \frac{-1}{2.5} \times \frac{9.5}{1} = \frac{9.5}{-2.5} = -3.8$$

The demand is less elastic than before, but it is still "greater than unity." Down below the middle of the given schedule, demand-elasticity becomes less than unity, as in the case where price changes from 6 cents to 5 cents and the demand rises from 6 units to 7 units.

$$E = \frac{-1}{6.5} \div \frac{1}{5.5} = \frac{5.5}{-6.5} = -.846$$

The elasticity formula may now be stated in more general terms. Instead of taking the average of the two prices and the average of the two quantities which requires adding in each case and dividing by 2, we can use simply the totals because the 2 in each fraction cancels out. Nor do we need to change the price by only one cent each time. A difference of any size is permissible, provided the corresponding quantities are used. Thus in the illustrative schedule, if the change is supposed to be from 7 cents to 5 cents, where the quantities rise from 5 units to 7 units, a difference of 2 in each case, the application of the elasticity formula gives a quotient of -1.00, as follows:[4]

$$E = \frac{-2}{7+5} \div \frac{2}{7+5} = \frac{-2}{12} \times \frac{12}{2} = -1.00$$

In symbols the demand-elasticity formula becomes:

$$E = \frac{Q_1 - Q_2}{Q_1 + Q_2} \div \frac{P_1 - P_2}{P_1 + P_2}$$

8. Demand-Elasticity, Third Method: Ratio Between Segments of a Tangent to a Demand Curve.—A third method of computing demand-elasticity is given in Note D. It is derived from this second method and is used to determine or to demonstrate the elas-

[4] This case may be tested, as may all the others, by applying the first method given for determining demand-elasticity. The total revenue remains unchanged at $35¢ = 5 \times 7¢ = 7 \times 5¢$.

ticity of demand at any price-quantity point on a smooth demand curve. Although useful in the interpretation of the diagrams which constitute the shorthand of economics, it is not applicable to problems faced by economists or businessmen who want to know the probable effects of a change in price.

NOTE D.—POINT-ELASTICITY WHEN DEMAND CURVES ARE PLOTTED ON CUSTOMARY ARITHMETICAL AXES. When the price change is infinitesimal, demand-elasticity at the point from which the infinitesimal movement occurs may be calculated by using a variant of the formula given in Section 7 which uses the quotient of two percentage changes. Denoting the midpoints between the two prices as P_m and between the two quantities as Q_m, the formula becomes:

$$E_m = \frac{Q_1 - Q_2}{Q_m} \div \frac{P_1 - P_2}{P_m}$$

inverting:

$$E_m = \frac{Q_1 - Q_2}{Q_m} \times \frac{P_m}{P_1 - P_2}$$

whence:

$$E_m = \frac{P_m}{Q_m} \times \frac{Q_1 - Q_2}{P_1 - P_2}$$

or:

$$E_p = \frac{P}{Q} \times \frac{\Delta Q}{\Delta P} = \frac{\text{Price}}{\text{Quantity}} \times \frac{\text{Infinitesimal change in quantity}}{\text{Infinitesimal change in price}}$$

since, in dealing with infinitesimal changes, any "midpoint" between prices or quantities is the same as either point.

Then, by applying the rule developed in Section 4 and Note A, the second term in the final equation above may be said to equal the marginal quantity. The final equation becomes:

$$E_p = \frac{\text{Price}}{\text{Quantity}} \times \text{Marginal quantity}$$

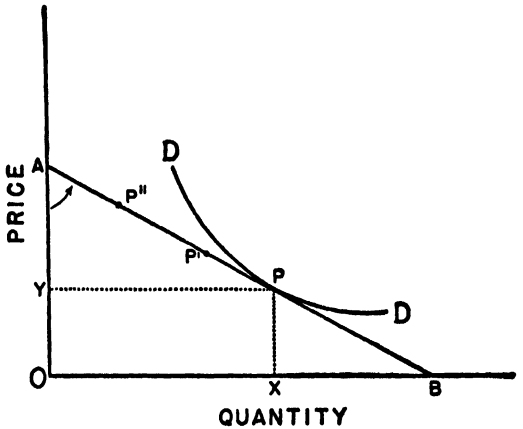
This formula may be presented diagrammatically and then generalized into a geometric rule for finding demand-elasticity at any point on a given demand curve. Thus in Figure 8 the marginal quantity at P, representing price OY, is the slope of the tangent APB with reference to the price axis. In terms of lines on the diagram this equals OB/OA. Then by geometry we find that $\frac{OB}{OA} = \frac{YP}{AY} = \frac{OX}{AY}$. The latter term, together with the price: quantity ratio $\frac{OY}{OX}$ may now be substituted in the elasticity of demand formula as follows: $E_p = \frac{OY}{OX} \times \frac{OX}{AY} = \frac{OY}{AY}$. By geometry, $\frac{OY}{AY} = \frac{BP}{AP}$. Therefore, $E_p = \frac{BP}{AP}$.

The rule for showing point-elasticity of demand in diagrams of demand curves may therefore be formulated as follows: draw a tangent to the given demand quantity *curve* at any point P, and divide the lower segment (BP) by the upper (AP). This rule is very useful in in-

FIGURE 8

POINT-ELASTICITY OF DEMAND

(To prove $E_p = BP/AP$)



terpreting conventional demand curves, which are often drawn as straight lines. At first sight one is tempted to confuse demand-elasticity with the *slope* of such lines. On the contrary, the point-elasticity changes throughout the entire length of a straight-line demand curve. The closer the chosen point (P'') is to A (i.e., the higher the price), the greater the elasticity. When it is half-way between the axes (P'), elasticity will be unity ($P'B/AP' = 1$). As the price approaches zero, the elasticity decreases rapidly. Incidentally, this indicates that when the conventional demand curve is drawn as a straight line, it is graphically realistic, for experience shows that the demand for most necessities tends to become more elastic at prices higher than customary and less elastic at prices below the usual ones.

It is now possible to combine some conclusions of Notes B and D regarding the interpretation of demand curves:

1. A tangent to an AR demand curve drawn on *arithmetical* axes makes an angle with the price axis which reveals the *marginal quantity,* or rate of change in quantity sold at that price. (In this it resembles the tangent to TR which was shown in Note A to reveal *marginal revenue.*)
2. A tangent to an AR demand curve drawn on *logarithmic* axes makes an angle with the price axis which reveals the *elasticity* of demand at that point.

9. Summary of Economic Principles Regarding the Effect of Price on Demand.—When other variables are held constant, the price-elasticity of demand may be measured in either of two ways: the change in total revenue or the quotient of percentage changes in price and quantity. Either method enables one to determine whether a given price change reveals demand to be "elastic" or "inelastic" by comparison with "demand-elasticity of unity." The second method gives a numerical measure of the degree of elasticity.

All finite changes in price reveal an *average* elasticity of demand between the two prices. By the use of infinitesimal changes in price, the elasticity of demand may be obtained for any given price or quantity point.

Demand and its elasticity may be shown by whichever one of four types of schedules or curves is the most useful in the problem at hand: AR (P) and Q, MR and Q, TR and P, or TR and Q. The most common demand curve is the AR type, which plots price on the vertical axis and quantity on the horizontal axis.

Chapter 5

DEMAND ANALYSIS DEMONSTRATED BY INDIFFERENCE CURVES

1. Outlay Indifference Curves.—Many of the demand concepts explained in the preceding chapters may be summarized by the use of a type of economic shorthand called indifference curves. These curves also suggest a few relationships not readily apparent in either the verbal approach or in one using utility curves. The argument may be divided into two main divisions: first, the individual's position relative to two commodities; second, his behavior when one of these two commodities is defined as money. This latter approach most closely resembles the spending problems treated in the demand analysis of earlier chapters.

The first type of indifference "curve" to be studied here is a straight line which may be called an "outlay indifference curve." It is based upon certain assumptions which seem peculiar at first glance, but whose significance will become clear as the argument proceeds. These may be outlined as follows:

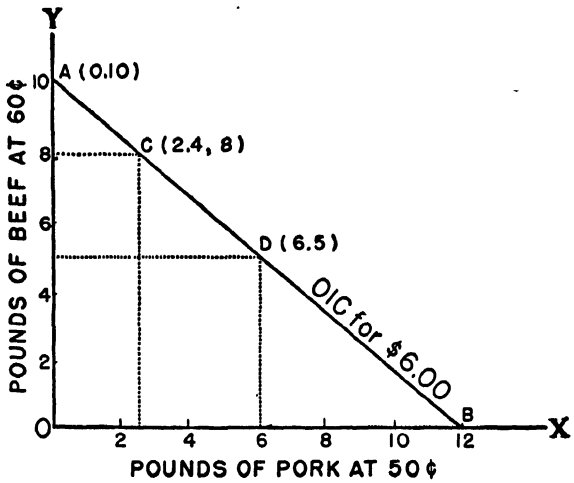
1. John Smith has $6 which he intends to spend on meat for his large family's Sunday dinner.
2. At the store he finds that beef roast costs 60 cents per pound and pork roast costs 50 cents per pound.
3. He does not care whether he gets a 10-pound roast of beef or a 12-pound roast of pork. Both meats are equally well liked by his family. The differences in bone, fat, etc., are such that the nourishment value of the two roasts is approximately the same in spite of their difference in weight.
4. He can, if he wishes, get some beef plus some pork instead of all of one and all of the other. It is really a matter of indifference to him what quantities of each he gets, provided only that the total purchase does not cost more than $6.

These assumptions are outlined in Figure 9. The vertical axis represents pounds of beef which Smith could buy with his $6 outlay. The horizontal axis represents pounds of pork. Point A on the OY axis shows Smith buying 10 pounds of beef. Point B on the OX

axis shows him buying 12 pounds of pork. The straight line connecting *A* and *B* shows all the possible combinations of beef and pork which he might buy for his $6. Point *C*, for instance, reveals the purchase of 2.4 pounds of pork for $1.20 and 8 pounds of beef for $4.80. At point *D*, Smith would buy 6 pounds of pork for $3 and 5 pounds of beef with his remaining $3. So far as outlay is concerned, it is a matter of indifference to Smith whether he spends his $6 for one combination or another on the line *AB*. This line is there-

FIGURE 9

EQUAL TOTAL OUTLAY POINTS ON AN OUTLAY INDIFFERENCE CURVE



fore called an *outlay indifference curve* or a "constant total outlay curve." It is defined as a line on which all points represent the same aggregate outlay for two commodities at constant prices but for varying quantities of each.

2. A Series of Outlay Indifference Curves.—If we now change the amount which Mr. Smith sets aside to spend for his Sunday meat supply, we will get different outlay indifference curves. A larger amount, say $8, will give a line drawn from 13.3 pounds on the *OY* axis to 16 pounds on the *OX* axis as shown between *E* and *F* in Figure 10. If Smith has only $3 to spend, his constant outlay line will be drawn between 5 pounds on *OY* and 6 pounds on *OX*, or *GH* on the adjoining figure. All points on *EF* represent a total outlay of $8, which is larger than any point on *AB* because that line shows an outlay of only $6. Similarly, movement from *GH* to *AB* means spend-

FIGURE 10

A SERIES OF *OIC*'s SHOWING DIFFERENT TOTAL OUTLAYS

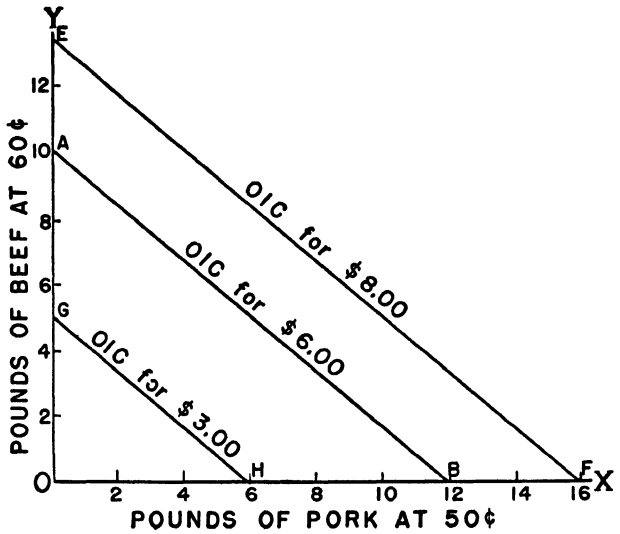
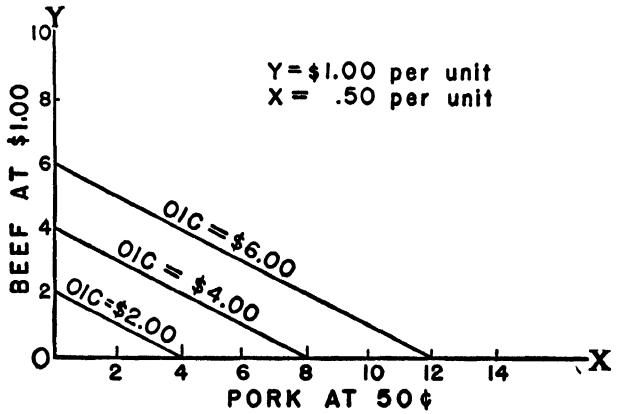


FIGURE 11

SERIES OF *OIC*'s WHEN PRICE RATIO IS 2:1



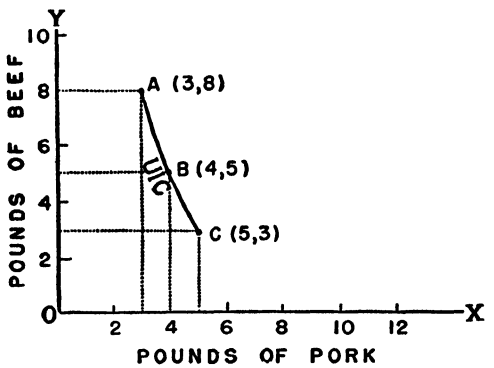
ing more money regardless of where one starts on *GH* or ends on *AB*. No point on any one line will be the same as that on any other, i.e., the lines cannot intersect. All three lines in Figure 10 and any other outlay indifference lines drawn with the same prices of beef and pork will be parallel.

If the prices of beef and pork are different from those of the original assumption, the slopes of the outlay indifference curves will be different. For instance, Figure 11 shows a series of such curves when the price of beef is $1 and that of pork is 50 cents. These lines are not so steep as those in Figure 10. The general rule is that the higher the price of the commodity plotted on *OY* compared to the price of the commodity plotted on *OX*, the less steep will be the lines. If the ratio of the prices were 10:1, the lines would be nearly horizontal. If the ratio were, say 1:20, the outlay indifference curves would be nearly vertical.

3. Utility Indifference Curves.—A second type of curve based on the same principle as the constant total outlay curve just described is one in which there is assumed to be constant total utility. Money outlay disappears from the picture and in its place is put the total utility or satisfaction derived from consuming beef and pork in different ratios. Smith is presumed to experience diminishing marginal utility as he increases his consumption of either of them. This has the corollary of *increasing* marginal utility if he *decreases* his

FIGURE 12

UTILITY INDIFFERENCE CURVE SHOWING THREE POINTS OF EQUAL TOTAL UTILITY



consumption of either beef or pork. Therefore, if Smith gives up some beef to get more pork, the marginal utility of beef rises and that of pork falls. But the *total* utility can remain the same, provided the amount of beef surrendered has a utility exactly equal to the amount of pork obtained.

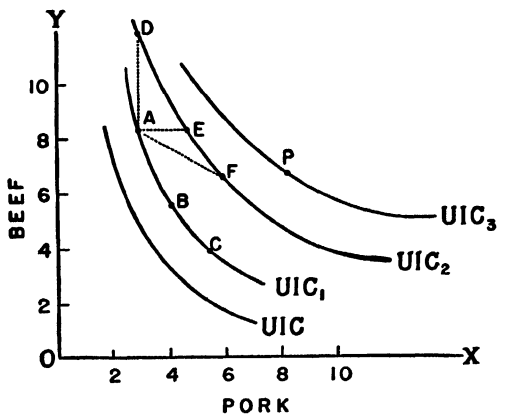
This line of reasoning furnishes the basis for drawing constant total utility curves like that shown in Figure 12. At point *A*, Smith

is consuming pork and beef in the ratio of 3 pounds of pork to 8 pounds of beef. At point *B*, Smith's meat consumption ratio has changed to 4 pounds of pork for every 5 pounds of beef. And at *C*, the ratio is 5 pork to 3 beef. By definition, Smith's total utility is the same at each point *A*, *B*, and *C*, that is, at each meat consumption ratio, 3:8, 4:5, and 5:3. In terms of utility, 3P + 8B = 4P + 5B = 5P + 3B. It is a matter of indifference to him which ratio he has. The line connecting *A*, *B*, and *C* is called a utility indifference curve. Intermediate points on this smoothed curve also have the same total utility. The argument may appear more logical if Smith is pictured as getting his meat free in whatever ratio he desires. Or Smith may be exchanging beef for pork on a barter basis without ever changing his total satisfactions.

4. A Series of Utility Indifference Curves.—Following the same line of reasoning developed in Section 2, we may now construct a series of utility indifference curves as in Figure 13. Each curve

FIGURE 13

A Series of *UIC*'s Each with a Different Total Utility



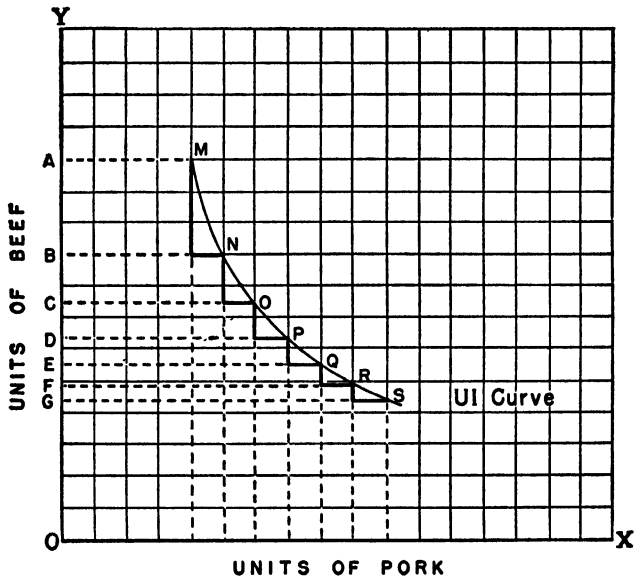
represents a different total utility comparable to the different total outlay curves of Figure 10. Movement *along any one curve* represents no change in total utility, but only a change in the ratio of consumption of the two meats. Movement *from one curve to another*, however, represents a change in total utility and also a change in the quantity of at least one of the two commodities. Thus in Figure 13, movement from *A* to *B* to *C* *along* UIC_1 does not change Smith's

total utility, nor from D to E to F on UIC_2. But a movement from A to either D, E, or F, puts Smith on a higher total utility curve. He would have no choice between positions A, B, or C, but he would definitely prefer to be at D, E, or F rather than at any point on UIC_1. Similarly, any point P on UIC_3 represents a position preferable to any point on UIC_2, UIC_1, or any other utility indifference curve to the left of or below UIC_3.

Since each utility indifference curve represents a different total utility from any other, the curves cannot cross. In this respect they resemble a series of outlay indifference curves. But utility indifference curves need not be parallel, and in fact they rarely are. They may have many different shapes, but are usually convex toward the origin like those in Figure 13. This means that each of the two commodities has diminishing marginal utility, as suggested above. In the indifference curve analysis, however, this term is dropped and in its place one speaks of a *diminishing marginal rate of substitution.*

FIGURE 14

DIMINISHING MARGINAL RATE OF SUBSTITUTION OF Y FOR X



The argument is that Smith may be considered as bartering Y for X (beef for pork). For each unit of pork obtained in exchange, Smith gives up a diminishing amount of beef. This may be seen by drawing a utility indifference curve on a grid of cross lines as in Figure

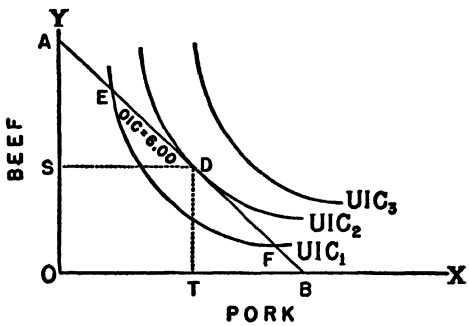
14. Moving from left to right by equal intervals represents the acquisition of equal additional units of pork. When these points are marked on the curve, as *M, N, O, P,* etc., the amount of beef that must be given up by Smith to keep his total utility constant may be seen by drawing horizontal lines from each of these points to the *OY* axis. The distance between these lines becomes smaller as one moves from *M* to *N* to *O* to *P*. This shows diminishing amounts of *Y* substituted for equal increments of *X* or, as stated above, a diminishing marginal rate of substitution of *Y* for *X*. By reversing the process one may also demonstrate a diminishing marginal rate of substitution of *X* for *Y*. Both relationships are based upon the diminishing marginal utility of the object whose rate of consumption is being increased, and the increasing marginal utility of the commodity whose rate of consumption is being decreased.

5. Maximum Total Utility for a Given Total Outlay.—It is now desirable to bring together the argument of the first four sections of this chapter. First we shall consider the problem of securing the maximum total utility for a given total outlay.

Assume that Smith has $6 to spend for beef and pork as in the first illustration. Retain also the assumption that beef and pork cost respectively 60 cents and 50 cents per pound.[1] Change the assumption that he has no particular preference for more of the one rather than the other. Assume instead that he has a preference schedule that may be represented by a series of utility indifference curves as shown by the three in Figure 15. The outlay indifference curve *AB* cuts UIC_1 in two places *E* and *F*, is tangent to UIC_2 at *D*, and does not intersect UIC_3 at all. Smith would prefer to be on the highest possible utility indifference curve, but he is held back by his limit of $6 to spend. Therefore, he cannot attain the total utility level of UIC_3 at all. On the other hand, he would not choose the beef-pork

FIGURE 15

MAXIMUM TOTAL UTILITY FOR A GIVEN TOTAL OUTLAY



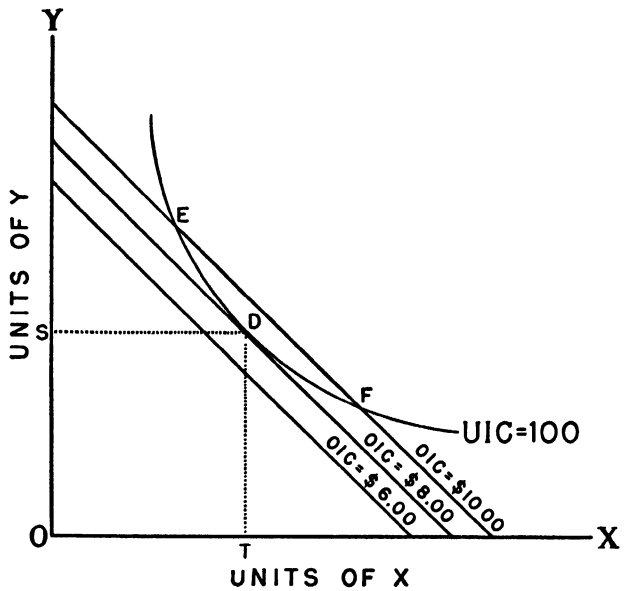
[1] Strictly speaking, an *OIC* is an exercise in arithmetic and need not be associated with Smith or any other particular individual. On the other hand, *UIC*'s must be associated with a specific individual.

combination represented by either point E or F because they are on UIC_1, which has a lower total utility than UIC_2, which is within his spending capacity at one point, D. That is, the *point of tangency* between the given OIC and the highest possible UIC solves our problem. Point D reveals the maximum total utility that it is possible for Smith to obtain when he spends \$6 for beef and pork at the same time, and at the prices given. This point also indicates, by perpendiculars to the axes, the quantities of beef (OS) and of pork (OT), which Smith would buy.

If it were possible to define utility in terms of definite units, a similar demonstration could be given with three OIC's and one UIC.

FIGURE 16

MINIMUM TOTAL OUTLAY FOR A GIVEN TOTAL UTILITY



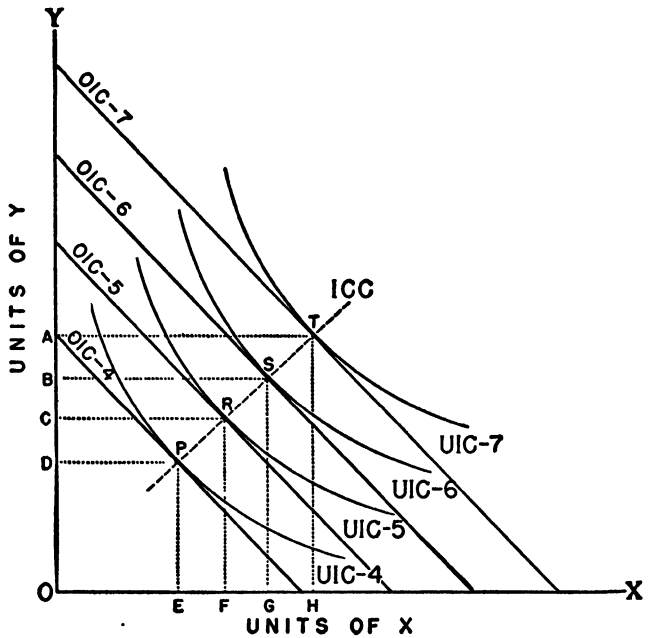
That is, if one could define for Smith a UIC representing, say, 100 utils, one could then demonstrate the minimum outlay required to attain a satisfaction of that amount (cf. Figure 16). It would be the point of tangency between UIC_{100} and the lowest possible OIC. UIC_{100} would intersect higher OIC's at two places, E and F, thus indicating the possibility of moving along UIC_{100} to an equally satisfying point, D, which would be on a lower OIC, i.e., which would require less total expenditure. The least outlay, \$8, would be shown

by the tangency solution similar to that of the maximizing problem treated in the preceding paragraph.

6. Income Consumption Curve: A Series of Tangency Points on Parallel Outlay Indifference Curves.—The next step is to plot on the same axes a series of *OIC*'s and *UIC*'s for a given individual and given market situation and to find their respective points of tangency. This is shown in Figure 17. Movement from *P* to *R*, *S*, and *T* may be interpreted as Smith's reaction to an increase in income which gives him more to spend on that Sunday meal. When his income is small, the best quantities of *Y* and *X* for him to buy are

FIGURE 17

INCOME-CONSUMPTION CURVE: A SERIES OF TANGENCY POINTS



shown by *OE* and *OD*, which are called the coordinates of point *P*. As his income increases, he should increase his purchases to *OF* of *X* and *OC* of *Y* as shown by point *R*, and so on. The line *PT* which connects these tangency points is called an income consumption curve, *ICC*. More logical terms would seem to be "outlay consumption curve" or "optimum outlay curve," but custom dictates "income consumption curve."

One must note the assumption that the preference pattern of Mr. Smith remains unchanged as his income increases. The *UI* curves define a surface which does not alter.[2] Only the *OIC*'s change as new ones appear further to the right. Their shape and slope is held constant by the unchanging prices of *Y* and *X*. The unrealism of the initial assumption does not invalidate the analysis, and it may not be very faulty if the increase in income is small and slow. Two practical applications of income consumption curves in economic analysis are best deferred to Section 8, where the assumption of two commodities is changed to one commodity and money.

7. The Price Consumption Curve: A Series of Tangency Points on a Fan-Shaped Series of Outlay Indifference Curves.—Returning now to the concept of outlay indifference curves discussed in the last part of Section 2, we may construct a fan-shaped series of curves by holding one price constant and varying the other. Beef may be held constant at 60 cents per pound while pork is assumed to sell at prices ranging from the first price of 50 cents down to 10 cents per pound. The fan of *OIC*'s shown in Figure 18 is the result. Each of these lines has a point of tangency with a utility indifference curve. When these points of tangency, *K, L, M,* and *N,* are connected, a price-consumption curve for pork results.

A *PC* curve of this type shows two things. First, it reveals the effects of a falling price of pork upon the demand for pork.[3] Second, and more important, it shows the effects of a changing price of pork upon the demand for beef. This is the *cross-elasticity* of demand problem discussed in Chapter 2, Section 5. If the *PC* curve slopes downward as it moves to the right, *Y* is a substitute for *X*. Smith buys less of *Y* when he buys more of *X*. If the *PC* curve is horizontal, *Y* is a good which is not associated with *X* in any significant way. If, however, the *PC* curve slopes upward, *Y* is revealed to be a complement of *X*. Smith buys more of it when he buys more of *X*. The ability to use indifference curves to demonstrate such relationships is one of their major advantages over ordinary demand curves. The principle, however, is not new and has already been stated in non-graphic terms in Chapter 2, Section 5.

8. Price Consumption Curves Become Demand Curves When Money Is Substituted for One Commodity.—It is now desirable to

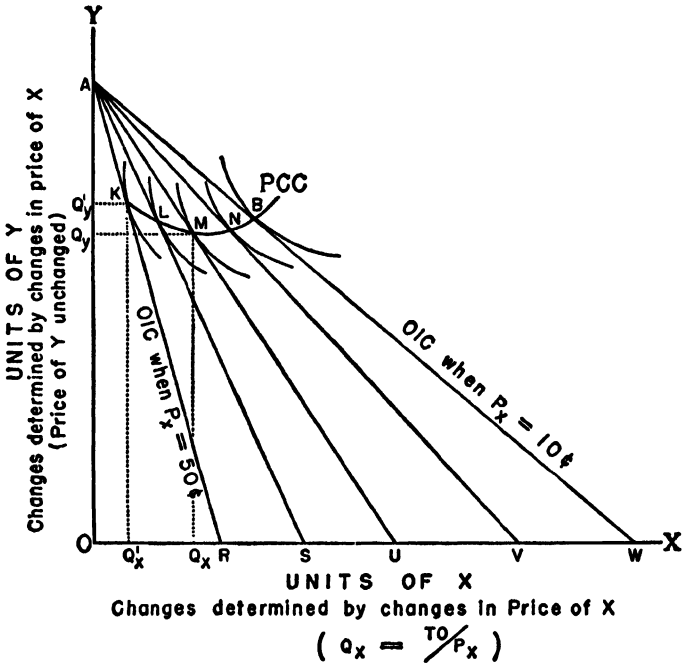
[2] Of course one might define the pattern of *UI* curves as one which represented Smith's changing preferences as his income increased, but that is contrary to the customary assumptions used here.

[3] Comments on the irreversibility of demand curves and similar points apply just as much to *PC* curves as they do in the demand analysis of Chapter 3, Section 6.

enter the second half of the indifference curve analysis. This is done by substituting money for the commodity whose quantity has been measured on the OY axis. To keep the discussion in commodity terms, OY sometimes is said to represent "all other commodities"

FIGURE 18

CROSS-ELASTICITY OF DEMAND REVEALED BY PRICE-CONSUMPTION CURVE FOR TWO COMMODITIES



than the one measured on OX. This has the advantage of indicating that money may be spent for many other things and is desired because of its ability to serve many uses.[4]

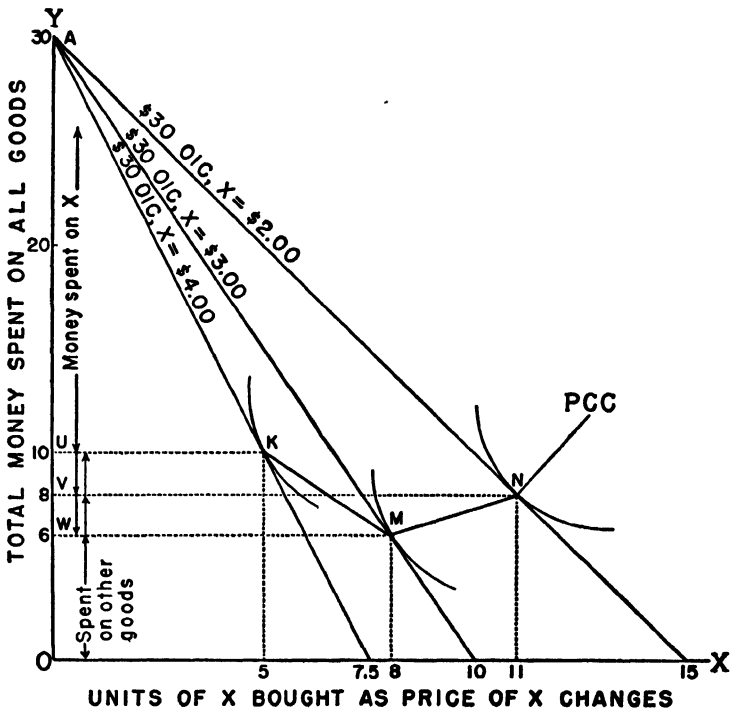
When OY represents money, there will be some point on OY such as A which shows the total amount of money which Smith has to spend in the period when his purchases of X might be influenced by changes in the price of X. The fan-shaped series of OIC's will be much the same as in Figure 18 although the OY scale will have to be changed to money units. For most items of low price, the "ribs" of the "fan" will be close together. To make relationships clear, Fig-

[4] But note that, in Fig. 19, OY is the total *money* spent on *all* commodities, not the quantity of all *other* commodities.

ure 19 therefore assumes an item whose price is high relative to disposable income. In this hypothetical illustration, a fall in the price of X from \$3 to \$2 per unit will increase Smith's demand for X from 8 units to 11 units when Smith's disposable income is \$30.

FIGURE 19

PRICE-ELASTICITY OF DEMAND SHOWN BY A ONE-COMMODITY PRICE-CONSUMPTION CURVE



A further attribute of the one-commodity price-consumption curve is its ability to reveal the price-elasticity of demand by its slope, which the conventional demand curve can do only when plotted on logarithmic axes (cf. Chapter 4, Note C). Since OA represents the total amount of money to be spent on OX and other commodities, the distance from O to U, V, or W represents the amount spent on those other commodities. By subtraction, the distance from A to U, V, or W shows the amount spent on X. When a fall in the price of X from \$3 to \$2 causes the PC curve to rise from M to N, it also shows the total amount spent on X decreasing from \$24 (AW) to \$22

(AV). But a decreased total expenditure on X when its price is falling is an evidence of inelastic demand. Between \$4 and \$3, however, a falling price raised the total amount spent on X from \$20 (AU) to \$24 (AW).[5] The generalization may be made, therefore, that when a PC curve slopes downward to the right (KM), the price-elasticity of demand for that good is greater than unity. When the PC curve is horizontal, demand has an elasticity of unity. And when the PC curve slopes upward to the right (MN), the price elasticity of demand is less than unity.

This brings us to a theoretical advantage of indifference curve analysis. It allows for changes in the marginal utility of money at the same time that it considers changes in the marginal utility of the goods bought with that money. The conventional demand curve is based upon a diminishing marginal utility schedule for the commodity. In order to translate utils into money, each util must equal an unchanging number of cents or dollars. The marginal utility of money must be constant, an assumption which is probably near enough to the truth in most cases to be acceptable. But there are cases of small amounts of available funds or purchases of high priced items which require the use of the more general assumption of the diminishing marginal utility of money. Indifference curves fit any type of case.

9. Superior and Inferior Goods Revealed by One-Commodity Income-Consumption Curves.—Another advantage of indifference curves is found in their ability to demonstrate the income-elasticity of demand. This is done by drawing a one-commodity income-consumption curve in which the vertical axis represents disposable income as in Section 8. The most common type of ICC is shown in Figure 20 which resembles Figure 17 except for the change on the OY axis. The upward-to-the-right slope of this ICC indicates that as Smith gets more money to spend he buys more of this commodity (the "Q" points move to the right). According to the argument and terminology of Chapter 2, Section 6, this means that the commodity is a "superior good" so far as Smith is concerned.

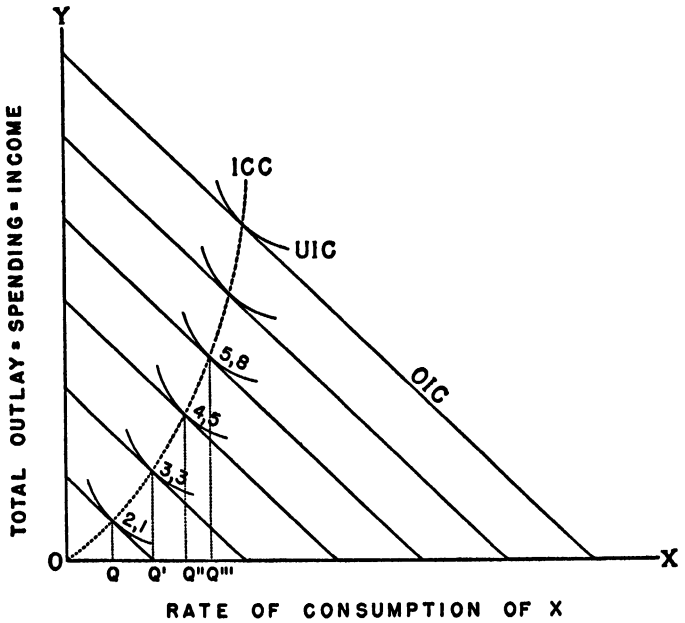
Under some circumstances Smith might buy *less* of a good when his income rose. Such a commodity is called an "inferior good" and may be shown diagrammatically by an ICC which slopes to the left (see Figure 21). At other times a commodity may be "superior"

[5] \$30 minus \$ 6 equals \$24 equals 8 times \$3.
\$30 minus \$ 8 equals \$22 equals 11 times \$2.
\$30 minus \$10 equals \$20 equals 5 times \$4.

for income increases at low levels and then become "inferior" as incomes rise still higher. This possibility is diagrammed in Figure 21 where an increase in weekly income from $10 to $25 per week makes the consumption of X increase. From $25 to $30 there is no change.

FIGURE 20

INCOME-CONSUMPTION CURVE (for one commodity)



But above $30 per week X becomes "inferior" and the *ICC* begins to slope to the left. A practical illustration might be found in the purchase of motion picture tickets to a certain theater. At higher income levels Smith might take his family to a more expensive theater, at least part of the time.

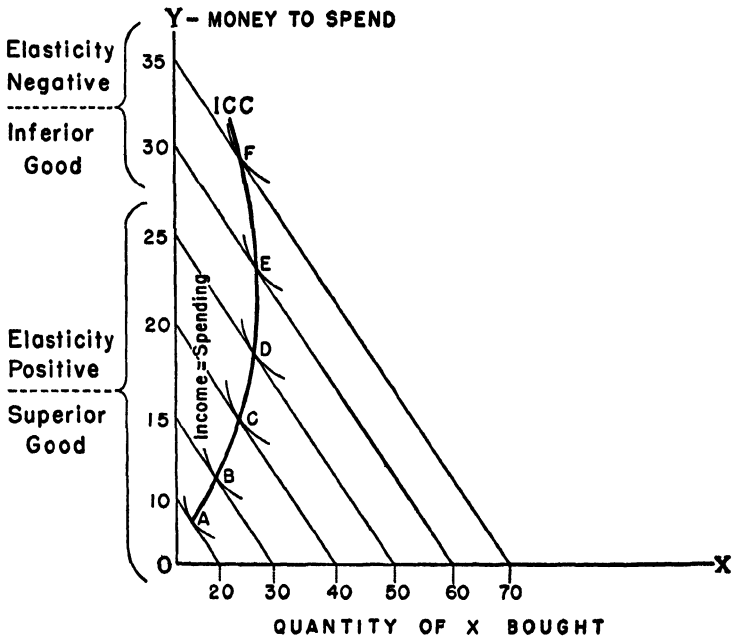
If one-commodity *IC* curves are drawn with the variables plotted on log-log axes, the income elasticity of demand may be read at a glance. An *IC* curve sloping upward to the right at 45 degrees will represent positive income-elasticity of unity because the percentage change in the quantity of X bought will equal the percentage change in other commodities bought.[6] If the slope of the *ICC* is less than

[6] If the percentage change in purchase of both X and Y occurs at the same rate, the change for total income will be of the same percentage since the total spent on X plus "all other commodities" (Y) must by definition exhaust the amount of income being spent.

45 degrees, the income-elasticity may be said to be greater than unity because Smith will be increasing his purchases of X more rapidly than he increases his purchases of other commodities. Similarly, a slope steeper than 45 degrees means income elasticity less than unity,

FIGURE 21

SUPERIOR AND INFERIOR GOODS SHOWN BY INCOME-ELASTICITY OF DEMAND (via *IC* curve)



diminishing to zero when the *ICC* is perpendicular, and becoming negative when the *ICC* turns backward to the left and indicates the "inferior good" type of response.

10. Income Effect and Substitution Effect of Price Change.— For some purposes it is convenient to separate into two parts the effect of the rise in price of a given commodity. The first is the tendency to decrease the consumption of the good whose price has risen in the same way that a person would have to do if his income were decreased. This is called the "income effect" and may be diagrammed by a movement from the original *OIC* to a hypothetical *OIC* below it. The second part is the tendency to decrease the consumption of a good and buy more of other goods when the first be-

comes more expensive relative to the latter. This may be diagrammed by a movement along a utility indifference curve from one point to another. As explained in Section 3 above, such a movement does not change a person's total utility from the commodities consumed. It does represent a change in the *OIC* whose tangency would determine first the one point and then the other. The substitution effect of a price increase is always negative. The income effect is negative in most cases, but in rare cases may be positive as in the case of inferior goods. In extremely rare cases, if the income effect is positive, it may be so great as to offset all of the negative substitution effect. When that happens, the good whose price has risen may be called an inferior good on two scores—income change and price change.

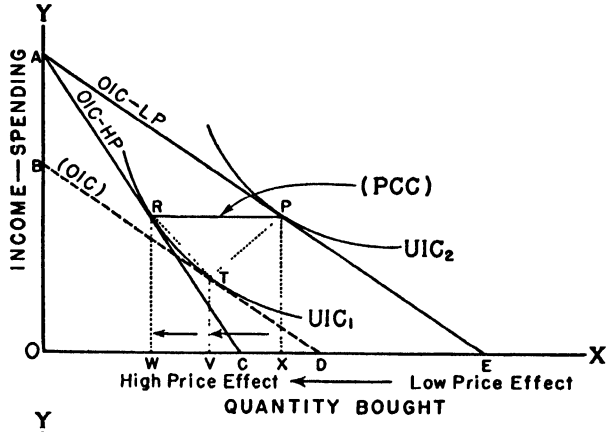
The accompanying diagrams of Figure 22 show three possible cases. In each of them the income effect is indicated by the dotted line *PT* which resembles the income consumption curves of Figures 17 and 20. The substitution effect is shown by the dotted line *TR*. The price consumption curve in each case is represented by the solid line *PR*, like the *PCC*'s of Figures 18 and 19. In the first two cases, movement from *P* to *R* represents a *decrease* in the total amount of *X* purchased as its price *increases*. These represent the most common type of market response to price change. The third case shows the highly exceptional "price inferiority" case of a rise in price causing an increase in the total amount purchased. Even the second case is rare. Each of the three diagrams is intended to represent a different commodity, and therefore the *UI* curves have a different pattern. In each diagram the broken line *BD* representing the hypothetical *OIC* is drawn parallel to the line *AE*, which represents the initial price-income situation and which determines by tangency with the highest *UIC* the starting point *P*.

11. Summary and Evaluation.—What has been gained by the indifference curve approach to demand? In what respects is it superior to, inferior to, or no better than the conventional demand curve based on diminishing marginal utility? Both diagrammatic approaches may also be compared with (1) a strictly verbal analysis like much of that in the earlier chapters, and (2) a purely mathematical type of functional notation. Two criteria may be applied: clarity in illustrating principles, and realism of assumptions.

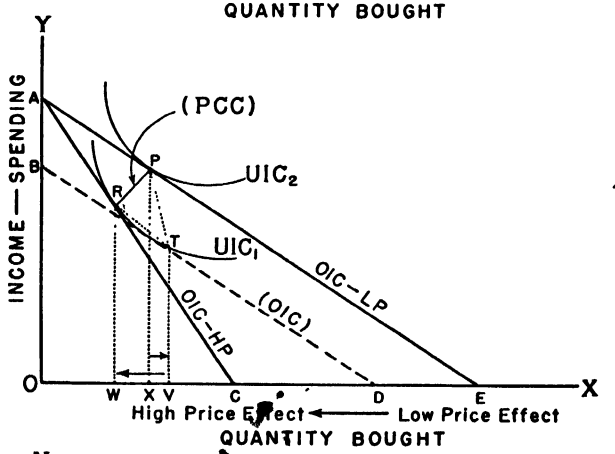
I. *Clarity*—Indifference curves and their derivatives (*PCC*'s and *ICC*'s) demonstrate certain concepts better than do conventional demand curves:

FIGURE 22

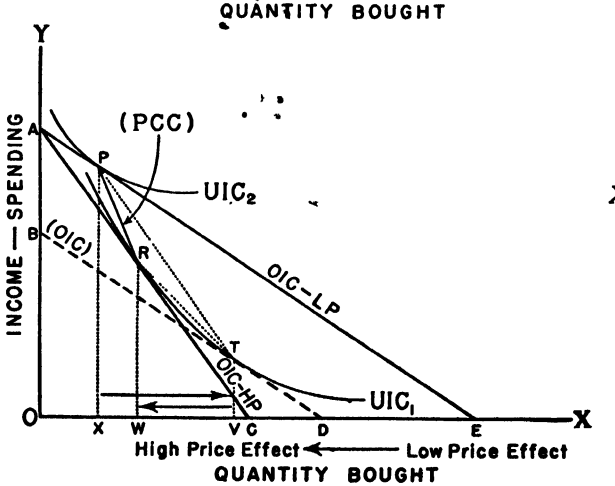
INCOME EFFECT (XV) AND SUBSTITUTION EFFECT (VW) OF PRICE INCREASE



I. *Superior Good*
XV neg. + VW neg.



II. *Superior Good*
XV pos. < VW neg.



III. *Inferior Good*
XV pos. > VW neg.

A. The income-elasticity of demand: effects of changes in income upon the quantities demanded by individuals. There are two types with varying degrees:
   1. Superior goods
   2. Inferior goods

B. The cross-elasticity of demand: effects of changes in the price of one commodity upon the demand for another. There are two types with varying degrees:
   1. Substitute goods
   2. Complementary goods

C. The twofold effect of price change upon demand:
   1. The income effect
   2. The substitution effect

II. *Realism*

A. *UIC*'s are *more* realistic in the following respects:
   1. May make allowance for the increasing marginal utility of disposable money income as larger amounts of it are spent for a given good
   2. Do not require the use of explicit or implied specific measures of utility, such as utils, but only *relative* preferences. This is often described as using ordinal numbers (1st, 2nd, 3rd, . . .) to replace cardinal numbers (1, 2, 3, . . .). The latter must be related to utility units, but the former need not be.

B. *UIC*'s are *less* realistic in the following respects:
   1. Imply that consumers evaluate relative *pairs* of commodities more commonly than they evaluate rival quantities of the same good or single units of rival goods
   2. Tend to obscure the unrelatedness of commodities which compete for the consumer's dollar, but are not otherwise substitutes (food and clothing, or salmon and shirts)

C. Indifference curves and their derivatives represent *no gain* in the following respects. One should note that these defects also apply to the conventional demand curves used to illustrate points of the theory of consumer demand. The verbal method of presenting the material *may* avoid many of these shortcomings of the diagrammatic or mathematical presentations.
   1. Overstress the importance of choice and conscious planning in most consumer expenditures as opposed to the force of habit, impulse, and regimentation

2. Overstress choices between some and more of a homogeneous good rather than between this and that heterogeneous good
3. Overstress the bases of individual demand schedules, although market behavior on the supply side is more often determined by expectations about collective demand schedules
4. Imply knowledge about preference or utility schedules at prices and incomes far beyond the probable and customary ranges
5. Fail to allow for variations in demand which result from differences in the rate of change of the independent variable (price or income), its direction of change, and the magnitude of successive changes (cf. Chapter 3, especially Sections 6 and 9)
6. Fail to allow sufficiently for product differentiation, monopolistic activity, and institutional controls [7]

[7] An excellent appraisal of indifference curves may be found in *The Theory of Consumer Demand* by Ruby Turner Norris, New Haven: Yale University Press, 1941, especially Chapter 3.

Chapter 6

SUPPLY AND PRICE—GENERAL CONSIDERATIONS

1. Statement of the Problem.—Transactions prices are the joint result of demand and supply forces. Sometimes one of these forces seems more potent than the other. But a transaction is an exchange of money for goods, and therefore both a buyer and a seller must be involved. Both demand and supply must be present. To paraphrase Alfred Marshall, it is as idle to think of demand or supply alone determining a transaction price as to conceive of either the upper or the lower blade of a pair of shears doing the cutting by itself. The demand side of the market has received considerable attention in earlier chapters. We now turn to an examination of the supply side. It is perhaps not more difficult than demand side analysis, but it has received more attention in economic literature. The following questions must be asked and answered. The first two lay the foundation for the others, which will be studied in later chapters.

1. What are the different concepts of supply?
2. In what sense and in what situations is *supply a cause* of transaction price?
3. What are the *causes of supply* quantities and supply prices?
   (a) What are the institutional determinants of supply prices? (Ch. 7)
   (b) How may cost estimates influence supply prices? (Ch. 8)
   (c) How may production policies influence supply quantities? (Ch. 9)
   (d) How may selling policies influence supply prices? (Ch. 10)
   (e) What is the supply side significance of various types of competition? (Chs. 11 and 12)
4. What are the *effects* of supply quantity changes upon production cost and price in different time periods? (Ch. 13)
5. How can individual price fluctuations be explained in terms of demand and supply changes in typical situations? (Ch. 14)

2. Supply as a Quantity Offered for Sale.—Supply generally means the quantity offered for sale. Sometimes supply refers to the

quantity offered at a certain stated or minimum price known as *the reservation price* or *supply price.* At other times it signifies the quantity offered for whatever it will bring, i.e., without any minimum price. Usually the reference is to the quantity offered by a group of sellers, whose size may be large or small. It may refer also to the offering of a single seller.[1] The time of supply is generally revealed by the context, but should be made explicit if confusion is likely to develop. The most common time period is the given instant of offering, or any time period shorter than that required to replenish (usually, to reproduce) the good offered. Finally, supply must be understood as referring to an amount at a certain place, although that place may be of any size, small or large, even the whole world. With agricultural products, for instance, the "supply" is often the entire amount produced by all the commercial growers in a certain region during a given crop year.

3. Supply as a Schedule or Series of Quantities.—Economists also use the term supply to refer to *schedule supply,* the *series* of quantities which individuals would sell at a *series* of prices. This corresponds with the concept of schedule demand explained in Chapter 2. Again the contrast should be noted between the schedule idea of supply and the market idea. The schedule is made up of a series of quantity-price pairs. Each pair represents a hypothetical market supply, i.e., a quantity offered at a price. All the comments made in Chapter 2 about demand schedules apply here. Supply schedules represent what someone *thinks* would be the amounts offered at the series of prices he chooses.[2] Only one of the quantity-price pairs can ever be known with certainty, and even that may be only an approximation. When a person himself is offering supply at a given time for a given price, he can describe *his* supply with precision, such as one used 1938 Ford coupé for $600. But the same person cannot now state definitely how many others are offering 1938 Ford coupés for $600 in a large city, or in the country as a whole. He must estimate that quantity. And if he wants to say how many *would be* offered at $700 or at $800, other estimates of still less accuracy would have to be made. The quantities would depend upon how the price is reached, whether by a rising price or a falling one, whether other prices rose at the same time, or did not, etc. Even the concept "1938 Ford coupé" is hard to define, since such cars may be in various stages of decrepitude.

[1] Cf. the concepts of demand outlined in Chapter 2, Section 2.
[2] They may also be historical, but these are less useful.

4. Supply in the Schedule Sense and in the Market Sense.—Nevertheless, the idea of supply in the schedule sense is a very useful one, particularly for staple commodities. The concept is indispensable when one wishes to compare two situations in which both supply price and supply quantity have changed or are expected to change. Thus, if American copper miners produce and offer for sale 500,000 tons in one year when the average price is 12 cents per pound of refined copper and 600,000 tons when the price in the next year is 15 cents, one can say only that "the market supply increased." One cannot say whether there was an increase in supply in the schedule sense until he knows what the copper mines *would* have produced in the first year at 15 cents a pound. If that amount were 575,000 tons, then the second year reveals an increase both in the market supply and in the schedule supply. Such a situation is revealed by the following figures.

HISTORICAL SUPPLY QUANTITIES MARKETED BY AMERICAN COPPER MINES

In Year 194X		In Year 194X plus 1	
Supply Quantity	Supply Price	Supply Quantity	Supply Price
500,000 tons	12¢	? tons	12¢
? tons	15¢	600,000 tons	15¢

HISTORICAL (AND HYPOTHETICAL) SUPPLY SCHEDULES OF AMERICAN COPPER MINES

In Year 194X		In Year 194X plus 1	
Supply Quantity	Supply Price	Supply Quantity	Supply Price
500,000 tons	12¢	(520,000 tons) *	12¢
(575,000 tons)	15¢	600,000 tons	15¢

* Added for completeness, though not absolutely necessary to argument.

The foregoing argument may now be summarized by saying that a change in supply *at any given price* represents a change in the *schedule sense*. This might be caused by a reduction in cost, by the entry of new producers, by a change in selling policies, etc. A change in supply which is *caused* solely *by a change in price* is a change in the *market sense*.

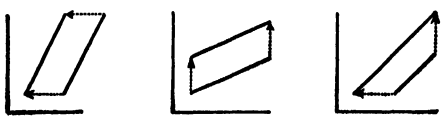
There is one other way of looking at supply which may now be examined. When the same quantity is offered at a higher or a lower

price than before, the supply may be said to have changed in the schedule sense. If the owner of the 1938 Ford coupé first offers it at $600 and then changes his mind and demands $700, the supply *at $600* has dropped to zero so far as he is concerned. Therefore, the price increase for the fixed quantity represents a decrease in schedule supply. This is a very important concept which is often misunderstood. If a gasoline station today raises the posted price of first-grade gasoline from 20 cents per gallon to 22 cents, the supply has decreased even though the station owner is willing to sell just as much as he did yesterday. Similarly, if some organized workers decide to refuse a renewal of their old contract at $1.25 per hour and demand $1.40, the supply of this type of labor has decreased regardless of the desire of individual workers to work as many hours as before. In short, a decrease in schedule supply occurs *either* when the quantity offered declines at a given price or when the requested price rises for a given quantity.[3] For increases in schedule supply the changes would be in the opposite direction.

5. Supply as a Price Asked by Sellers.—Just as the supply quantity is that amount which is offered at a given price, so the *supply price* is the money demanded by a seller for a given quantity offered for sale. The supply quantity concept often refers to the quantity that people are willing to sell at a given price *or higher*. Occasionally the supply price concept refers to a given quantity *or lower* which people are willing to sell at a given price. Thus, copper sellers will be willing to take more than 15 cents per pound for their 600,000 tons if anyone offers it to them. A gasoline station owner posting a price of 21 cents wants to sell all his station's daily capacity, say 5,000 gallons, but he is willing to sell any smaller amount if he cannot reach the maximum.[4]

6. Supply Price as a Cause of Transaction Price.—The foregoing discussion has shown that the transactions price may be a cause of the supply quantity. Our attention now shifts to three general situations in which the supply price may be a cause of the transac-

[3] On a diagram, the movement of a supply curve to the left is the same as moving it upward:



[4] Of course, a station with that capacity cannot long continue in business if it sells only 50 gallons daily.

tions price. In the first, the supply price of an individual seller determines the transactions price at which each exchange takes place. Thus the Standard Oil Company sets the price at which its service stations sell Standard gasoline. This is the market situation most frequently encountered by individual consumers. It will be studied more fully in Chapter 7.

In other situations, the supply price of an individual seller is only one of several proximate causes of the transactions price. For instance, when bargaining occurs between buyer and seller, each names at first a more favorable price than he finally agrees to accept. During the bargaining interval there may be a secret minimum reservation price in the seller's mind, but his declared minimum changes as the higgling proceeds and he need never reveal his "actual" minimum.

A similar situation exists when there are many individual buyers and sellers whose demand and supply prices become revealed as the bidding and offering proceeds to the point where sales take place. At this point the aggregate demand quantity sought at a certain price equals the aggregate supply quantity offered at that price. The sellers are those willing to take that price or less. If the price determined by the *balancing of aggregates* is 10 cents and some seller was willing to have accepted 8 cents, the same person would have taken 9 cents, 10 cents, or any price higher than 8 cents. Therefore, he is an included seller. On the other hand, sellers who wanted 11 cents or 12 cents are excluded and do not make a sale at this time. In other words, some supply prices influence the transactions price, and some do not.

The supply price of sellers does not determine the transactions price at all in three situations. In the first, the seller has no reservation price whatever and offers his goods for auction sale to the highest bidder. In the second, the buyer sets the transactions price much like the seller in the first case described in this section. Sellers either accept that price or make no sales at all. In the third case, the government or other third party sets the price. Both buyers and individual sellers must accept it if they wish to trade. In this respect it resembles the price set by the balancing of aggregates described in the preceding paragraph.

7. Causes of Supply Prices.—Supply prices are determined in many ways. The first group of causes may be designated as institutional since they include such things as laws, customs, trade agreements, and attitudes. Chapter 7 explains the leading cases in this category and examines also some effects of changes in the determin-

ing institutions. The degree of price competition among sellers is particularly important since it differs from industry to industry and changes frequently. Chapters 10, 11, and 12 will shed further light on this subject.

The second group of determinants of supply price centers around the analysis of cost of production. Attitudes toward cost as a price determinant are also important. Much price theory is written in terms of the cost-price-quantity relationships which would maximize profit under various conditions. If most businessmen followed price policies based on these theories, their actions would deserve to be studied first in a book of this type. But most supply prices spring from institutional origins, including rule-of-thumb markup percentages which use cost as a base. Therefore, cost-price theory has been deferred to Chapter 8, following the institutional-price theory of Chapter 7.

8. Causes of Supply Quantities.—The supply quantity offered by any one seller will depend upon at least three things: (1) the price he expects to receive, (2) the quantity he can buy or produce at a cost less than his selling price, (3) the quantity he has on hand from prior production or purchase. The determinants of a seller's rate of production will be analyzed in Chapter 10.

The supply quantity for a group is a function of (1) the supply schedules of individual sellers, and (2) the number of sellers. The total supply quantity is cumulative upward through a scale of rising prices.[5] For instance, at the lowest price there may be only one seller, A, in the market. A slightly higher price may induce seller B to enter the market. The total supply at the price then becomes the amount offered by A plus that offered by B, and so on.

9. Summary.—The supply side of the market is very complex and will require several chapters for its elaboration. Supply generally means the quantity offered by suppliers, and is best thought of as the quantity offered at a certain price. The number of sellers, the time, and the place should be identified. Supply schedules may be constructed to show what some one thinks would be offered at a short series of prices. Supply changes in the schedule sense must be distinguished from supply changes in the market sense. Supply also may be considered as the minimum or supply price of a given quantity. This supply price may influence the transactions price in many situations. These include bargaining, the balancing of aggregates, and

[5] Cf. the explanation of collective demand in Chapter 3, Section 4.

the action of certain sellers in fixing a price and taking whatever volume of sales can be made at that price. When the buyer or the government sets the transactions price, or when auction sales occur, supply prices have no immediate effect. Supply prices have institutional and cost determinants. Supply quantities are related to output plans of producers. All of these points will be elaborated in subsequent chapters.

Chapter 7

INSTITUTIONAL DETERMINANTS OF SUPPLY PRICES

1. Supply Prices When Institutional Forces Dominate.—In Chapter 6 the term *supply price* was defined as the price at which a person declares himself willing to sell something. The supply price for certain neckties at the Honest Haberdashery is marked on their tags as $2. This is the most common situation. The marked price is the minimum price at which the seller will sell. Therefore, it becomes also the transactions price, the price at which sales take place. In a few exceptional cases, the seller's actual minimum price at that time is less than his posted or asking price. Pawnshops and other stores selling second-hand merchandise often fall in this group. In some foreign countries retail bargaining is the rule, not the exception as in the United States today. And American manufacturers sometimes shade their list prices in a form of secret price competition.

The present chapter will examine six situations in which sellers determine what supply price to ask. It will assume that the supply prices announced to potential buyers are the actual minimum prices at that time. Occasionally the synonym *reservation price* will be used. The following situations are called "institutional" because the economic environment of the seller determines his supply price decisions with very little need for conscious and independent price calculation on his part:

1. No reservation price: the auction situation
2. Supply price determined by custom
3. Supply price determined by law
   (a) Specific prices
   (b) Fixed maximum or fixed minimum prices
4. Resale supply price determined by prior seller
5. Imitation
6. Markup determined by custom, or by law

The final case to be considered in this chapter requires some calculation of unit cost before the application of the institutionally determined markup. Therefore, it serves as a very appropriate bridge to the next chapter on cost as a determinant of supply price. The

cases examined in Chapter 8 deal with conscious calculations of profit-maximizing supply prices under assumed conditions of demand, cost, and competition. Institutional forces are present, of course, but are less dominant than in the situations described here.

2. Motives Underlying the Determination of Supply Prices.— In fixing their supply prices, sellers are influenced by a variety of motives and objectives. In capitalistic society the "profit motive" is usually assumed to be dominant. Unfortunately, this phrase has many connotations. It has been used as a synonym for material self-interest so as to include the worker's quest for higher wages. More narrowly, it applies to the desire of business managers to maximize the net income of their enterprises. But even that statement is indefinite as to scope and time. It fails to indicate whether profit maximization is sought for dealings in a particular product, for all goods produced by a certain plant, for everything sold by the enterprise, or for the managers and their friends who may have connections with other enterprises. Actions may be quite different as attention shifts from one of these profit objectives to another. So, too, with time. A profit-seeking seller may determine his supply prices according to their effect upon his current profits, his profits next month, next year, or over the next decade. The further ahead he looks, the more factors he will usually take into consideration in shaping his price policy. He is also more likely to do things which sacrifice present profits in hopes of securing larger profits in the future.

Profits, however, are usually a means to an end, not an end in themselves. The ultimate motive is the satisfaction or happiness of the business owner or manager. This basic objective not infrequently conflicts with profit maximization. Thus, when sellers are making what they consider to be a satisfactory profit, many of them are content to "let well enough alone." They do not engage in constant calculation and change in an effort to squeeze the utmost possible profit out of the situation. They accept the prices dictated by custom, law, or monopolistic pressure. This method reduces the number of variables which managers have to worry about. Some enterprising individuals prefer to depart from institutionally determined prices. They get satisfaction from the thrill of experiment, or of opposing the crowd. Even the most lethargic sometimes have to choose between retaining, raising, or lowering their prices. But even in these cases, impulse often outweighs cold, careful calculation. These are some of the reasons why this book begins with the institutional ap-

proach and defers until later a consideration of supply prices determined by cost-revenue calculations.

3. **No Reservation Price.**—The first type of institutional pricing for us to consider is the somewhat paradoxical one where the seller fixes no minimum supply price whatsoever. He may do this for two reasons. In the first case the seller may want money quickly. He does not have time to wait for the arrival of a buyer who will pay the customary asking price. Therefore, he offers his good for sale to the highest bidder or to the first buyer who appears on the scene. The auction is a familiar example of the first and a realty owner selling to a speculator illustrates the latter. In somewhat the same category is the instruction which a person owning a listed stock or traded commodity may give to his broker to "sell at the market." He will get the price that is quoted for the next transaction even though his contribution to the total sold is very small. In such sales through the exchanges the seller usually knows in advance approximately what he is likely to get, and might not tell his broker to "sell at the market" if he thought the market price was going to fall sharply for the next sale. Nevertheless, the fact remains that sellers who give such orders announce no reservation prices and therefore differ from those who do.

In the second case, we find sellers who believe it is futile to set a minimum supply price. Sellers may feel weak as compared to a buyer who has much larger financial resources. Many farmers, for instance, sell their produce or their milk to one buyer. They find it more difficult or less pleasant to deal with anyone else. So they accept the price he offers. Often the buyer argues that he is not a free agent and that his buying price is set by a selling price over which he has no control.

When buyers are few, they may organize and agree not to bid against one another in the purchase of certain goods. Unorganized sellers are impotent in the face of such agreements, and no one of them can force a higher bid by waiting and talking. It is futile for the individual seller to ask more. It is also foolish for him to accept less.

This feeling of urgency or impotence which leads sellers to name no reservation price should not be confused with the market situation described as the "balancing of aggregates" (Chapter 6, Section 6). In the latter case the naming of supply prices is very common, even though a few of those participating in a given sale on a stock or commodity exchange may be offering "at the market."

The equality of quantities bought and sold is established at a price set by some of the buyers and sellers, but not by all of them.[1]

4. Customary Price.—The easiest thing for sellers to do in many market situations is to offer to sell at the customary prices. This is true of old established producers and also for new firms entering the field. Some of the general motives involved were presented above in Section 3 of this chapter. Here we shall discuss the types of market situations which most frequently reveal the power of custom in price determination.

In the retail field customary prices tend to be retained for long periods on commodities of frequent purchase and low unit price, such as the 5-cent cup of coffee, cigar, candy bar, or soft drink. The fact that 5- and 10-cent prices can be paid with single coins seems to make these prices particularly stable. Twenty-five-cent commodities are purchased less frequently, and the price is less stable. Another obstacle to the change of customary prices is the frequent presence of many sellers. If one of them raises his price during a period of rising costs, some of the others are not likely to follow. This may be fatal for the first man to try it. In more technical language, the demand for the product of any one seller is very elastic when his price is raised above the customary one (cf. Chapter 4, Section 1).

An interesting variety of customary prices is to be found in the customary *percentages* involved in certain types of sales. These include the 10 per cent restaurant tip, the 2 per cent discount for prompt payment, the salesman's commission, etc. Piece-work wage rates are also a form of percentage pricing. Employers seem to resist union demands for piece rate increases more stubbornly than demands for time rate increases.

Manufacturers often maintain posted prices for a longer time than would seem warranted by changing demand or changing costs. They are much less likely than retailers to engage in "sales." Reasons for price rigidity at this level include the expense of preparing new price lists, the inconvenience and cost of re-educating salesmen, the fear of starting a "price war" among competitors, and the danger that downward adjustments when costs fall cannot be reversed readily when costs later go up again. Further explanations of these last two reasons will be given in the next chapter.

A final comment should be made to the effect that the longer a

[1] We ignore the exceptional case where all the sellers offer their goods "at the market." Even then potential sellers at higher prices probably exist and will sooner or later reveal their presence by offering goods at prices which will ensure sale.

price has remained in effect, the greater the difficulty in changing it. In this respect customary prices are like other customs. Ethical and moral arguments against price increases reinforce buyers' natural reluctance to pay more. Price decreases do not encounter such opposition, but neither do sellers feel price-cutting necessary to retain sales volume. Competition shifts to advertising, quality changes, and the like.

5. Prices Fixed by Law.—A few prices are fixed by law, either by statute or by decision of an administrative body given authority to act. Such are the government buying price for gold, certain public utility rates, and specific prices fixed by the Office of Price Administration during the second world war. In such cases the law-abiding seller has no decision to make about his supply price. He would be foolish to offer his goods for less. All he can do is to try to get the price increased. He may protest its "unfairness" or its harm to the public welfare, or he may proclaim the great good that would result for people in general if only his legitimate selling price were raised. During the depression of the 1930's the gold and silver producers talked like this. During the war nearly every seller did.

Other price-fixing takes the form of maximum or minimum prices. When the demand for goods rises more rapidly than the supply, governments may seek to hold down prices by maximum price legislation. The specific dollars-and-cents maxima described in the preceding paragraph are less common than "price freezing." This device establishes ceilings at the highest price asked by each seller for sales made during a certain base period (e.g., March, 1942). Individual adjustments are then permitted in cases of genuine hardship and public harm. Most sellers can sell their capacity output at the ceiling and see no gain from charging less. For them there is no supply price problem. Their task is to keep down costs by skillful buying, bargaining, quality debasement, etc.

During the depression which preceded the second world war, the United States tried to impose *minimum* prices. The NRA codes were soon invalidated by the Supreme Court, but for a time they reduced the amount of price-cutting. At best they did not have so great an effect upon supply prices as did wartime ceilings. Many sellers chose to offer goods at prices above the minima.

In recent years there has been a rise in the number of "guild prices." These are minimum prices imposed by quasi-governmental bodies controlled by particular trades, such as the barbers, the cosmetologists, the milk producers, etc. Other sellers are unable to get

favorable legislation of this kind and arrange, instead, monopolistic agreements of various types. Violators of legal guild price minima may be fined, or their licenses may be revoked. Violators of monopolistic agreements run the risk of retaliatory price-cutting. In both cases the area of institutional price determination is being widened. Songs in praise of free competition are becoming muted.

Public utility rates have long been supervised by governmental commissions. The companies themselves must initiate most new rates and rate changes upward. This requires price policy decisions on their part. Hundreds or thousands of rates must be determined for different types of sales. Since the allocation of overhead must be somewhat arbitrary, custom and the imitation of competing rates are perhaps more frequent supply price determinants than careful cost calculations. The companies must also meet the requirements of the law that no rates may be discriminatory and that the total net profit must not be higher than a "fair return on a fair value."

6. Resale Price Maintenance.—Another type of price setting which does not require independent policy decisions upon the part of sellers is called resale price maintenance. The middleman is required to sell at a price fixed by the manufacturer. This is usually a minimum price, but it may also be a maximum, as in the retailing of books. The compulsion has various institutional forms. It may be a contract which requires the seller to ask a certain price upon penalty of losing the right to buy the product in the future. The selling prices of wholesalers often are fixed by manufacturers who publish price lists showing fixed discounts for wholesale and retail buyers. Or there may be laws requiring all distributors to abide by the stipulated minimum to which some distributors are bound by contract. Other laws merely state that price-cutting may not go below a certain percentage of the list price. Nearly all states have some type of "fair trade law." Legal protection of this type has been sought by producers of drugs, toiletries, books, liquors, films, etc.

7. Imitative Price.—Sometimes sellers copy the supply prices of other sellers. A small seller may not want to put his price higher than that of a larger firm for fear of losing business to it. At the same time he does not dare to put the price lower for fear of starting a "price war." When the large firm changes its list price, the small firm plays "follow the leader." A new firm without cost experience may imitate predecessors upon the assumption that their prices will be high enough to cover its expenses and yield a reasonable profit. Too often the new firm fails to realize that profit also depends upon

achieving a certain minimum volume of sales. If there is some product differentiation, by brands or quality, as in most packaged foods, the small seller or the new one may be slightly independent. He may advertise an alleged qualitative superiority and ask a higher price. Or he may claim equal quality and offer slight price cuts as an inducement.[2] But the fundamental approach is imitative; it is not a mathematical calculation of the maximum profit price.

In the retail merchandising of standardized or nationally branded products, as in grocery and drug stores, one firm may adopt the policy of meeting every price cut made by competitors. "We refuse to be undersold" is a claim frequently made and sometimes followed in actual prices. Such stores often hire professional shoppers to purchase goods at the counters of their competitors in order to find out just what prices are being charged. With this information, or when a customer makes a protest, the "cut-rate" store reduces prices where necessary. Occasionally they also find their own prices lower than others and proceed either to raise them or to publicize them by advertising. The knowledge that weaker Store B will follow the price cuts of stronger Store A has sometimes led the management of Store A to reduce prices below total cost on certain volume items, hoping to exhaust the resources of Store B and thus force it out of business. This accomplished, Store A may then raise prices, even to a level higher than before the reduction, thus recouping its losses.

8. Markup Methods: (a) Markup for All Overhead and Profit Added to Direct Cost.—A method of determining supply price which combines imitation and calculation is the markup. A middleman may figure his unit cost of obtaining a good and then add a traditional percentage markup to allow for overhead and profit. Thus, if a grocer purchases a certain brand of spaghetti at 10 cents per pound package, he may add a customary 20 per cent and ask customers to pay 12 cents. Grocers use many different markup percentages, depending roughly upon the rate of turnover of each class of goods. Bread and milk have a rapid turnover and a small markup as compared with slow-moving items like spices, bleaches, and tea. There may be several dozen such groups with inventory turnover running from once a day to once a month or longer. This explains why the markup on article A which has an average turnover of once a week should be higher than that on article B which turns over twice a week. But it does not explain why the Reliable

[2] Much more will be said in later chapters about problems of price-cutting in various market situations. Cf. Chapter 9, especially Section 5.

Grocery Company chooses a 15 per cent markup for A and 10 per cent for B instead of some other figures such as 16 per cent and 9 per cent.

When dozens of articles are sold by a single seller, he can never be sure that he has precisely the most profitable markup on each item. Hence he is usually guided by custom and by the price decisions of his competitors. His minimum essential task is to cover his total overhead. His total revenue must exceed his total outlays for rent, wages, taxes, etc., plus depreciation allowances, if he is to remain in business. To make a profit he must make something more than this. Unfortunately for him, however, there is no single markup rule which will guarantee a profit. Where the number of articles being sold is large, the markup combinations which may yield a profit are legion. That is clearly revealed by the random variety of items chosen for price reductions by retail sellers when they wish to attract customers by advertised "sales." It is shown also by the low markup policy pursued by most chain stores at the same time that independent neighborhood stores have high markups.

9. Markup Methods: (b) Markup for Administrative Overhead and Profit (after Cost Accounting).—A second markup method of arriving at supply price is used by many manufacturers. Most of the overhead costs are allocated by accounting techniques and then a flat percentage markup is added to cover any remaining costs and to provide a profit. The chief accomplishments of cost accounting lie in distributing among their various uses the proper fractional parts of such joint costs as power, supervisory or incidental labor (foremen, clerks, janitors), and certain materials used in more than one article or process. The markup percentage becomes smaller as the cost base becomes more inclusive. But cost accounting cannot allocate expenses like research, property taxes, and administrative salaries without becoming highly arbitrary. Therefore an institutional element remains: the rule-of-thumb for the allocation of these costs or for the profit percentage added.

There is another reason for not putting too much reliance upon the cost accounting approach to supply price. Its negative value is greater than its positive. A newly introduced cost accounting system may show that a certain article has been sold at less than the estimated cost of production. But no amount of accounting can show that the volume of production used in cost estimates can be sold at a price which covers cost and the profit markup. This can only be determined by trial and error or, in a few rare instances, by market surveys which

discover potential demand. Similarly, no markup percentage is *the* correct one, guaranteed to yield the maximum profit, or any profit, for that matter. Markup rules are as numerous and as diverse as formulas for beating the stock market or choosing the winning horse at the Derby. Sellers in the same line of business will follow different rules, and during prosperity each will defend his rule as the best one.[8] When a seller fails, his chief trouble may lie elsewhere than in his markup percentage.

Individual sellers themselves often use different markups at different times or for different customers. Retail stores frequently stage "sales" and offer merchandise for less than former prices, i.e., at smaller markups. The reductions are not often connected with the need to cover overhead costs. More common reasons are the desire (1) to sell remainders of old stock quickly in order to make room for new merchandise of particular seasonal or style appeal; (2) to attract new customers who may buy other goods at nonsale prices; or (3) to raise cash to pay creditors. Many retailers also depart from their "one-price" rule when selling to employees. At the wholesale and manufacturing levels it is very common for sellers to charge different prices (use different markups) for different customers by a series of special discounts from list prices.

10. Markup Methods: (c) When Producing to Order.—A variant of the markup method of setting supply prices is found in those cases where the seller produces to order. In some cases the buyer offers to pay whatever it costs. The seller may then figure "time and materials." The hourly rate charged for the time required includes an allowance for overhead. In other cases the buyer says he wants something made to order and asks the seller to quote a price in advance. The producer must then decide the size of his markup. If he expects the order to be the only one from that buyer, he may charge more than if he is trying to build goodwill for future sales. Other influences are the age of the firm, its competitive position, the number of orders in proportion to the total business, etc. Production to order, one must remember, is not confined to custom tailoring. It is quite common in the purchase of certain types of semimanufactured raw materials, specialized equipment for factories, and even some retail merchandise. Some machinery manufacturers concentrate on producing to specification.

[8] They may also differ in cost accounting procedures, as in the treatment of inventories, depreciation, etc.

A variant of this situation occurs when a potential buyer states his needs to several possible suppliers and asks them to submit sealed bids. Under these conditions the potential seller arrives at his bid price in much the same way as before, i.e., a calculation of cost plus a profit markup. But other considerations may intrude. A producer who is currently running near capacity may be less anxious to get the job than one who is much in need of additional business. Profit margins may be adjusted accordingly in the bids submitted. There will also be differences in the estimated cost of production. Some firms are more advantageously equipped, stocked, or located than are other firms. The low bidder does not always set the price, for the buyer may find some method of playing favorites or may object to the bidder's materials, financial weakness, or lack of experience. Governments and some large private buyers use the sealed bid technique in purchasing both standardized commodities like fuel oil or copper pipe and nonstandardized things like factory buildings, post offices, and ships.

11. Summary of the Institutional Determinants of Supply Price.—Sellers determine their supply prices in a variety of ways depending upon the situation in which they find themselves. Often they exercise little conscious choice, but follow the dictates of custom, law, example, or the peculiar circumstances of the case. These institutional forces are widespread and potent. Even when some costs are calculated, the markup percentages often remain institutional or impulsive. Most of the prices we encounter in daily life spring from these institutional backgrounds and therefore we have studied them first. Now we may proceed to examine how prices may be set by using profit-maximizing formulas based upon hypothecated cost and revenue schedules.

Chapter 8

SUPPLY PRICE: DERIVED FROM MAXIMUM PROFIT CALCULATIONS

1. Calculating the Maximum Profit Price Under Assumed Cost and Demand Conditions.—Another group of supply prices is that in which the maximum profit price is calculated from assumed cost and demand conditions. This process differs from the methods described in the last part of the preceding chapter. It does not use any customary, legal, or imitative markups. It assumes that cost varies with output and that output varies with the rate of sales. Therefore, the price setter should estimate two functions and consider them both at the same time. The first is the cost function, or the costs of production at different rates of output (and sales). The second is the revenue or demand function. This was explained in Chapter 4 as the quantities that would be demanded of the seller at the different prices which he might set.

Many different assumptions are possible. Production cost may be viewed in various ways. Demand situations may be of one type or another. The following sections consider only a few of the possible combinations. Most of the cases show what *would be* the best price *under the given assumptions.* They do not say that sellers in real life actually do reach their supply prices in the manner shown. They say, rather, that if a seller were to seek his maximum profit price and output in this way or that way, he would obtain them by such and such a method. The major questions to be asked and answered are the following:

1. How to maximize gross revenue when cost of production is ignored?
2. What are the different types of cost schedules for the individual firm?
   (a) Total fixed cost, total variable cost, total total cost
   (b) Average fixed cost, average variable cost, average total cost
   (c) Marginal cost
3. How may these cost schedules be derived from input-output data?
   (a) In agriculture; the principle of diminishing returns

(b) In manufacturing; the possibility of constant marginal cost

4. Why are the maximum profit price and output revealed by the point of equality between marginal revenue and marginal cost?
   (a) When the selling price is beyond the seller's control
   (b) When the selling price may be set by the seller
5. What is "capacity" output and how is it related to the maximum profit output?
6. What is the effect of increases in equipment or other "fixed" expenses as output expands?
7. To what extent are the actual transactions prices of modern capitalistic society determined by these principles?

2. Seeking to Maximize Gross Revenue When Cost of Production Is Ignored.—The simplest type of supply price calculation in the present group is that which ignores cost of production. This occurs when goods have already been produced and the stock on hand is fixed. The seller is assumed to be able to determine the selling price by setting his supply price. His task is to determine the supply price which will bring him the maximum gross revenue within a given time. If that price is so high that some goods remain unsold, these can be destroyed, or given away. They may also be offered for sale in another time period at the same or a different price.

The solution of the problem is simple, once the demand for the good has been estimated. Set up a three-column demand schedule (cf. Chapter 4, Section 2) whose largest quantity is the maximum amount on hand for sale. Find the maximum total revenue amount in that schedule and charge the price indicated. In the following hypothetical example and diagram, the seller would make $81 by charging only 90 cents per unit. This will leave 10 units unsold. But the seller is better off than if he had tried to sell his total stock of 100 and had reduced his price to 80 cents in order to do so.

DEMAND SCHEDULE FOR 100 GADGETS OFFERED ON ONE DAY BY SELLER SMITH

Price	Quantity	Total Revenue
$1.20	60	$72.00
1.10	70	77.00
1.00	80	80.00
(best price) .90	90	81.00 (maximum revenue)
.80	100 (total stock)	80.00

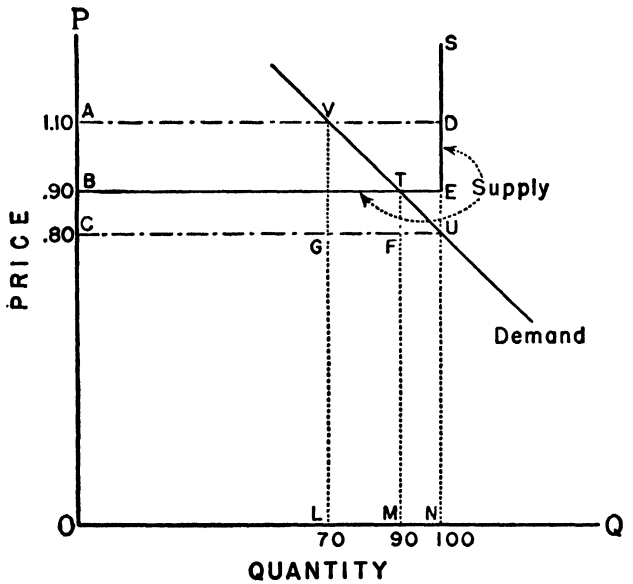
SUPPLY SCHEDULE OF SMITH AFTER PRICING TO MAXIMIZE GROSS REVENUE

Price	Quantity
$0.90	60
.90	70
.90	80
.90	90 (best quantity)
.90	100

The diagrammatics of the situation is shown in Figure 23. Note that the supply curve is a horizontal straight line of limited length turning vertically upward at 100. Its position is unknown until the seller determines his supply price. Under the assumed conditions it

FIGURE 23

PRICING TO MAXIMIZE GROSS REVENUE WHEN STOCK IS FIXED AND COST IS IGNORED



will extend to the right from 90 cents and will intersect the demand curve at *T*. If Smith had charged a lower price such as 80 cents per gadget, his demand and supply curves would have intersected at *U*. He would have sold all his 100 gadgets, but his gross revenue would

(b) In manufacturing; the possibility of constant marginal cost

4. Why are the maximum profit price and output revealed by the point of equality between marginal revenue and marginal cost?
   (a) When the selling price is beyond the seller's control
   (b) When the selling price may be set by the seller
5. What is "capacity" output and how is it related to the maximum profit output?
6. What is the effect of increases in equipment or other "fixed" expenses as output expands?
7. To what extent are the actual transactions prices of modern capitalistic society determined by these principles?

2. Seeking to Maximize Gross Revenue When Cost of Production Is Ignored.—The simplest type of supply price calculation in the present group is that which ignores cost of production. This occurs when goods have already been produced and the stock on hand is fixed. The seller is assumed to be able to determine the selling price by setting his supply price. His task is to determine the supply price which will bring him the maximum gross revenue within a given time. If that price is so high that some goods remain unsold, these can be destroyed, or given away. They may also be offered for sale in another time period at the same or a different price.

The solution of the problem is simple, once the demand for the good has been estimated. Set up a three-column demand schedule (cf. Chapter 4, Section 2) whose largest quantity is the maximum amount on hand for sale. Find the maximum total revenue amount in that schedule and charge the price indicated. In the following hypothetical example and diagram, the seller would make $81 by charging only 90 cents per unit. This will leave 10 units unsold. But the seller is better off than if he had tried to sell his total stock of 100 and had reduced his price to 80 cents in order to do so.

DEMAND SCHEDULE FOR 100 GADGETS OFFERED ON ONE DAY BY SELLER SMITH

Price	Quantity	Total Revenue
$1.20	60	$72.00
1.10	70	77.00
1.00	80	80.00
(best price) .90	90	81.00 (maximum revenue)
.80	100 (total stock)	80.00

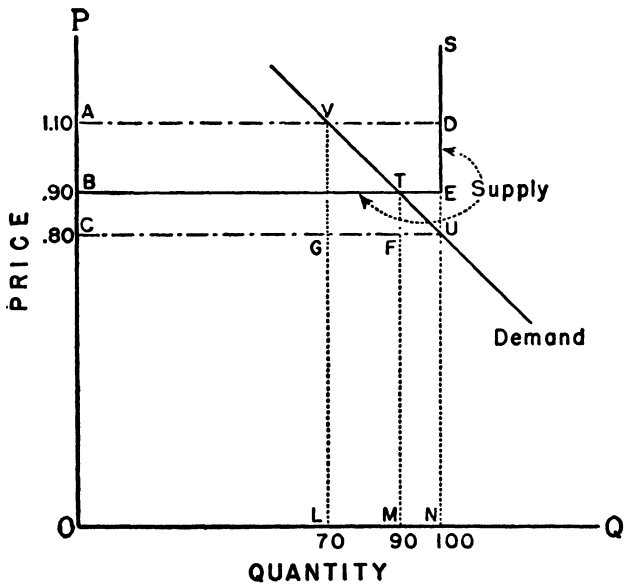
SUPPLY SCHEDULE OF SMITH AFTER PRICING TO MAXIMIZE GROSS REVENUE

Price	Quantity
$0.90	60
.90	70
.90	80
.90	90 (best quantity)
.90	100

The diagrammatics of the situation is shown in Figure 23. Note that the supply curve is a horizontal straight line of limited length turning vertically upward at 100. Its position is unknown until the seller determines his supply price. Under the assumed conditions it

FIGURE 23

PRICING TO MAXIMIZE GROSS REVENUE WHEN STOCK IS FIXED AND COST IS IGNORED



will extend to the right from 90 cents and will intersect the demand curve at *T*. If Smith had charged a lower price such as 80 cents per gadget, his demand and supply curves would have intersected at *U*. He would have sold all his 100 gadgets, but his gross revenue would

have been less. At a higher price, $1.10, the quantity sold would have been less than 90 as indicated by the dotted line through *V*. This diagram also illustrates two other concepts described in earlier chapters. The higher the supply curve, the lower the supply in the schedule sense (cf. Chapter 6, Section 4). Raising the supply price reduces the supply, even though Smith insists he still has 100 units he would like to sell. And the demand schedule is obviously elastic above *T* and inelastic below (cf. Chapter 4, Section 4). If price units were in ones instead of tens, the transition might be shown as occurring through a phase with elasticity of unity just above or below 90 cents.

3. Cost Schedules for the Individual Firm: (a) Total Cost.— All the remaining cases of calculated maximum profit prices require the use of cost schedules. These must now be explained. Cost schedules are constructed much like the demand and supply schedules described earlier. For each quantity there is a cost figure instead of a price. There are seven possible costs which may be used. The present section deals with three types of *total* cost: fixed, variable, and total. The next will describe three similar varieties of *average* cost. Section 5 will explain *marginal* cost.

To understand these different types of cost one must first distinguish between fixed and variable costs. These are technical terms whose definitions in economic analysis must be accepted even if they are different from those of other sciences, such as accounting. *Fixed costs* are those whose total does not change within a given range of change in output. *Total variable costs* do change with output. Administrative salaries, bond interest, real property taxes, and fire insurance usually fall in the first class of fixed costs. Raw materials and direct labor are usually good illustrations of variable costs.

Some costs fall in one category or in the other according to the degree of change in output. When the capacity of a given machine is reached, another must be added. Depreciation on the first machine was a fixed cost for a certain range of output change. Beyond that range, further expansion of output requires the addition of a second machine. This outlay may be called a variable cost. Once the second machine is acquired, however, its depreciation may be treated as a fixed amount for the range of expansion of output it makes possible. The distinction between fixed and variable costs may also depend upon the decision of the person constructing the cost schedules. If he wishes to designate as overhead a certain expense, such as that of time clerks, and treat it as a fixed cost, he may do so. But if his cost accounting system allocates this expense to units of goods pro-

duced, it may be considered a variable cost. The person who is making the profit-maximizing calculations is the one who must decide where the various expenses should be put.

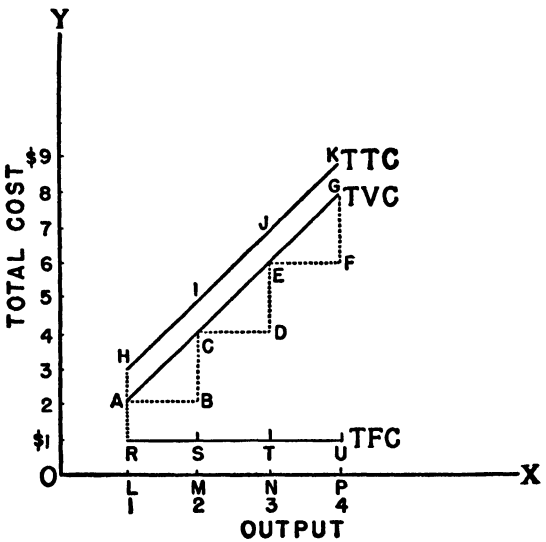
The following table and figure show a hypothetical case. Total fixed costs are $1 and do not change as output rises from 1 to 4 units. Total variable costs rise from $2 to $8. This is shown as a dotted "staircase" which has been smoothed into a solid straight line for simplicity. The top line represents the sum of the total fixed and variable costs at each output. *Total total cost* sounds redundant, but the double adjective is useful to help distinguish it from total average cost, which will be explained in the next section.

HYPOTHETICAL *Total* COST SCHEDULES AND CURVES FOR A FIRM

Total Output	Total Fixed Cost	Total Variable Cost	Total Total Cost
1	$1.00	$2.00	$3.00
2	1.00	4.00	5.00
3	1.00	6.00	7.00
4	1.00	8.00	9.00

FIGURE 24

THREE TOTAL COST CURVES



A note may be added for those who find mathematical formulation of such material interesting or useful. The equations of the curves are as follows:

$y_1 = \$1$ (total fixed cost is constant at \$1)
$y_2 = x(\$2)$ (total variable cost is a multiple of \$2)
$y_3 = \$1 + x(\$2)$ (total total cost is the sum of $TFC + TVC$)

Diagrammatically, TFC is a horizontal straight line. TVC, when smoothed, is an upward sloping straight line. TTC is an upward sloping straight line parallel to TVC and separated from it by the distance (\$1) between TFC and the OX axis. For instance:

$$LH = LR + LA$$

$$MI = MS + MC$$

$$\text{also } LR = AH = MS = CI = \ldots$$

4. Cost Schedules for the Individual Firm: (b) Average Cost. —The average cost figures are generally more useful than the total cost figures. They are derived from the latter by simple division, using total output as the divisor. This gives average fixed cost, average variable cost, and average total cost as shown in the following table based upon the schedules given in Section 3:

HYPOTHETICAL *Average* COST SCHEDULES AND CURVES FOR A FIRM

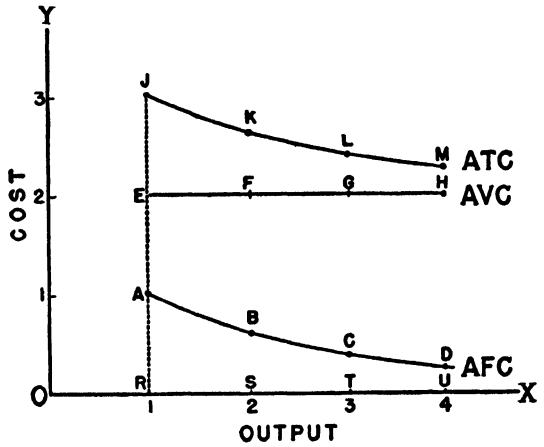
Total Output	Average Fixed Cost	Average Variable Cost	Average Total Cost
1	\$1.00	\$2.00	\$3.00
2	.50	2.00	2.50
3	.33	2.00	2.33
4	.25	2.00	2.25

Three characteristics of the new schedule should be noted. Average fixed cost declines, sharply at first and then more gradually. Average variable cost remains constant, since the total variable cost function was constructed in the first case as a simple multiple of \$2. (A later case will give an illustration of both TVC and AVC as curves, instead of straight lines, the TVC curve bending upward and the AVC curve bending downward.) Average total cost declines with increasing output, following the pattern of average fixed cost. Average total cost also may be obtained by adding AFC and AVC.

In Figure 25 showing the three average cost curves, $RJ = RA + RE$, resembling the distances added to get totals in Figure 24. In similar fashion, $AJ = RE = BK = SF = \ldots$. The diagrammatic

FIGURE 25

THREE AVERAGE COST CURVES



contrasts between the two figures are best seen by putting them in parallel columns as follows:

TOTAL COST CURVES		AVERAGE COST CURVES[1]
TFC	—Horizontal—	AVC
TTC and TVC	—Parallel—	ATC and AFC
$TTC = TFC + TVC$	—Totals—	$ATC = AFC + AVC$

As a mathematical note the equations of the average cost curves may be given as:

$y_1 = \frac{\$1}{x}$ (average fixed cost is \$1 divided by the output)

$y_2 = \frac{x(\$2)}{x} = \$2$ (average variable cost is constant at \$2)

$y_3 = \frac{\$1 + x(\$2)}{x} = \frac{\$1}{x} + \$2$ (average total cost is the sum of $AFC + AVC$)

5. Cost Schedules for the Individual Firm: (c) Marginal Cost. —The seventh cost schedule and curve for the individual firm are of the marginal variety. Marginal cost is an addition to total cost that

[1] In other cases AVC may be curved. When this occurs, ATC and AFC are not parallel, nor are TTC and TVC, though nearly so.

may be described in several ways. These should be compared with the similar definitions of marginal revenue given in Chapter 4, Section 4. In every case, marginal cost must be thought of as occurring at or just before a certain output, x.

When output increases one unit at a time, then marginal cost is the amount added to the total total cost at $x - 1$ units of output to make the new total total cost at x units of output. In other terms, it is the difference between the total total costs at the two rates of output and may be expressed by the formula:

$$MC_x = TTC_x - TTC_{x-1}$$

In our hypothetical cost schedule one case might be that when output, x, is 4 units, then $MC_4 = TTC_4 - TTC_3 = \$9 - \$7 = \$2$.

When changes in the scale of production by a firm are estimated by its manager to occur in larger jumps than just one unit, marginal cost becomes *average* additional cost. When the divisor was 1, as in the first case, it could be ignored, since the average increment and the increment itself were the same. Not so when increments occur of 100 units, 1,000 units, or larger amounts. Let the increment be called m, then the new formula becomes:

$$MC_x = \frac{TTC_x - TTC_{x-m}}{m}$$

In a numerical illustration, a firm producing 9,000 units per month at a total cost of \$90,000 might step up its output to 10,000 units and find that its new total cost was \$101,000. The *average* additional cost is $\frac{\$101{,}000 - \$90{,}000}{1{,}000}$, or \$11 per unit *added*. This marginal cost must be contrasted with the average *total* cost at 9,000 units of output which was $\frac{\$90{,}000}{9{,}000} = \$10$, and the ATC at 10,000 units, which was \$10.10.

For diagrammatic purposes, a smoothed curve is more convenient than the more realistic staircase zigzag. Therefore, economists often use the convenient fiction of infinitesimal changes in output. Under this assumption, m is made to diminish until it approaches zero as a limit. The division now gives the *rate of change* in total cost at a certain output, x, instead of the average *amount* of change *between* two outputs, $x + m$ and x. It is this rate concept of marginal cost at a point (an output) which economists usually have in mind when they use the term. Businessmen, on the other hand, if they use the mar-

ginal cost concept at all, would be more apt to think of the average additional cost of an expanded output as described above.

6. Marginal Cost Usually Differs from Average Variable Cost. —Marginal cost may be the same as average variable cost, or it may differ. The simple illustration used in Sections 4 and 5 made them both the same. That is, *MC* was always $2 and *AVC* was likewise always $2. This resulted from the assumptions of the schedule, which can be described either as (1) a straight-line (linear) total cost function based on variable costs of the simple multiple type, (2) an unchanging average variable cost, or (3) a constant marginal cost.[2]

In most cases found in actual business practice, marginal cost and average variable cost differ at some if not at most output levels. There are two major reasons. First, there are changes in the efficiency with which the variable factors are combined with the fixed factors. Illustrations include improvements such as increasing physical returns, better coordination, reduced waste, and greater specialization. As the scale of output gets very large, decreases in efficiency may occur because of diminishing physical returns, duplication, and "red tape." A second major reason for changes in average variable cost is not related to changes in efficiency in the use of the factors but rather to their unit cost. It may result from the buying economies obtained by purchasing in larger lots. Or the increased volume of buying may raise unit costs through the need to pay freight from longer distances, pay overtime wages, etc.

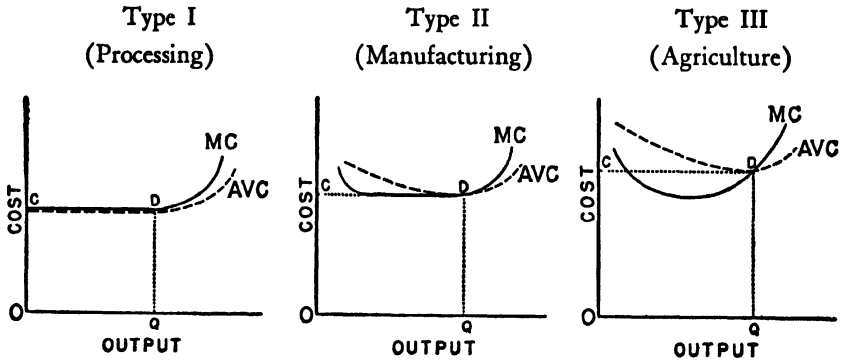
There are three major types of relationship between marginal cost and average variable cost. These are shown in Figure 26. In the first, $MC = AVC$ throughout most of the output range. This was the assumption of our first hypothetical examples and might be found in many simple processing operations. Eventually, however, *MC* rises and *AVC* follows, but less rapidly. In the second case, *MC* falls at first, remains constant for a large range of outputs, and then rises. *AVC* falls slowly until it intersects *MC* just after *MC* starts to rise. Many factories show this pattern. The third illustration is taken from agriculture. It shows *MC* falling and then rising in a U-shaped curve, *AVC* falls less rapidly and again intersects *MC* after

[2] Because of the arithmetic relation between five of the cost schedules of a firm, *TTC, TVC, ATC, AVC,* and *MC,* a schedule of any one of them with output may be used to obtain any one of the others. Only the fixed costs, *TFC* and *AFC,* remain independent of the other five. It is interesting to note that $TTC_n - TTC_{n-1} = TVC_n - TVC_{n-1}$, since the *TFC* component of *TTC* does not change. Therefore, *MC* is *either* the addition to total total cost or to total variable cost, but the former idea is usually the more useful of the two.

MC starts to rise. In both the second and third cases, the curves cross at the *lowest point* on *AVC*. In the first case the curves separate at that lowest point which, of course, is constant until the upturn begins.

FIGURE 26

THREE TYPES OF MARGINAL AND AVERAGE COST CURVES



7. Increasing and Decreasing Returns in Agriculture.—The traditional U-shaped cost curves of many authors were derived chiefly from studies of agricultural experiment stations which dealt with the effect upon output when different quantities of a single variable were combined with a composite fixed factor. For instance, increasing quantities of a certain kind of fertilizer have been added to a series of test plots where seed, soil, and cultivation have been held constant. On the various plots, the output per pound of fertilizer added has been observed to rise with the first few additions, reach a peak, and then decline as in the hypothetical table on page 103.

This table reveals *three* points of diminishing returns. Total product reaches a peak at 34 and begins to decline thereafter. This is the point of diminishing *total* returns. Note that at this point the additional product (column D) is zero. Second, there is a point of diminishing *additional* returns when three pounds of fertilizer are used and total product is 17. Beyond this point additional pounds of fertilizer increase the total product each time by less than 7 units of product. The 7 unit increase is the maximum. The third point of diminishing returns is that of the *average* returns column, E. This reaches its maximum at 5.75 when there are 4 units of fertilizer input and 23 units of product output.[8]

[8] Some synonyms should be understood: product = output = returns; and additional = marginal.

INCREASING AND DIMINISHING RETURNS IN AGRICULTURE

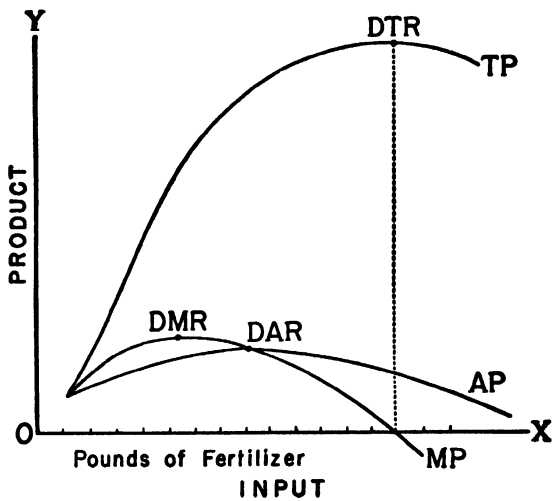
(Land, seed, cultivation, etc. constant; fertilizer variable)

A Composite Units of the Fixed Factors	B Pounds of Fertilizer (n) (Given)	C Total Product in Bushels (Given)	D Additional Product per Additional Pound of Fertilizer (Marginal Product) $(C_n - C_{n-1}) \div (B_n - B_{n-1})$	E Average Product per Pound of Fertilizer (C/B)
1 unit	1 pound	4 bushels	4 bushels	4.00 bushels
1	2	10	6	5.00
1	3	17	7 (*DMR*)	5.67
1	4	23	6	5.75 (*DAR*)
1	5	28	5	5.60
1	6	32	4	5.33
1	7	34	2	4.86
1	8	34 (*DTR*)	0	4.25
1	9	31	− 3	3.44

These points of diminishing total, marginal, and average returns may also be shown by diagrams. Figure 27 presents three typical

FIGURE 27

POINTS OF DIMINISHING PHYSICAL RETURNS



curves. They are drawn to show relationships and do not represent the figures of the preceding table. The point of *DTR* must occur directly above the point where the *MP* curve is zero, i.e., where it cuts the *OX* axis. The point of *DAR* must occur where *AP* is cut by *MP*. Additions which raise the average must exceed that average (*MP* points to the left of *DAR*). The point of *DMR* is more difficult to observe visually, but the calculus may be used to prove that it occurs where *TP* stops being convex to the *OX* axis and begins to become concave.

8. Transposing Diminishing Physical Returns Into Increasing Physical Cost.—If these product curves are turned upside down, they become the familiar U-shaped cost curves. More accurately, the cost figures are the reciprocals of the product figures. For instance, column E gave the physical product per unit of fertilizer used. These average products were obtained from successive quotients of C/B. The average cost figures of column G are the reciprocals of column E, or the quotients of B/C. In briefest terms, column E shows output per unit of input while column G shows input per unit of output. Similarly, columns F and H may be considered the reciprocals of columns D and C respectively. They also may be calculated independently by the formulas given. One must remember that all of these figures represent *physical* cost. They are *fractional* pounds of fertilizer used. Not until the next section will we be dealing with *money* cost.

DECREASING AND INCREASING *Physical* COST IN AGRICULTURE

A Fixed Factor	B Variable Factor	C Product	F Marginal Cost $(1/D)$ or $(B_n - B_{n-1}) \div (C_n - C_{n-1})$	G Average Variable Cost $(1/E)$ or $(B_n \div C_n)$	H Average Fixed Cost $(1/C)$ or $(A_n \div C_n)$
1 unit	1 pound	4 bushels	0.250 pounds	0.250 pounds	0.250 units
1	2	10	.167	.200	.100
1	3	17 *	.143 *	.176	.059
1	4	23 †	.167	.174 †	.043
1	5	28	.200	.178	.036
1	6	32	.250	.187	.031
1	7	34	.500	.206	.029
1	8	34‡	Infinity	.235	.029 ‡
1	9	31	− 0.333	.290	.032

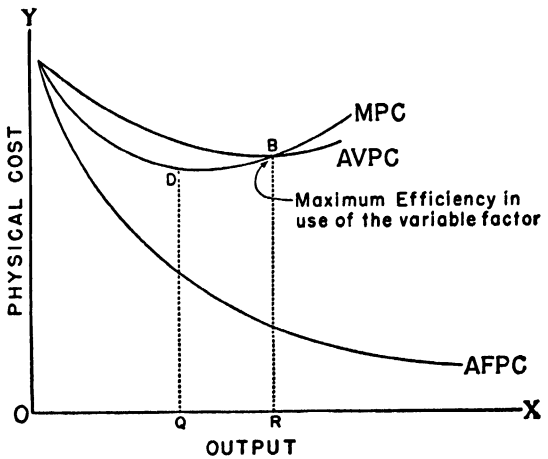
* Point of lowest marginal physical cost.
† Point of lowest average variable (physical) cost.
‡ Point of lowest average fixed (physical) cost.

The three physical cost functions may now be diagrammed. Representative curves are shown in Figure 28. They resemble the output curves of Figure 27, but are upside down. Instead of points of diminishing returns, there are now three points of increasing physical cost. Only one of these is very significant, the point of minimum average variable physical cost. This occurs where the *AVPC* curve is intersected by the *MPC* curve. The reasoning is the same as that given in the preceding section for the intersection of *AR* and *MP*. This point of lowest average variable physical cost is sometimes called the *point of maximum efficiency* in the use of the given variable. It represents the scale of production at which the least input is required per unit of output, so far as that variable is concerned. But it should not be confused with the point of lowest total *money* cost, nor with the output which maximizes profit, both of which will be described later in the chapter.

One other change should be noted. The *OX* axis in Figure 28 is changed from inputs to outputs. This changes the shape of the

FIGURE 28

THREE PHYSICAL COST CURVES



curves somewhat, but the fundamental relationships remain the same. There is merely a shift from plotting pairs of figures which include column B to pairs which include column C in the tables given above. The output approach is essential to the total money cost analysis soon to be given.

9. From Physical Cost to Money Cost.—It is a simple matter to translate physical cost into money cost by multiplying the indicated

pounds of fertilizer by the cost price per pound, say 10 cents. The composite fixed factor may be treated similarly, when the appropriate prices are known. This permits adding the *money* cost of the variable factor to the money cost of the fixed factor to get a total money cost of one unit of the product at each output. (No total could have been obtained by trying to add pounds and acres.)

Another minimum point may now be shown. It is the point of minimum average total money cost, sometimes called the lowest unit cost. A few writers call the output which reveals this minimum cost the "capacity" output. The volume of production may be expanded beyond this point, but unit costs rise. This use of the word *capacity* should be contrasted with the popular connotation of the maximum amount that can be produced in a given time period regardless of cost. Furthermore, technical "capacity" exists at a different output for each factor considered as a variable. Thus, there is one capacity output for fertilizer inputs, another for seed, a third for cultivation labor, etc.

The following table shows the principle in relation to the hypothetical figures of the preceding illustration with the additional assumption of a 10 cent per pound cost of fertilizer and a 50 cent cost for the composite unit of the fixed factors.

MARGINAL MONEY COST AND AVERAGE MONEY COST IN AGRICULTURE

B Variable Factor (Inputs)	C Product (Outputs)	I Marginal Cost in Dollars ($F \times \$0.10$)	J Average Variable Cost in Dollars ($G \times \$0.10$)	K Average Fixed Cost in Dollars ($H \times \$0.50$)	L Average Total Cost in Dollars ($J + K$)
1	4	$0.0250	$0.0250	$0.1250	$0.1500
2	10	.0167	.0200	.0500	.0700
3	17	.0143	.0176	.0295	.0471
4	23	.0167	.0174	.0215	.0389
5	28	.0200	.0178	.0180	.0358
6	32 *	.0250	.0187	.0155	.0342 *
7	34	.0500	.0206	.0145	.0351

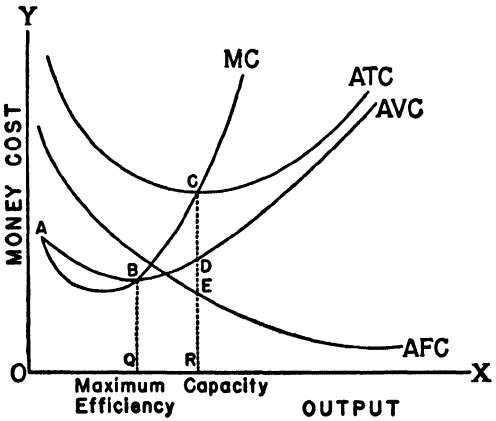
* Capacity, or point of lowest average total money cost.

Figure 29 shows representative curves for a farm product table like that above. The vertical axis is now money cost. This permits adding the OY distances of the average fixed cost curve and the average variable cost curve to get a higher curve known as the average total cost ($RE + RD = RC$). This latter curve approaches the

AVC curve as the *AFC* curve approaches zero. The minimum point, *C*, of *ATC* occurs at a larger output than the minimum point, *B*, of the *AVC* curve. These two minima are shown on the *OX* axis as

FIGURE 29

FOUR MONEY COST CURVES



R and *Q* respectively. Their designation as "capacity output" and "maximum efficiency output" must be considered technical definitions as explained above.

The marginal cost curve intersects *both* the *AVC* and the *ATC* curves at their lowest points. This is important in the analysis to follow, which explains the use of the *MC* and *ATC* curves in calculating the maximum profit price. The shape of these curves is determined by the conditions of physical productivity peculiar to agriculture. In manufacturing and merchandising the curves may be much flatter, but their typical relationship is most clearly seen in this example.

10. Use of Cost and Revenue Curves in Setting Price.—In the preceding six sections we have completed a survey of the cost curves of the individual firm. Now we can proceed to the solution of our basic problem. We want to know how a business manager might determine the price which would maximize his profit if he knew his cost and revenue schedules. The answer may be stated formally as follows. The maximum profit price is that at which the additional cost of the last unit sold is just equal to the additional revenue derived from that sale. More briefly, it is the price of the output for which marginal cost equals marginal revenue.

The reasoning is simple. At any lower price the marginal revenue would be less than the marginal cost. A loss on that marginal sale would occur. Therefore, the volume of sales-output should be reduced and this means raising the price. At any higher price a profit would be made on the marginal unit. This would suggest an expansion of sales-output, but that would require a reduced price. The maximum profit price is the only one at which no change in sales-output would benefit the producer. That is why it is sometimes called the *equilibrium price* for the individual firm. Another name might be *equilibrium quantity,* but this phrase is less common. It has the advantage of pointing out the necessary identity between output and sales in this approach. Cost relates to a quantity output which is equal to the quantity sold at the price which maximizes profit.

The application of the foregoing rule for determining the maximum profit price for an individual firm may be either numerical or graphic. If a manager were to make such calculations for his own firm, he would use figures. He would begin by drawing up four schedules: (1) average revenue and quantity, (2) marginal revenue and quantity, (3) average total cost and quantity, and (4) marginal cost and quantity. For simplification, quantity might be put in the first column and then four others for *AR*, *MR*, *ATC*, and *MC* ranged along side. Next he would look at *MR* and *MC* to find the point where they are equal. This will occur at a certain quantity. If the *ATC* is subtracted from the *AR* for that quantity, the profit per unit will be obtained. The final step is to multiply this unit profit by the quantity to obtain the total net profit. This is the maximum profit obtainable under the assumed circumstances. The solution may be proved by adding two more columns to the table. The sixth would be profit per unit at each output level. The seventh would be total profit, obtained by multiplying unit profit times quantity. The largest figure in this seventh column would be seen to occur at the output which also reveals *MC* to be equal to *MR*.

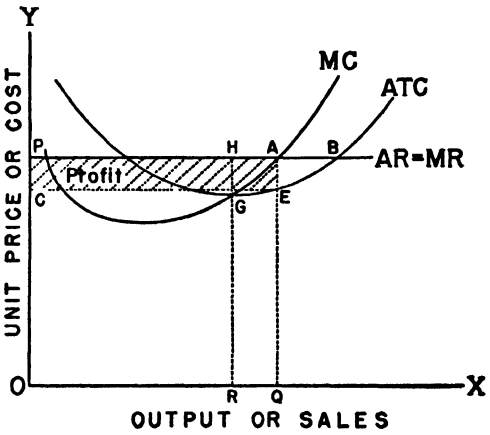
Curves may be used to arrive at a similar result. Businessmen would have to use figures since they want precise numerical answers. Economists merely want to demonstrate the principle involved and hence resort to their graphic shorthand. The symbols of this shorthand have been presented above in the form of typical cost and revenue curves. With the use of these symbols it is now possible to present quickly some of the many different situations which occur in real life. Although the cases can be defended as real, one must not assume that most managers of business units use these techniques in

determining their actual prices or outputs. Further comments on this qualification will be made after the typical cases have been examined.

11. Graphic Demonstrations of Profit-Maximizing Situations When Price Is Constant.—The first case is taken from agriculture. Farmer Smith is assumed to be one of thousands of farmers selling the same product. The price which he expects to get for his product will not be affected by any variations in his output or sales. He may expect the price to rise or fall between planting and harvest, but somehow he gets an idea of a probable minimum price at which he will be able to sell. Smith is also assumed to know his cost curves so that he can plot them as *MC* and *ATC* in Figure 30. Since the price is constant regardless of sales by Smith, the marginal revenue is also constant. The horizontal line *PAB* represents both, $AR = MR$.

FIGURE 30

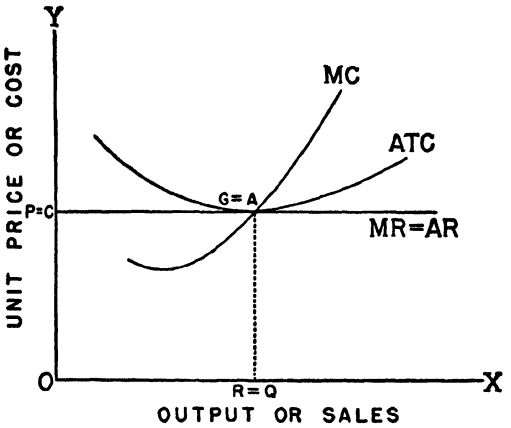
MAXIMIZING PROFIT WHEN SELLING PRICE IS FIXED



The maximum profit output is indicated by the point *A* where $MC = MR$. The price for this output (and all others that Smith might determine) is *OP*. The profit per unit is $AR - ATC$, or $QA - QE$, which is *AE*. Output *OQ* is also equal to *CE* or *PA*. Therefore, total profit is shown by the rectangle *PAEC* whose area is obtained by multiplying output *CE* by unit profit *AE*. Under the cost-revenue conditions assumed for Smith, he benefits by producing more than *OR*, which would have given him his minimum unit cost. *AE* is less than *HG*, but this loss in unit profit is more than offset by the gain from selling the additional product, *RQ*. The amount of this extra profit is shown by the triangle *GHA*. Each unit beyond *G* (*OR*) has an additional cost shown by the height of the *MC* line *GA*. But each unit brings additional revenue shown by the height of the *MR* line *HA*. The difference in each case is positive up to output *OQ* or point *A*. Beyond point *A*, *MC* is above *MR* and additional output would be produced and sold at a loss.

FIGURE 31

ZERO PROFIT WHEN SELLING PRICE IS CONSTANT



At the same time that Smith makes profit of the magnitude $PAEC$, Brown may be just breaking even, as shown in Figure 31. Here MC intersects MR and ATC at the same point, $G = A$. This point also equates maximum profit output OQ with minimum unit cost output OR. Brown cannot improve his position by producing more or by producing less, because moving in either direction from $G = A$ will make his unit cost exceed his unit revenue.

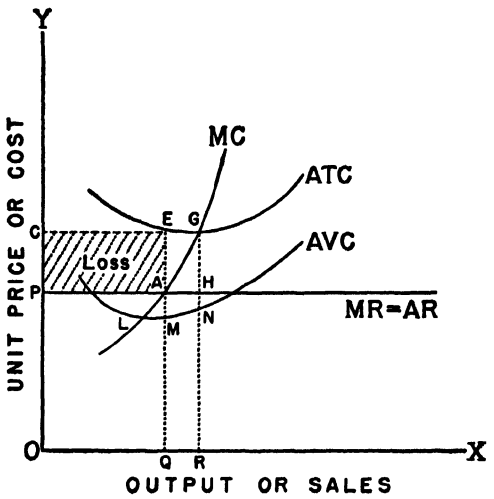
Jones, however, is in a still worse position. As shown in Figure 32, Jones cannot break even at any output. His task is to minimize his loss. This is done by producing at output OQ indicated by point A where $MC = MR$. This is better than producing OR which would give minimum unit cost RG but would bring greater loss.

Someone might ask why Jones produces at all. The answer could be that his AVC (dotted line) curve cuts MC below PAH and therefore he makes something more than his out-of-pocket expenses. This operating profit may be used to pay part of his fixed costs which otherwise would be entirely unmet. Or the answer might be that Jones is just getting started and expects to have lower costs in the future.

Other reasons might be given which often ap-

FIGURE 32

PRODUCING AT MINIMUM LOSS WHEN SELLING PRICE IS CONSTANT



pear in agriculture, but they are contrary to the assumptions used here. The cost and revenue curves are expectation curves, not realization curves. If the latter were the case, Jones might start with the expectation of a higher AR curve or a lower ATC curve and find afterwards that he had made a serious mistake in his estimates. The labor of the farmer is implicitly included as a cost in AVC and ATC above. Most farmers do not figure costs that way and hence have much lower curves. As a matter of fact, most small farmers do not do any calculation planning of this type at all. They go rather by hunches, reactions against failures, imitation, rules of thumb, health of livestock, and the limits of human strength. A few college trained men and large scale farmers may do differently, but not many of these have any concept of MC and MR curves.

12. Graphic Demonstrations of Profit-Maximizing Situations When Price Varies Inversely with Quantity Demanded.—The next three sets of graphs deal with cases where the demand as viewed by the individual seller is a function of price. Marginal revenue differs from average revenue (cf. Chapter 4, Section 4). But the problem remains the same. The manager must decide what sales-output quantity will maximize his firm's profits. The solution is usually given, not in terms of output, but in terms of the profit-maximizing price. Since the price and the quantity are inseparably connected, the choice of the one determines the other. There is only one choice, not two, and there is no fundamental difference between this choice and that of the farmer.

The most common graph in this group is that of Figure 33. It is like Figure 30 in that Ross, the producer, is making a good profit. Again the crucial spot on the graph is point A where $MC = MR$. This determines the maximum profit quantity OQ, from which a line may be projected vertically through A to B on AR. Moving horizontally to the left, we find point P on OY. OP and OQ, the price and quantity, are simultaneously determined by point B. When Ross finds his best quantity, he has also found his maximum profit price. The profit itself may be shown by finding the point E which shows the unit cost of the output OQ (E is at the intersection of QB and the ATC curve). Unit profit then becomes QB minus QE, or EB. Multiplying by the output, the total profit becomes the shaded rectangle $PBEC$.

One significant difference should be noted between the solution in this case and that of Figure 30. The most profitable output, OQ, is *less* than the capacity output, OR, whereas in Figure 30 it was

greater. Some of the implications of this difference will be explored in a later chapter.

The second figure in this group shows producer Adams just breaking even (Figure 34). His *ATC* and *AR* curves are tangent at point $B = E$ on the quantity perpendicular QA. The third figure, 35,

FIGURE 33

MAXIMUM PROFIT PRICE
(U-shaped marginal cost)

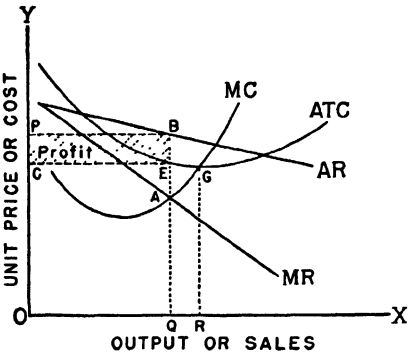


FIGURE 34

MAXIMUM PROFIT AS NORMAL PROFIT
(Selling price varies inversely with output)

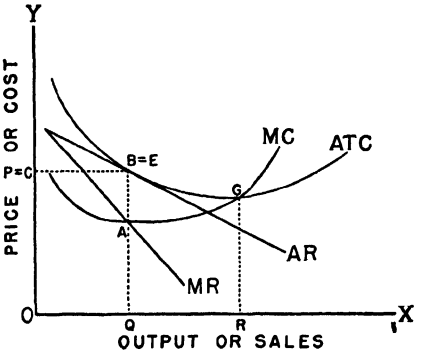
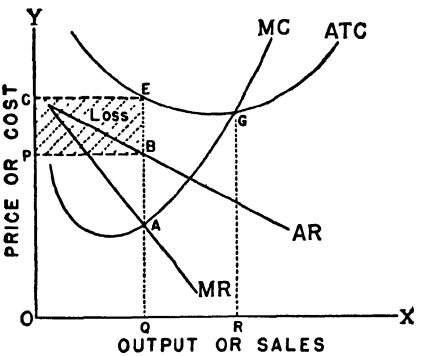


FIGURE 35

MINIMIZING LOSS



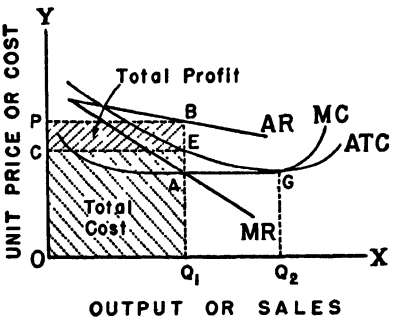
shows producer Thomas operating at a loss, *CEBP*, but minimizing that loss by choosing output OQ (price OP) whose $MC = MR$ at A.

13. Special Situations May Be Shown by the Shape of the Curves.—Most manufacturing and distributing firms do not have deeply convex cost curves. *MC* curves may be flat-bottomed for a considerable range, as in Figure 36. The solution is the same as

before. The only difference is that OQ and OR may be further apart when firms operate under constant marginal cost. The same type of MC curve also may be used to show a manufacturer just breaking even or operating at a loss.

FIGURE 36

MAXIMUM PROFIT PRICE WITH CONSTANT MARGINAL COST



Another situation of interest is that in which the best output is zero. That is shown in Figures 37 and 38 where the expected revenue is so small that out-of-pocket costs cannot be regained. Farmers, particularly fruit growers, sometimes find themselves in this predicament. The demand falls so low that the market price will not cover the costs of hiring the labor to pick and ship the crop. In Figure 37 the marginal cost and the average variable cost are both higher than the potential net revenue shown by $MR = AR$. The manufacturer who sells in a market where his demand is a function

FIGURE 37

BEST OUTPUT IS ZERO

I. Constant Selling Price
(An orchard owner's view at harvest time)

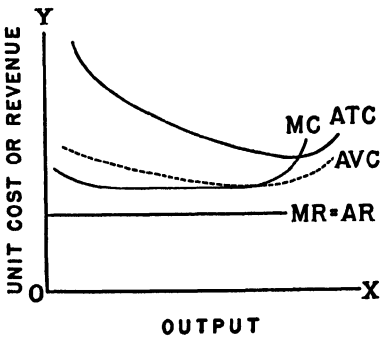
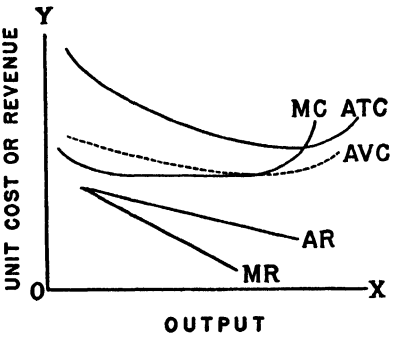


FIGURE 38

BEST OUTPUT IS ZERO

II. Price Falls as Quantity Rises
(Manufacturer's view)



of the price faces the situation shown in Figure 38, where AR and MR slant and diverge, but at no place exceed MC or AVC.

Again the reader must remember that these are planning graphs, not records of historical events or experiments. They portray how

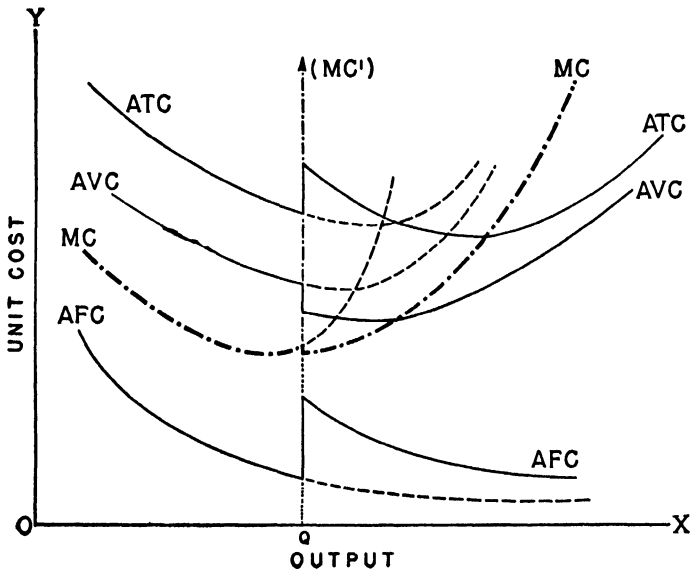
the probable cost and revenue situations look to two different individuals who are potential producers. Production occurs in hopes that the situation will be like that of Figure 30, 33, or 36. But if the producer finds he is mistaken, a different picture must be drawn, such as Figure 37 or 38.

Cost curves will have upward jogs whenever new "fixed" factors are added. The simplest group of cost curves assumes that there is only one fixed factor which remains constant throughout the entire range of output. In reality the only difference between "fixed" and "variable" factors is the greater frequency with which inputs have to be increased as outputs expand. Direct labor and materials are continuous variables. Indirect labor, like that of foremen, must be increased intermittently. Hand tools and smaller machines are added less frequently than larger machines. And so on. Graphically this introduces upward jogs in the *AFC* and *ATC* curves. *MC* rises abruptly to a point *MC'* above the upper limits of the usual graph at the moment the new "fixed" expense is treated as a variable. After this point *MC* falls below its previous level and continues its rise as shown in Figure 39A. Similarly *AVC* has a small downward jog, but not enough to offset completely the upward jog in *AFC*.

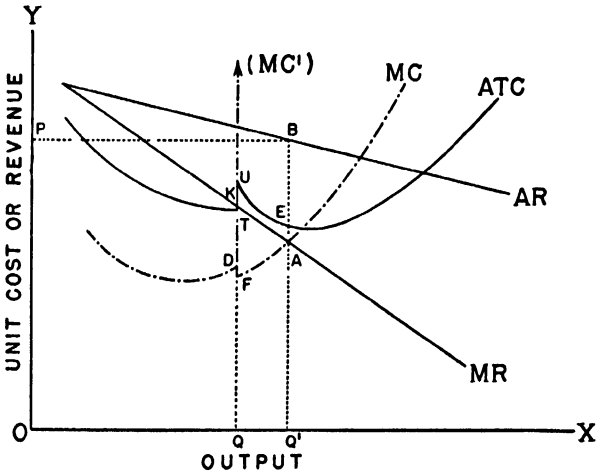
FIGURE 39

INCREASES IN "FIXED" FACTORS

A. The Cost Curves



B. Maximum Profit Point



Graphs of this type are clearly irreversible. They are pictures of cost changes with increasing output and would be different were output to decline. MC crosses the MR curve at two points, K and A as in Figure 39B. The former is clearly an exceptional case caused by the method of treating the additional "fixed" expense as an incremental expense at that particular point. Therefore the maximum profit position is determined by A and the optimum output is OQ'.

Another problem of cost curves in price setting is the question of whether or not to include normal profit. In the foregoing diagrams this cost was omitted, but in a later chapter the same sort of curves will be presented with it included. The whole subject is too involved for more than mentioning at this time. (See Chapter 10, Section 13.)

14. Some Observations on the Meaning of Cost Curves for the Firm.—In conclusion a few observations may be offered to introduce whatever realism is possible in the general subject under discussion. Few business managers use the $MC = MR$ approach in determining their most profitable selling price. The MR schedule is the hardest one to estimate. It requires business forecasts which are far beyond the capacity of most firms. Even expert estimates of probable future demand are largely guesses. Demand fluctuates fairly rapidly for most products. If the $MC = MR$ formula were used rigorously, the selling price would have to change frequently. These changes would themselves affect future demand. Therefore most firms which independently set their own selling prices for new products usually

follow some form of the markup rule. They hire cost accountants to tell them the base figure and then guess at the best markup. A few enterprisers estimate the price at which they can sell a certain output and then make plans to produce that much if they think costs can be kept below that price.

Once prices have become established, the pricing problem becomes one of change. Many times demand rises or falls and variations occur in cost without any change in the asking price. The inconvenience, expense, and market disturbance of changing price lists discourage application of the $MC = MR$ rule even by those who know it. Even when major changes in demand or cost do force changes in prices, many long-run considerations are influential in determining the new figure. Some of these will be examined in a later chapter.

Yet profit maximizing does remain a motive. Even crude price-setting methods strive to achieve it in either the short or the long run. Perhaps the $MC = MR$ equilibrium point is better seen as the norm about which guesses cluster than as the price-determining mechanism itself. Whether realized or not, norms are useful concepts, but they should not be confused with the actual economic behavior of real people.

15. Summary.—If producers could determine cost and revenue schedules for their future business they could compute the maximum profit output by use of the $MC = MR$ formula. These schedules are usually very difficult to obtain with any accuracy. Other considerations usually cause departure from the formal solution even when it can be calculated. Certain useful principles, however, may be derived from a study of this approach.

1. The *quantity* of a good which one seller will produce and offer for sale at a given place within a given period of time is a function of:
   (a) The seller's expected cost and revenue schedules
   (b) Chance factors which may upset plans
2. The *price* which one seller will ask for the goods he produces and offers for sale is a function of:
   (a) The power of the seller to set his own price independently
   (b) The seller's expected cost and revenue schedules
   (c) The formula applied to these schedules. The $MC = MR$ point indicates a probable norm, but there are many departures from it.

Chapter 9

DETERMINANTS OF SELLING POLICIES

1. Statement of the Problem.—Often the choice of the best selling price involves much more than the simple methods of the three preceding chapters. Selling cost may be added to production cost. Products may be improved or debased. Prices may be set high or low relative to current cost of production. Prices may be kept constant when costs and demands change, or they may be adjusted accordingly. One cannot understand the workings of our modern capitalistic economy without knowing the determinants of these choices.

Selling policies may be defined as those ways of action an entrepreneur chooses to maximize his profit by influencing his volume of sales. They fall into two main groups: (1) price policies and (2) demand stimulation policies. The former include markup percentages, stability, and discrimination. The latter refer to advertising, product improvement, and other ways of attempting to increase buyer desire for the product. Demand stimulation affects demand intensity, or demand in the schedule sense. Efforts in this field cost money and must be compensated by increased revenue if they are to be continued. Price policies affect schedule demand indirectly through their long-run influence on consumer desire. Their direct effect is upon demand in the market sense: the higher the price the lower the demand, etc.

The various facets of this complex problem may be seen from the following list of questions to which answers are given in this chapter:

1. What are selling costs and why do producers incur them?
2. Under what conditions does advertising pay?
3. How do selling costs influence the maximum profit price computed from cost and revenue curves?
4. How does product differentiation affect the pricing policies of sellers?
5. What long-run perspectives may influence sellers' price policies?
   (a) The attitude of buyers
   (b) Price policies of competitors
   (c) Entry or exit of competitors
   (d) Public regulation
6. What are the different types of price discrimination, their causes and their effects?

2. From Production Cost to Selling Cost.—In the preceding chapter the influence of cost upon supply price was discussed in terms of production cost. For simplicity it was assumed that the only cost decision to be made by suppliers was an estimate of the effect of changes in the volume of output upon unit cost, either marginal or average. On the other side of the market, the demand schedule was presumed to be known. It was either a fixed price for unlimited sales by an individual producer or a schedule in which the demand varied inversely with the price. It is now necessary to move from this over-simplified picture to introduce additional variables for greater realism.

The first step is to introduce the concept of selling cost. Producers do not need to accept the demand schedule as given. In the modern world they often try to increase the intensity of demand by various types of selling effort. Most firms do not sit back and merely accept orders as they are received, but reach out either by salesmen or by advertising, to influence buyers who otherwise would have bought elsewhere, or conceivably not at all. At times there even is pressure to sell some assets so as to buy others, such as the advertising which indirectly seeks to induce people to liquidate their war bonds to purchase homes, automobiles, refrigerators, etc.

Total profit is the difference between total income and total outgo, and each of these categories includes quantities as well as prices. On the revenue side this means that total income is a function of the quantity sold as much as it is a function of the price per unit sold. On the cost side there are not only the problems of the number of units purchased and the price paid per unit, but also certain other expenses associated with the general conduct of the business which are often loosely called overhead. Sometimes selling expense may be linked directly with the number of units sold, as when salesmen are paid a commission which varies with the volume of their sales. More often selling expense cannot readily be linked with each particular commodity and must be considered a part of overhead. In the latter category it represents an expense which is fixed in total amount, but declines per unit as the volume of sales increases. Selling expense must be estimated and budgeted in advance of the selling period or calculated in retrospect as though it had been. Selling cost, however, differs from most other overhead costs in that the latter usually are planned relative to an expected volume of sales, whereas selling efforts are intended to increase that volume, to raise the price at which goods may be sold, or both.

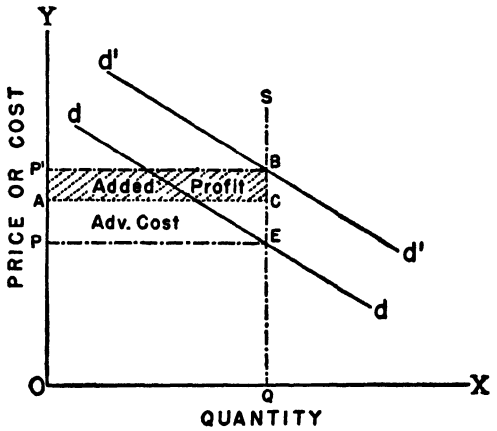
3. Advertising and Other Forms of Selling Expense.—At the retail level the chief form of selling expense to try to increase sched-

ule demand takes the form of advertising. The retailer uses newspapers and radio to tell potential customers about himself, his store, or his products, hoping thereby to induce more people to buy from him than otherwise would have bought. He may stress the high quality of his merchandise, the courtesy of his salespeople, the convenience or prestige of his location, or the generosity of his credit terms, if he wishes to appeal to those potential customers who do not buy with a thrifty eye to price. If his aim is to attract the latter, the retailer will stress the cheapness of his product, invite price and quality comparisons, tell how much the price has been reduced from former levels, "refuse to be undersold," etc.

If the advertising does increase demand, the seller will reap additional profit, provided the increase in gross revenue exceeds the additional cost. The simplest case is one in which the increased demand from *dd* to *d'd'* permits the seller to charge a higher price without loss of customers, as shown in Figure 40. Since the volume of sales remains constant at *OQ*, the only additional cost is that for advertising, *PACE*. This is less than the additional gross revenue, *PP'BE*, and, therefore, the seller makes a net gain from his price and advertising policy as shown by *AP'BC*.

FIGURE 40

ADVERTISING MAY PAY THROUGH HIGHER PRICES



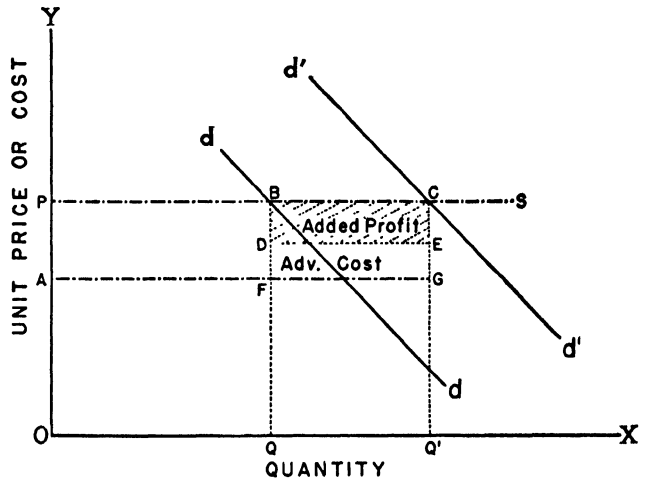
Under other circumstances the price may remain constant either because of voluntary choice or because the maximum price is set by the government, the trade association, or a traditional policy of "follow-the-leader." In such cases individual sellers (or trade associations) may benefit from advertising and other selling efforts which increase the volume of sales.[1] The effect upon profits may be shown under either of two conditions: constant cost (of production), or decreasing

[1] In the case where there are many sellers of a homogeneous good such as wheat or copper, the maximum price is fixed by "the market," and an individual firm would undoubtedly find it too expensive to try to increase demand when the effect of its efforts would be felt by all sellers and could not be concentrated upon its own (undifferentiated) product.

cost. For instance, in Figure 41 the unit cost of operation remains the same, $OA = QF = Q'G$, when sales expand from OQ to OQ'. The area $QFGQ'$ represents the additional operating expense. $FDEG$ is the advertising expense which increases demand from dd to $d'd'$. This leaves $DBCE$ as additional profit.

FIGURE 41

ADVERTISING MAY PAY THROUGH LARGER VOLUME
(Constant Cost and Price)



However, it is more likely that additional sales will reduce average total unit cost through the process of "spreading overhead" and even possibly through economies in operating at the larger volume. This situation is diagrammed in Figure 42 where the unit cost drops from OG to OA as output rises from OQ to OQ', price remaining constant at OP. Advertising cost is represented as $AEFB$ by spreading it over the entire output.[2] If ATC' is the new unit cost curve, the increase in net revenue is $EGHJ$ on the former volume OQ, plus $JCDF$ on the added sales, QQ'.[3]

Both these illustrations assume that the advertising campaign is "successful," that it more than pays for itself. But that is not always the case. Sometimes the increase in gross revenue is less than the extra selling expense. This probably occurs most frequently in

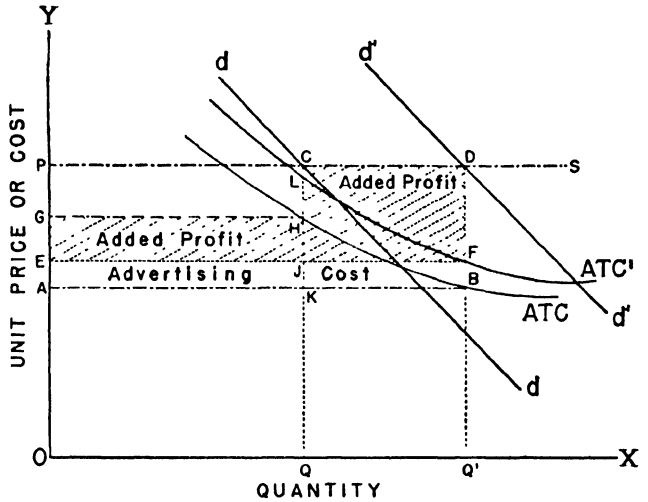
[2] When selling expense is a variable expense like salesmen's commissions, a different diagrammatic technique must be used.

[3] ATC' will not be parallel to ATC because the total advertising expense will have to be distributed over fewer units if sales are OQ than if they are OQ', i.e., LH will exceed FB.

cases of "defensive" advertising where competitors are taking away a firm's trade by their own successful selling efforts or by offering a better product, and where the injured firm tries unsuccessfully to stem the advance of its rivals. Instead of increasing demand by advertising, the result may be merely to prevent it from declining or

FIGURE 42

ADVERTISING MAY PAY THROUGH LARGER VOLUME AND LOWER COSTS
(Price Constant)



from declining as much as it otherwise would. Some cigarette advertising would seem to fall in this category. A similar unfortunate result may sometimes occur in the case of an advertising campaign designed to bring customers to a new store or to persuade people to buy a new article. In short, an outlay for selling expense may or may not pay for itself, depending upon the article, buyers' willingness to purchase, the effectiveness of the particular type of advertising chosen, and the effectiveness of the advertising of competitors.

Although the advertiser generally wants to increase schedule demand by increasing buyer information and desire, he may also wish to change the price-elasticity of demand at the same time. There are two possible alternatives. If an advertiser is contemplating a price cut, as in a "sale" planned by a retailer, he may be content to *increase* the elasticity of demand for the product whose price is to be reduced. In fact, it is very difficult in such cases to distinguish between the increase in intensity and the increase in elasticity of the demand sched-

ule. On the other hand, an advertiser may wish to *de*crease the elasticity of demand so that he may raise his price without suffering a considerable reduction in market demand. From a different approach, there is also the possibility that the advertising may be designed chiefly to decrease the *cross*-elasticity of demand, i.e., to prevent adverse shifts in schedule demand that might result if a competitor should cut his price or improve the quality of his product.

There are, of course, other forms of selling expense besides advertising. They include hiring and training salesmen, providing attractive and convenient places and methods of sale, furnishing customer services such as credit and delivery, etc. Sellers at all levels may spend money to promote the sales of their goods. Manufacturers of branded goods often advertise, and middlemen likewise. In fact, middlemen count on the advertising campaigns of manufacturers to help them sell to retailers, and the latter receive selling help from the advertising of all who previously handled the good. Such advertising does not help a retailer competitively against other retailers handling the same good, particularly if the manufacturer insists that all retailers in a certain area quote the same price. But it will help if a retailer has the only dealer contract for a given area, like a Ford dealer in Dogpatch, and is trying to sell the product in competition with other products which are claimed to be equally good, or better.

Expenditures for advertising seem to follow a law of diminishing additional returns. A small amount of selling expense may be relatively ineffective and the money wasted, but larger amounts may bring increasing additional revenue. Somewhere a peak occurs and thereafter further expenditures reveal diminishing additional revenue. Eventually a point is reached at which the selling expense no longer pays for itself in expanded demand and added revenue. Business firms seek to budget their selling expenditures so that they just reach this point, but its exact location must always remain a matter of guesswork since it varies with competitors' efforts, changes in buyers' incomes, etc.[4] The most profitable amount to spend on selling effort also depends upon the type of sales appeal. Some slogans are far more effective than others, some advertising media give more contacts for the money, some salesmen are better builders of repeat business than other salesmen.

[4] The seller must also decide how much emphasis to place upon selling expense in relation to other methods of increasing sales, such as price cuts, product improvement, etc.

Whether selling expense does increase demand in most cases cannot be proved, since much sales activity is defensive as in the case of rival brands of cigarettes. Each tries to get some slogan or sales appeal which will prove more attractive than the other and their efforts therefore largely cancel out. However, selling expense is a cost of doing business and in the long run must be compensated by income from sales. That is, unit prices asked and received by sellers must cover all costs, including selling expense, if the seller is to remain in business. Where advertising, for instance, does increase demand for a given article, unit costs of production may be sufficiently reduced by the expanded volume of production to pay the cost of that advertising and still permit the seller to quote lower prices than before. (See Figure 42 where $OA + AE$ is less than OG.) But the seller need not lower his price merely because his costs are reduced. It is more likely that he will retain or even increase his asking price to take advantage of the increase in schedule demand. And if the selling effort actually costs more than the decrease in per unit production expense, then a price increase is to be expected. Finally, industry abounds with trade-marked articles whose superiority over similar articles is chiefly an illusion created in the minds of nondiscriminating buyers by the clever and oft-repeated slogans of advertisements. Proof of this contention may be found in the large profits made by certain sellers in this group, such as the producers of cosmetics, as the result of being able to maintain a price far above the cost of production.

4. Product Differentiation Aids Price Control by Sellers.—Most selling effort is conducted by individual firms in favor of their own products. This requires that they be distinguished from similar products in some way so that the buyer can tell rival products apart. For many nondurable consumers' goods, such as food and clothing, the use of trade-marks or brands is the chief method ("None genuine without this signature" or "Look for the red label"). Minor differences in design or "quality" may be introduced to make the distinction more obvious. These become more pronounced among durable consumers' goods such as automobiles, radios, and refrigerators.

Product differentiation of the quality type often involves additional expense so that the seller is faced with the problem of deciding how much he can afford to put into such extras ("chromium grille," "dual horns," "humidor compartment"). His decision on that score is closely related to his ideas about the probable effectiveness of his contemplated selling campaign. A gadget is worth adding only if it

can be magnified by advertising into a very important feature "well worth the extra money" it costs the buyer. To be profitable to the seller, it must yield enough additional revenue to cover more than both the manufacturing and the advertising expense.

Precise calculation of "the most profitable price" under such conditions is obviously impossible. The seller plans his production design and the amount and type of his selling expense according to the best guesses he can make and not according to specific schedules of estimated material costs and marginal revenues. On the expenditure side the number of variables is large, and on the revenue side the imponderables of the ever-changing market confound attempts at arithmetical formulation of maximum-profit prices. If the expected rate of sales does not materialize, the seller may increase the amount or change the type of selling expense. He may reduce the price, offer more liberal credit terms, or do nothing at all, hoping that the current selling efforts will eventually prove successful as the product catches the public's fancy or consumer spending rises. An improvement in the design or a change in the packaging or the name of the product are other possible expedients.

The presence of so many alternative ways of expanding sales makes the precise explanation of asking prices difficult. There is no standard formula or rule-of-thumb which will tell the price at which profit will be maximized in the short or long run. The seller who spends much on selling expense or in improving his product need not make large price cuts to achieve a desired sales volume. Another possibility is that he may set his price high above production costs and rely on selling efforts to evoke sufficient buying to permit him to cover both production and selling expense and make a good profit.

5. Long-Run Perspectives in the Determination of Selling Variables.—When a seller decides upon the magnitude of one or more of the variables involved in sales policy, he ordinarily looks farther ahead than merely the time to dispose of his current output. It is important for him to estimate demand intensity months and years ahead if he takes certain steps in the present. The importance of the effect of price changes upon the desires and buying habits of the people has already been discussed at various points (cf. Chapter 3, Section 7). The effects of variations in product are probably longer lasting, because they build up or tear down the reputation of the seller for turning out "good goods." Advertising and other elements of selling expense often do not have their full effect for months afterwards. For instance, the cumulative effect of a slogan ("They

Satisfy") repeated again and again may ultimately attract customers who were not stimulated to purchase at first. And the word-of-mouth advertising of those customers who actually are satisfied also takes time to get around.

The level at which prices are initially set or the changes which are subsequently made may have long-run effects in four fields with which sellers are frequently concerned: (1) the attitude of buyers, (2) the price policies of competitors, (3) the possible entry of new competitors into the field or exit of existing ones, and (4) the possible stimulation of public regulation. Each of these topics deserves separate discussion.

6. Long-Run Effect of Price Changes upon Attitude of Buyers. —A new retail establishment often features initial prices lower than its competitors', hoping thereby to wean away their customers and get people in the habit of shopping at the new store. After a period of a few weeks or months, prices may be raised gradually to more profitable levels, but there is always the danger that the price-conscious customers originally attracted will then go off to other stores for reasons of longer familiarity or greater convenience.

Established retail stores often feature "sales" of various kinds hoping to draw customers from competitors. Some of the advertised "sale" articles may not be priced lower than before, but are merely played up in display advertising to bring them to the attention of potential buyers, who would not otherwise have thought of buying them. Where price cuts do occur, the seller must always recognize that a part of the business he attracts is made up of old customers, who take advantage of sales to stock up on the featured merchandise. In the course of a year or two they may not buy more of the cheapened goods, but merely space their purchases differently from what they otherwise would have done. The long-run effect, therefore, may be diminished buying in the months ahead by these old customers, and increased buying by new customers, who return to buy other goods after the sale is over (or even form the habit of patronizing the sales of the particular store). Price cuts also may be made defensively to meet those initiated by competitors and to make the consumer believe that he always can shop in that store with confidence in getting the lowest possible prices.

The significance of price competition in the retail field must not be exaggerated, however, since many if not most consumers are not insistent upon getting the most for their money in durability, design, and general satisfaction. Style features catch the eye more often

than price tags, and consumers may not bother to inform themselves about rival products in other stores. Careful judgment is too inconvenient, takes too much time, or requires technical competence to appraise the worth of goods which only a few possess. Hence, a store may rely upon convenience of location, completeness of stocks, efficiency of sales force, generosity of terms, etc., instead of price-quality competition. The advertising of such stores may stress these points, feature a very few genuinely cheap articles, or merely publicize prices with regularity and rely upon repetition to convey the illusion of good buying opportunities.

Manufacturers have different problems to face than retailers when setting their asking prices. In the first place their buyers are more price-conscious, whether they be those who buy for resale or for use in further production. The latter also examine the quality of their purchases with greater skill and care, since business success depends as much upon efficient and economical buying as it does upon diligent selling. Sellers of industrial equipment and raw materials often stress design, technical performance, durability, and convenience more than price. Those who sell finished products to middlemen must consider the effect of their initial prices or price changes not only upon the ultimate consumer, but also upon those intervening buyers who handle the products. A manufacturer sometimes seeks to encourage wholesalers and retailers to carry his line of merchandise rather than a competitor's by conducting a vigorous advertising campaign intended to make the ultimate consumer want the products and ask for them by name at the retailers. Or he may grant "advertising allowances" to middlemen in which he cuts his list price by an amount which the buyer is supposed to use in local advertising. He may fix the final selling price so high that retail margins or commissions of middlemen are larger than usually accompany his type of product. To encourage stocking and pushing new products, special deals of various types may be employed, such as giving one case free for every dozen cases purchased. The variety of expedients is so great that no significant generalization is possible regarding the optimum price in relation to cost.

7. Effect of Price Changes upon Price Policies of Competitors. —The fact that price changes made by Seller A may cause Seller B to make similar or offsetting changes has already been noted in several connections and deserves further analysis. In the first place, although A's price change is obviously a causal factor so far as B is concerned, the quantitative reaction of B is unpredictable. If the product in-

volved is one in which there is high transference of demand from one seller to another when price differentials emerge between sellers (as with many industrial raw materials and a few standardized retail products like gasoline), Seller B is likely to make precisely the same cut (or increase) that is initiated by Seller A. The firm making the first change in such cases usually is the major seller in the field and the other smaller firms merely "follow the leader." If A happens to be a relatively small seller or if most of the firms are operating near capacity, competitors may not follow price cuts initiated by A.

On the other hand, if the industry is one which has much excess capacity, or if the major sellers want to punish firm A which has become too independent to suit them, the reactions of B, C, etc., may be to make price cuts larger than A did. This may prompt A to still further cuts and a "price war" occurs in which most of the firms, if not all of them, soon sell at a loss. Illustrations may be found in the history of retail sales of gasoline, bread, milk, and other relatively standardized articles of wide consumption. At one time railroads likewise engaged in this type of "cutthroat competition" before the Interstate Commerce Commission was granted the power to impose minimum rates and to prevent rebates and other forms of secret price-cutting. A price war ends when continued losses discourage further price-cutting, or bankruptcy forces the weaker firms out of business. Some firm may then raise prices a bit and find that others are tired of the war and willing to follow. The group eventually may restore prices which approach those prevailing before, or exceed them, as in the case of some American railroads in the decade following the Civil War.

In the second place, the possibility that a price cut by A might stimulate similar or greater price cuts by B, C, etc., is something that is likely to be taken into account by the management of A if it is farsighted. Ordinarily such a prospect will be a deterrent, but not always. New firms or those with particularly aggressive management may be willing to take the risk. By secrecy or surprise they may get the selling advantages of low prices for some time before competitors follow suit. One result of such price wars or the possibility thereof is that sellers may get together in a formal or informal way to agree not to do such things. They hope that such an understanding, though unenforceable in the courts, will act as a greater deterrent than purely independent caution. The possible boomerang effect of price action makes for still greater uncertainty about demand schedules than ordinarily exists. Equilibrium theorists point out the logical impossibility of stable prices under such con-

ditions and describe the situation by terms such as *oligopoly, duopoly, circularity,* or *the small group case.*

If products are highly differentiated in fact or in the minds of the buyer (as through effective advertising about the alleged superior quality of a given brand), sellers will be less prone to reduce prices to meet price cuts of a rival. Instead they often try to discredit the rival product as being inferior by bringing out what they call "second grade" brands to sell at the lower price while refusing to lower the list price of their "first grade" brands. Such has been the case in certain retail markets, such as that for automobile tires (chain store competition), gasoline (small refiners and casing-head gasoline), and refrigerators. Major canners have often put "off-brand" labels on their surplus pack to prevent having to make the price cuts which would be required to market their entire supply under the nationally advertised brands. Because of the lack of grade labeling or public testing laboratories, consumers cannot tell the difference. Therefore, the more prosperous among them, or the less thrifty, continue to buy the highly advertised brands instead of others which offer greater value for the money. "Fighting brands" are almost exclusively a feature of retail trade. Industrial buyers are more careful and less gullible. However, the prestige of "big names" undoubtedly does make it hard for new firms to break into the market at any level even though their products are superior in design or quality or are lower in price.

8. Effect of Price Policies upon Entry or Exit of Competitors. —If firm A raises the price of its product very far above the unit cost of producing it, new or existing competitors will try to take away its customers by offering similar articles or effective substitutes at lower prices. The most common case is where A is deterred from raising his price by fear that existing firms will undersell him and seriously reduce his sales volume. To prevent this he may get an agreement among his chief competitors to raise their prices at the same time—or rely on their voluntarily following his lead. Monopolistic agreements of this type to reduce price competition are generally kept secret because of fear of prosecution under antitrust statutes. Ostensible competitors may have interlocking directorates, considerable stock ownership in common, or potent trade associations which secure unity of action in regard to price. This restricts the field of competition to selling efforts and to product superiority. Even these may be limited by agreement, as in the "cartel" arrangements between firms in different countries, which frequently limit the areas within

which members are free to seek buyers, so that each member has his own home territory free from foreign competition of any kind.

Potential sellers may be prevented from becoming actual competitors in various ways. These include the large amounts of capital required to get started (alloy steels); the situation where patents are held tightly by one firm or made available only at high royalty (electronics); exclusive franchise (radio broadcasting); lack of sales outlets (automobiles); lack of skilled labor or management experience; and tariffs (against foreign competition). Any firm which contemplates asking prices which others deem "high" must consider the effect of such prices upon its future competition if it wishes to arrive at a wise decision. Sometimes it will choose to set the high price and at the same time hunt for means to secure cooperation or retard entry.

Stated in general terms, there are probably two main reasons for entry: prices and profits. The prospective entrant is probably attracted most often by high prices which exceed the cost he thinks he would have to incur in producing a similar product or service. In other cases, however, the new entrant is attracted by reports of high profits in the given field. Prices are obvious, profits may be concealed or distorted, and financial reports demonstrating their presence do not usually appear until after high prices have been in effect for some time. Other reasons for entry include a desire on the part of a hired manager to become his own boss, old firms taking on new lines to make fuller use of existing capacity, etc.

When entry occurs, the newcomer is beset with many difficulties, not least of which is the task of convincing buyers that he has a good and dependable product. Since he usually cannot point to his past record, he must offer some special inducement to get his first sales made. This usually means offering goods at less than prevailing market price, granting more favorable terms of sale, or throwing in some "extra" for the standard price. That is why many firms operate at a loss during the years in which their product is being introduced in the market. They spend more for selling expense of various types or they have to operate at such a low percentage of capacity that each unit sold carries more overhead than is customary in the industry.

If the newcomer is successful, he may take away so much of the business of the price-raising firm or firms that they are reduced to a smaller scale of output. To avoid such damaging competition, an existing firm which is tempted to raise its prices may reduce them instead. This will tend to discourage entrants and may even bring

greater profits through the economies of large-scale production which were impossible at the higher price.

Defenders of the capitalistic system of free enterprise acclaim self-restraint and free entry as automatic checks to monopolistic price raising. They are similar to the price competition which allegedly reduces the prices of all firms in a group when one of them introduces a process for reducing the cost of production and cuts prices to exploit its advantage. Rivalry between firms already in a field plus free entry of new competitors is supposed to keep prices so low in relation to cost that no firm or group of firms can long receive abnormally high profits. The picture is undoubtedly true in its main outlines, but these are increasingly blurred as obstacles to entry multiply and innovators seek profit through price maintenance rather than price-cutting. Established producers may also buy out the upstart rather than engage in price competition with him. Therefore, conclusions regarding desirable social policy must be reached in the light of a careful appraisal of each given situation and not by the application of sweeping generalizations either about the benefits of free competition or about the harmfulness of government regulation and public price-fixing.

9. Effect of Price Policy upon Public Regulation.—In deciding his proper price policy the business executive also must take into consideration the possibility that what he does may provoke public regulation harmful to him. In earlier periods of our national history, excessive prices charged by railroads, water companies, gas, light, and power companies, and street railways provoked public regulation designed to curb increases and price discrimination. In more recent years, however, the field covered by the term *public utilities* has become fairly well stabilized as an area for public control and two other trends have developed. The first dates back to 1914, when the Federal Trade Commission was set up to curb various acts known as "unfair methods of competition." Since this form of public regulation is designed to protect the seller, not the consumer, it sometimes seeks to prevent price-cutting. The argument is that a seller should not engage in temporary price-cutting in a given area in order to squeeze out an upstart competitor and then, after that has been achieved, return the price to the former level. The purpose of that kind of price competition being to eliminate future price competition, it is considered antisocial. Other forms of unfair competition are banned by orders of the FTC, but only price matters concern us here.

The second trend is for competitors to *seek* public regulation in order to avoid the profit-destroying effects of price competition. The milk distributors and producers have secured such protection in a number of states by legislation creating commissions to set minimum prices. In California the barbers have similarly organized and protected themselves, with the result that the public is paying almost twice as much per haircut as formerly. Neither of these industries has been declared a public utility, and the consumer has no representative to guarantee that prices will be held down to a "reasonable" level like that vaguely defined for public utilities as a "fair return on a fair value." Since public regulation of this type often has grown out of conditions of destructive price competition, it is like going through a war in which the survivors emerge into the promised land flowing with the milk and honey of guaranteed minimum prices.

There are some retailers, however, especially chain stores, which may read the handwriting on the wall if they press down too heavily upon their weaker competitors. Already most states have enacted "fair trade" laws to prohibit retailers and wholesalers, in certain lines, from selling at prices less than "cost" plus a minimum markup determined in various ways. Small businessmen also have secured the passage of the federal Robinson-Patman Act (1936). This amends the Clayton Act and forbids sellers to grant large buyers any more reduction from the price charged small buyers than is warranted by the economies in manufacture, sale, or delivery which result from selling in larger lots. In similar fashion the price-cutting tendencies of large retail sellers have been curbed by state and federal legislation (Miller-Tydings Act) permitting producers to set minimum resale prices for their goods and to refuse to sell to those who will not sign a contract to maintain these prices. In addition to these laws which limit freedom to cut prices, aggressive selling may also bring down upon chain stores a special tax, varying directly with the number of stores in the chain, which certain states have already inflicted.

During the depression that followed the 1929 crash, the downward trend of prices and spending became so alarming that one of the devices adopted by the New Deal administration to stop it was the National Recovery Administration. Under the NRA *minimum* prices and wages were established and industries were organized into groups to draw up codes of fair competition which limited members' rights to cut prices. The law was later declared unconstitutional, but a precedent was set and many industrial firms still feel the effect of the "price-stabilizing" codes in spite of declarations by the Department of Justice that the antitrust laws are still on the statute books.

10. Types of Price Discrimination.—As suggested above, certain sellers offer large lots of goods at lower prices than small lots. This price discrimination in favor of large buyers is defended on the ground that merchandising costs are less per unit of sale. But that is not the only reason, nor is it sufficient explanation of the spread between low and high asking prices for the same goods. Buyers of the same size are sometimes charged different prices which have no conceivable cost variations to justify them. This is called personal or firm discrimination and is due either to the desire to favor the low price buyer or to his more effective higgling over the price. In times of rising prices, new customers are often charged more than old customers, although they do not always know it. When business is light, the reverse may be true, and the seller may shade his price secretly to get new business. Another common practice is to grant cash customers a lower price than those who take time to pay. If we assume that prices should be proportionate to costs, then whenever they are not, discrimination is made against some one.

11. The Basing Point System Reduces Location Advantages.—If all buyers of steel or other commodity of a certain type and quantity paid the same F.O.B. price at the mill, the nearest buyers would have an advantage over the more distant ones. Their transportation costs would be less. Buyers who could use cheap water transportation would have an advantage over those paying more expensive rail freight rates. These buyers are discriminated against when they are deprived of this transportation cost advantage.

The most famous method of discriminating in this way is the *basing point system.* It is best known in connection with the steel industry. For many years most steel buyers in the United States were compelled to pay a Pittsburgh, Pennsylvania, F.O.B. price plus rail freight from Pittsburgh, regardless of where they bought the steel. A Chicago firm might buy steel from Gary, Indiana, but it had to pay the Pittsburgh price plus freight from Pittsburgh ten times farther away. A firm in Laramie, Wyoming, might have the same treatment when it bought steel from Pueblo, Colorado, nearby. These buyers lost the advantage of being near steel mills.

Discrimination of this type depends upon some type of monopolistic price-fixing. There may be only one producer with several factories geographically scattered. The Ford Motor Company, for instance, has many assembly plants, but all Ford dealers pay a price which includes freight from Detroit. Or there may be some type of understanding among competing producers of a standard commodity

like portland cement or linseed oil. Such firms would all quote the same prices to all buyers in a given city even though purchases were made from different sellers in different cities. Each quoted price would be based upon an agreed F.O.B. price plus an agreed transportation cost from an agreed basing point.

Another form of location discrimination occurs when manufacturers establish the same delivered price for all buyers wherever they may be located. The most extreme case is found in retail commodities sold at the same price throughout the nation, like Hershey bars or the latest popular novel. This form of uniform delivered price has sometimes been called "postage stamp pricing" because of the uniformity of rates on first class mail regardless of distance.

The size of the market area in which all sellers quote identical prices depends upon the rail freight structure. Some mail order sales have two zones, one east of the Rockies and another west of the Rockies. If the rail freight rate on a certain type of steel pipe is the same from an eastern basing point, such as Sparrows Point, Maryland, to Los Angeles, San Diego, and San Francisco, the uniform delivered price region will be very large. Ordinarily, however, the transportation rates to different cities as far apart as these vary considerably and therefore their buyers do not pay the same price.

The major economic significance of uniform delivered prices lies in the fact that buyers do not gain from being located close to a producer. A San Francisco buyer might get his steel from Pittsburg, California, only thirty miles across the bay, but he could not get it any cheaper than the Los Angeles buyer, 450 miles to the south. Both would be charged the base price at Sparrows Point plus freight to the west coast, even though the steel originated in Northern California, and this western producer would pocket the extra transportation charges, often called "phantom freight."

12. Other Aspects of the Basing Point System.—A uniform delivered price system of one type or another has been used in thirty or more major industries, including asphalt roofing, cast iron pipe, cement, copper, gypsum board, linseed oil, lumber, soap, steel, sugar, tiles, and zinc.[5] Although its historic origin has differed in different industries, the chief purpose seems to have been to prevent price competition which arises when new firms commence business at points closer than old firms to important consuming centers. Buyers

[5] An excellent summary and criticism of the uniform delivered price system may be found in TNEC Monograph No. 42, *The Basing Point Problem*, which contains the views of both the United States Steel Corporation and the Federal Trade Commission.

otherwise would tend to buy from their nearest supplier in order to save on transportation costs. This would cut out the older firm from markets near the new one and prevent the newcomer from invading the territory near the old one. The chief reason why new firms cooperate in such a monopolistic system probably is fear of local price cutting by the old firms, which could afford to sell at a loss for a while if they thought it necessary to prevent losing all of the business in a given area. The new enterprises must rely for their selling appeal either on speedy delivery or on buy-at-home-from-those-you-know arguments, since the article involved is ordinarily identical in form and quality regardless of who produces it.

Sometimes the new firms become so well established that they are able to force the older concerns to grant them the privilege of having their own basing point, although they usually continue to use the base prices of the industry leader. Pressure from buyers and from the government also have been responsible for introducing a multiple basing point system in place of a single basing point. The delivered price in localities near the new basing point will fall because freight charges from that point are less than from the former one, but uniformity of delivered prices will remain. Sellers in the old basing point city and elsewhere will revise their price lists to adopt the delivered prices of the sellers in the new basing point cities. This means that if they succeed in making sales in the natural market area of the latter, they will not be able to charge the full freight from their own shipping point, but will have to absorb it. For instance, if Birmingham, Alabama, is considered the new basing point, Pittsburgh, Pennsylvania, the old basing point, and New Orleans the place of sale, then the change gives New Orleans buyers a price reduction equal to the difference between freight rates from Pittsburgh and from Birmingham. Pittsburgh sellers, however, must absorb this difference, approximately equal to the freight between Pittsburgh and Birmingham.

It is obvious that the most economical procedure for the country as a whole would be for buyers to purchase from the source which can produce and ship goods the most cheaply.[6] Under the basing point system this rule is violated. Goods are shipped from west to east which could be bought in the east, and from east to west which could be bought in the west. There is much of this "cross

[6] It is often wise, however, not to concentrate all orders on the cheapest source of supply in case it might be closed by a fire, strike, or other difficulty. In periods of generally tight supply and rationing those with many suppliers have generally fared better than those with only one.

hauling," with attendant waste of manpower and resources in transportation and selling efforts, plus a higher average level of costs to the consumer.[7]

The basing point system has many variations, all of which involve price discrimination. If the number of basing points is large, the amount of discrimination is small, but it does not disappear unless every production center is a basing point for every item, in which case there would be no reason for having the system since it would be equivalent to F.O.B. pricing. The base prices at different basing points are often uniform, but uniformity in this respect is merely monopolistic, not discriminatory. Where base prices differ, they may reflect differences in cost of production, and to that extent the market price structure is more competitive.[8]

13. Requisites for Price Discrimination.—When a seller charges one buyer or group of buyers a higher price than another, he must rely upon the inability of the low-cost purchasers to resell at the higher price. This is achieved in various ways: (1) by keeping the price differential less than the transfer costs between the two buyers (includes costs of handling, transportation, and taxes, such as excises or tariffs); (2) by refusing to install the necessary equipment for the buyer who does not buy directly (telephone service); (3) by contracts prohibiting resale; (4) by the perishable nature of the article; or (5) by keeping the high-price buyer in ignorance of the discrimination against him. Efforts may also be made to convince the high-price buyer that the apparent discrimination is really to his benefit.

The latter argument is based upon the following line of reasoning: (1) The high-price buyer A is originally charged a price which, when paid by all buyers, is just sufficient to cover direct costs, indirect or overhead costs, and a normal profit. (2) If sales can be made to B only at a lower price than that charged A and others like him, A may yet gain if two conditions are fulfilled: (a) the price charged B is in excess of the direct costs of supplying B, and (b) the excess is used to reduce the total overhead costs that must be charged to A. However, if the excess is used only to swell the profits of the seller, A does not benefit. Another approach is to assume under (1) above that

---

[7] As a make-work policy during the depression of 1929–1939 some government agencies, it is reported, made a practice of ordering supplies from the most distant firm of those submitting identical bids.

[8] The basing point system of pricing was outlawed for the cement industry in a 1948 decision of the U. S. Supreme Court. The ban of this decision may be extended to other industries, but evasionary tactics are probable.

the revenue from firms charged A's high price is not sufficient to cover all costs and that the seller is faced, therefore, with the necessity of discontinuing business or raising A's price if sales cannot be made to B which will meet the deficit in overhead costs. The difficulty lies in convincing A that B really is being charged as much as he should, or that the seller is as nearly bankrupt as he contends.

Prominent illustrations of long-continued price discrimination are found in charges for railroad transportation, water, electricity, gas, and telephone service. In rail history the most famous example is the long-short haul controversy. To meet competition from water carriers or other railroads at terminal points, railroads have often charged more *per ton-mile* for hauling identical merchandise a short distance to a nonterminal point than for a long distance which includes the short one (from New York to Denver as opposed to from New York to San Francisco). Protesting inland towns were told that they benefited from the discrimination, since otherwise they would have no rail service at all. Somehow they could not envisage the rails being torn up, but the best they could do was to secure the passage of legislation prohibiting the extreme discrimination which had existed when the *total* charge for the short haul exceeded that for the long haul (larger charges per ton-mile were still allowed). Less obvious examples in the field of railway rates are the higher charges for articles of high value and small bulk (silk, tea) than for cheap articles of large bulk (cement, coal). The allegation that railroads and other public utilities "charge what the traffic will bear" attests the fact that those who must pay the high rates usually feel that the sellers are raising the price to them instead of just lowering the price to others for the benefit of those who protest. It is an axiom of monopolistic price-fixing that the monopolist sets his single or multiple rates as high as he can in the light of the prospective demand for his product by various buyers.

There are also many cases of departure from the published basing point prices discussed in the preceding section. When purchasers want the product modified in size, quality, or delivery date from what is customary, the seller is involved in additional expense. This has led to the imposition of charges known as "extras." If the leading producers operate monopolistically, they may consult with one another about these charges and set up a scale of extras which is supposed to apply to all sales, like the underlying basing point prices themselves. In such a case the extras will only approximate the additional costs involved and may be shaded upon occasion. Departures from the basing point prices themselves have also been common in

some industries, notably steel. In efforts to get orders from very large consumers, price concessions are frequently offered, but the smaller firms and governmental buyers are forced to pay the full list price. These price cuts, usually secret, cannot be justified on the grounds of the economies of large-scale operation, since they bear no uniform relationship either to average or to marginal costs and since the government, though a large buyer, gets no cuts at all. The extent of price concessions varies instead with the degree of economic concentration in the consuming industries, i.e., with their bargaining power.[9]

14. Limits of Price Discrimination.—If we assume that a seller knows the demand and cost curves for his product throughout a wide range of quantities, which is unlikely, we can state with precision the prices which he should charge to maximize his profit. Having determined the most profitable uniform price, he should then subdivide his customers into as many groups as possible at prices higher than this one. He should also seek additional customers at prices below the most profitable uniform price until his total sales have expanded to the point where marginal cost equals market price. The problem may be diagrammed by Figure 43 which should be compared with Figure 33. If all goods must be sold at a single price, the best price is $OP = QB$, determined by the $MC = MR$ rule of Chapter 8, Section 12. If the seller can discriminate against the buyers willing to pay more than OP, he should seek to exploit as much as possible the demand represented by section SB of the demand curve AR. The more finely he can subdivide these buyers and ask each group the full amount it is willing to pay, the closer he will come to the theoretical maximum profit increment, the triangle SPB. Having done this, the seller might seek also to exploit the demand at prices below $OP = QB$ on the segment BU of the demand curve. This is often called "dumping." The maximum gain to be derived from this type of discrimination is shown by the triangle BAU.

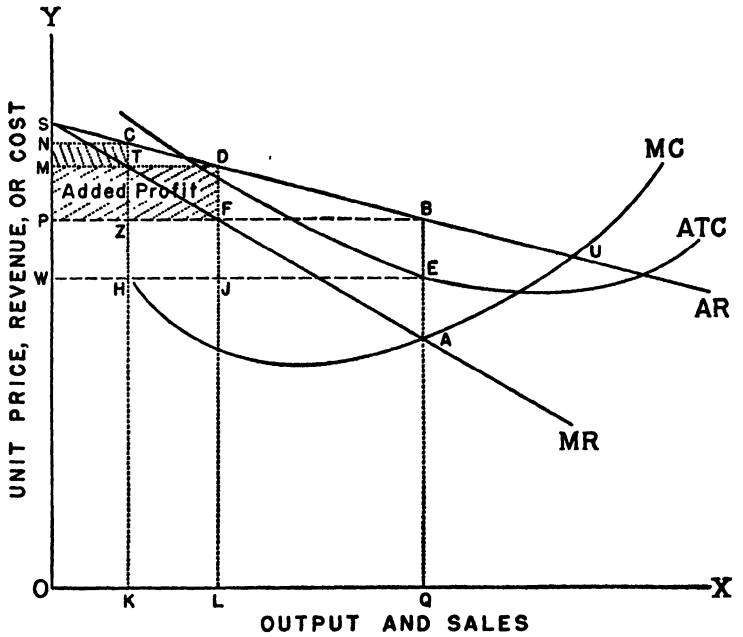
In actual practice the seller does not know the full shape of his demand curve nor is it constant from day to day. Hence the best he can do is to make guesses and use trial and error methods. This

[9] John M. Blair and Arthur Reeside make the comment in TNEC Monograph No. 41, *Price Discrimination in Steel,* Washington, 1941, that "the obvious effect of these concessions to large buyers is that small purchasers are placed at a competitive disadvantage. No elaboration is needed in describing the position of a small manufacturer who in February 1939 was forced to pay for a ton of cold rolled strip a delivered value of $154.69, netting the mill $145.41, as against a large competitor, who, for this product, paid a delivered value of only $76.19, a mill net of but $70.88" (p. 26).

usually results in prices that are too high for his own good rather than too low. For instance, if the seller in Figure 43 aims at the group which would pay *OM* for *OL* units but asks them to pay *ON*, he loses entirely *KL* units of sale so that his gain (*PZCN*) from trying

FIGURE 43

DISCRIMINATORY PRICING



to segregate buyers is less than expected (*PFDM*), and may even prove to be a loss (if *PZCN*, the approximate gain in profit, is less than *ZFJH*, the approximate loss of profit).

15. Summary of Economic Principles Regarding Selling Policies.—The three groups of economic principles explained in this chapter continue the supply-side exposition of the three preceding chapters, but some of them refer back to the chapters on demand. They clearly are supplementary to the principles described earlier and in no wise constitute a sufficient explanation in themselves. The three groups are as follows:

A. Demand intensity and elasticity are a function of:
   1. The amount and type of selling effort by the seller
   2. The amount and type of selling effort by rival sellers and those selling complementary goods

B. The asking-selling price of any given seller is a function of:
   1. Selling expense
   2. Price policies of competitors
   3. Expectations about reactions of competitors to own price policy, and the ultimate effect of those reactions upon himself (oligopoly)
   4. Expectations about reaction of public control agencies to price policy. This will be a function of:
      (a) Changes in particular prices in relation to the general price level
      (b) Profits relative to other firms and other times
      (c) Degree of price discrimination

C. The degree of price discrimination between buyers is a function of:
   1. Differences in cost of sale to different buyers (vary with quantity per purchase, credit terms, delivery and freight charges, etc.)
   2. Monopoly power to separate buyers and prevent resale among them
   3. Elasticity of demand by classes of buyers
   4. Elasticity of supply (ease of entry of new competition as price is raised for certain buyers)

Chapter 10

PRODUCTION POLICIES INFLUENCE PRICES

1. Production Choices of Business Executives.—Prior discussions of supply and price have assumed that the decisions to be made by business executives were limited to (1) price to be asked, (2) output to be offered for sale, (3) proper amount and types of selling expense, and (4) best design or quality of product. Decisions must also be made in three other fields: (5) what to produce, (6) how to produce it, and (7) when to produce it. Although these problems are among the first to be tackled by managers when an enterprise is begun, they are less familiar to the average person than the ones already examined and hence have been deferred to this point. Much of the previous discussion will prove useful background material for this further analysis of the problems of supply and price.

2. Deciding What to Produce and How Much.—To state that an enterpriser will choose to produce that product which will bring him the most money is to state an axiom of capitalist economics. But it ignores other possible motives and it does not tell us anything about the conditions on which such a decision depends. Some individuals choose to produce that which they enjoy producing, even though it is relatively unprofitable. Such choices, as well as pecuniary ones, usually are related to the prior experience and skill of the enterpriser. The possession of specialized equipment is also likely to influence his decision. Many decisions to begin small enterprises are made impulsively without careful appraisal of costs and revenues. In larger firms more care is exercised and exhaustive engineering and cost surveys often are made. The larger the investment needed to start producing a given product, the more likely is the new entrant to be a firm which already is producing something else. If relative market prices change so as to make product B appear more profitable than before, and A less profitable, a producer whose equipment is not too specialized will shift his output from A to B. This often occurs in farming areas where soil and climate permit producing a variety of crops. A change in relative costs of production will produce similar results.

If only one article is being produced, the problem of how much to produce can be solved in the abstract according to the rules outlined in earlier chapters. Those who must accept or who choose to accept the selling price given by the market, market leader, monopolistic agreement, custom, or government, ordinarily will plan to produce only as much as they can sell at the given price. At times, however, the rate of production may exceed the rate of sales or fall below it as management chooses to build up or to diminish inventories. In order to keep valued employees, many firms produce to stock during slack periods or business recessions, hoping for future sales to expand and reduce accumulations. The output of annual crops is planned with less care, partly because the price is not clearly predictable at the time of sowing and partly because the weather may prove to be a more important determinant of output than human decisions about planting acreage. Finally, those who are in a position to set their own selling prices usually adjust production rates to current sales, plus or minus changes in inventory.

If two or more articles may be produced alternatively with the same material or labor, the most profitable quantity of each to produce will be a function of the technical efficiency of production and of the price of each, assuming that these are given and constant. A high price may be offset by high cost (low efficiency) of production. Assuming that unit cost of production varies with output, the relative output of each should be adjusted until the marginal costs of each bear the same ratios to one another as do the selling prices to one another (the marginal rate of substitution equals the ratio of their prices). This is reasonable only under the further assumption that the producer can sell as much of each as he wishes to produce without changing the market price when he varies the size of his offerings—an assumption of the theory of pure competition, but rarely encountered in the business world. It is more frequently true that if he wishes to vary his output and sell without inventory accumulation or decumulation, he must change his asking price (or selling expense). Under these circumstances one can repeat only the earlier generalization that profit will be maximized when marginal cost equals marginal revenue for each of the items produced.

3. Entrepreneurial Adaptation to Shifts in Schedule Demand.—At this point it is again necessary to emphasize the distinction between two types of entrepreneurial choice. In the preceding section and most earlier ones we have taken demand as fixed and have assumed that the entrepreneurial problem was to choose the price and

output which would be most profitable. Business managers also have to face the problem created by shifts in schedule demand caused by competitors' price and selling policies, by seasonal change, by general advances or declines in business activity, and the like. In the multiple products case examined in the last section, we may assume, for instance, that the demand for product A rises so that the former volume can be sold at a higher price, or a larger volume can be sold at the same price (or both price and volume may rise). The problem now is how to adapt production rates of each of a group of products (A, B, C, etc.) to the increase in schedule demand for A.

Under conditions of pure competition, theory states that the price will rise and that the individual firm must decide whether to shift some of its efforts from producing B, C, etc., to increase the output of A, or to expand A while keeping B, C, etc., constant. The latter alternative would require an increased employment of the factors required to produce A (more labor and land used for alfalfa). The former might involve merely a transfer of factors and a diminished output of one or more of the products B (barley), C (corn), etc., which are technically substitutable for A in the production process. If, however, one of the other products, such as B, is complementary with A in production (cf. wool and mutton), then its output will expand, not contract. The most profitable expansion of A (and its complements) will be reached when the marginal cost for A equals the new market price for A. Regarding other products, the most probable situation is one in which the firm maintains its former volume of output for them. If, however, their prices decline as a result of the increased demand for A, the firm should produce less. Finally, it is possible that the total supply of a limited factor (such as a farmer's land) may be diverted from producing other products to concentrate exclusively on A, whose price has risen. Or if the transferable factor is one which can be employed only at a rising marginal cost of production of A and is also limited in supply (such as labor in an isolated region), then output of A may expand to the $MC = MR$ point and the factor which remains after that volume is reached may be devoted to producing the less remunerative products B, C, etc.

4. Deciding How to Produce: Choosing Proper Amounts of the Factors.—Wherever price competition is keen, business managers are forced to give much thought to methods of reducing costs. Sometimes this is achieved by driving hard bargains with those from whom purchases of materials or services are made. But more often the economies arise from devising better combinations of the factors,

that is, inventing new processes for factory, farm, mine, or office which permit producing the same amount of product with less total outlay for the factors used. These inventions may save either labor, capital, or land, or any combination of these in their various types. This is popularly known as securing "increased efficiency" and may be measured either in terms of decreased physical inputs (or total dollar cost of inputs) or in terms of increased total output if the total expenditure for inputs remains constant.

It is obvious that two things chiefly determine the most economical combination of the factors: their unit prices and the technical knowledge and skill with which they can be made to work together. Physical efficiency determines the physical cost of each unit of product, and through unit prices, the money cost. Therefore, with prices and efficiency given, the firm will make the best use of available factors when it uses them in such quantities that the price ratio between any two factors equals the marginal rate of substitution between them. The advantage of substituting one factor for another lies in the relatively greater value product per dollar spent for the former. Therefore, when the marginal value products per dollar of input for each factor purchased are equal, there is no further incentive to substitute. This concept of equilibrium may be formalized by the equation:

$$\frac{\text{Price of A}}{\text{Price of B}} = \frac{\text{Marginal value product of A}}{\text{Marginal value product of B}}$$

If we assume, for instance, that the price of one of the factors changes (wages, material costs, interest, etc.), there will be a change in the quantity used and therefore in the ratio between the quantities of the various factors employed. Specifically, let the price of factor A decline for some reason or other. A firm will tend to adapt itself to this change by increasing the use of A both absolutely and relatively to other factors. The quantity of substitute factors will decrease absolutely, while complementary factors will be employed in larger quantities than before, until a new equilibrium is reached according to the above formula.

It is very important to realize that most factors are both complementary and competitive. They are complementary in that both are needed in production, but competitive in that substitution occurs between *marginal* units. A farmer needs both land and labor, but if the price of land rises more rapidly than the price of labor, he will change the ratio between the units of each factor that he employs by using less land or by hiring more labor. In some cases the tech-

nology of production requires that certain factors be used in fixed proportions, such as one engine per auto frame, or one driver per truck, and changes in relative price are ignored. These perfectly complementary factors may be considered as single units in applying the formula for equilibrium of substitution in relation to other factors where competition at the margin occurs. In most instances of apparently fixed proportions, however, the fixity is due to the smallness of the price changes that are likely to occur. If, instead, the price of one factor goes way up as a result of a war or embargo, there is a strong incentive to economize it, and ratios that were considered fixed may be changed. Or an additional variable may intrude: a price-induced invention or change in technology which brings a substitution of another type. Instead of changing from a ratio of 5A:3B to 5A:2B, the new combination may be 5A:6C, or even 5A:1B:4C.

5. Deciding When to Produce: The Problem of Future Output.—In addition to the foregoing problems of deciding what to produce, how to produce, and how much to produce, the enterpriser also must consider the problem of when to produce. Sometimes the decision can be made separately from the others; for example, the cotton farmer must decide only what is the best time for planting in order to reap his autumn harvest. Mostly the decision regarding timing is linked with other decisions such as that regarding how to produce, or how much, and obviously it is related to the choice of products, which may differ in the minimum length of time required to produce them. Probably the shortest time between inputs and outputs is found in certain lines of merchandising where the turnover is very rapid, and the longest time in cases where large investment in capital equipment is necessary, as in the mining of low-grade nonferrous metals or the development of hydroelectric projects. These latter illustrations also point to the relationship between choice of what, when, and how much, since they belong to that group of enterprises which ordinarily must be conducted on a large scale if they are to be profitable and, because of their huge capital investment, require long-range planning.

If, for purposes of analysis, the time factor be isolated and considered by itself, certain important generalizations may be made. In the first place, the decision to produce outputs for sale at some time in the future requires a corresponding decision to make inputs prior to that time. These inputs of materials, labor, and land use must be bought and paid for by the enterprise, except where contributed by

its owners. This requires the expenditure of money and is commonly referred to as an investment, although this word has several other uses. If the output date is known, the investment in inputs must be made at some earlier time, determined by technical factors associated with the chosen method of production. Animal and vegetable products have a rather definite life cycle often connected with the seasons, so that there is little room for choice of input timing once the output date is decided. In certain manufacturing processes, however, and in merchandising, the enterpriser may vary the length of his production period considerably. He may speed up the manufacturing process by hiring more labor, working longer hours, or getting speedier machinery, and he will do so if he thinks it worth his while. The remarkable production records of American industry during the war indicate what can be done to shorten the production process when it seems desirable to bring the date of output closer.

In normal times, the decision regarding the timing of initial and subsequent inputs may be impulsive, rule-of-thumb, determined by availability of funds, or calculated so as to maximize profit. Established firms often have a long record of past operations to guide their judgments about the future, and the technical and trade journals offer the inexperienced manager much information of value in appraising alternatives. Many production processes become so standardized that there is only one "right" way to do them. There remain, nevertheless, a considerable number of production problems that cannot be reduced to routine. All new products present problems of timing in addition to those of factor combination, and new market situations are continually developing for old products. Small producers, as has been frequently pointed out in these pages, rarely calculate their actions as carefully as they should. Large producers are more likely to analyze cost and probable revenue with care, particularly when large investments are in prospect. In either case some element of guesswork is involved.

Interest cost often is ignored in small-scale enterprises and in firms where the funds are contributed largely by the owners. The interest cost for the time interval between input and output should be considered in rational calculations. An item of major uncertainty is the future price of the future output, or if the price is to be fixed by the seller, the future volume of sales at that price. At times the quantitative relationship between inputs and outputs is also conjectural, especially in farming, or with new enterprises, new methods, or new products. These elements of risk are sometimes allowed

for by assuming that the lowest of several possible prices or volumes of sale will be realized, or the highest probable expense.

The economist attempts to generalize the essence of the calculating process by assuming that the production planner decides upon certain reasonable or probable figures for inputs and outputs and then proceeds to calculate which of several possible constellations of cost and revenue would be the most profitable. This is done by much the same method as was used above for analyzing which factors it is best to use to get desired outputs. Each input is dated and considered as a separate factor, such as one day's labor on June 23rd, or one day on June 24th. Outputs similarly are dated in order to keep them distinct. The input cost is subtracted from the output revenue on each of the planned future dates to determine the net revenue for that day, whether positive or negative. The time element is then further introduced by discounting these amounts to determine the present value of each of these net revenues on the planning date. An algebraic total is then computed of the positive and negative remainders or "surpluses." The aim of the business manager will be to select that time distribution of future inputs and future outputs which will give the largest possible total present value, thus choosing the production plan which will maximize profit in the future. Earlier analysis indicates that this maximum will be reached when (1) the marginal rate of substitution between outputs of any two dates equals the ratio of their discounted prices, (2) the same rule holds for inputs, and (3) the discounted marginal revenue of any output equals the discounted marginal cost of any factor inputs with which it may be technically associated in the production process.

It is obvious that calculations such as these would be too laborious for the average enterprise to conduct in complete detail, and also that the figures involved are too conjectural to warrant treating them with such exactitude. Near-future planning data have a lower margin of probable error, but such planning hardly requires the elaborate discounting technique suggested above. Furthermore, long-range planning is perhaps more closely connected with preparing or obtaining the capital equipment required for production than estimating the precise output of that equipment on specific dates. This is particularly true under monopolistic conditions when the possession of capacity to produce is not equivalent to production at capacity. Nevertheless, the foregoing theoretical analysis helps to point out the many details which businessmen should or could take into consideration in deciding when to make expenditures to get desired future output. The process of discounting future costs and revenues

also indicates the often overlooked importance of the interest rate. Whether the funds to be invested are borrowed or owned, their use in preparing for future production involves a cost, either explicit cost or opportunity cost.

6. Altering Plans for Future Production.—The next problem related to plans for future production is the changes in such plans that arise from the changes in the conditions which gave them birth. Because of technical reasons and long-term contracts, changes often cannot be made quickly; and when they are made, they may have long-lasting effects. Therefore, managers must look far ahead to estimate future obligations, future inputs, and future outputs. These estimates are all subjective: entrepreneurial ideas regarding future price or sales volume, ideas concerning physical efficiency, ideas about future cost of factor units, etc. The personal ideational basis of judgments in this field is so obvious as to render hardly necessary the oft-repeated caution that demand and supply activities in the market are the result of guesswork, impulse, and imitation more often than they are precise, mathematical computations based on objective data. There even has been developed a "theory of expectations" which tries to explain, or at least to classify, human judgments regarding future economic events. For instance, if the price of a given product has risen during the past week or month, that may be taken by some to mean that this price rise will, in the future, continue at the new level, at the same rate, or in some other way. Others may interpret recent experience as forecasting a decline, before long, to the prior level or even below.

Whatever the specific expectation, it is important to realize that people do judge the future by the present (or rather by the recent past). Present price changes, therefore, alter not only current production rates as explained in previous sections of this chapter, but also alter plans regarding future production. They raise or lower the expected net revenue of future dates and therefore may change the optimum production plan as measured by the maximum present value of expected future surpluses. Changes in expectations of physical productivity do likewise, as when reports are received of the success or failure of similar processes adopted by others or of new inventions. The oil-drilling industry with its dry holes and its gushers offers abundant illustrations in this field. Finally, there must be mentioned the prospect of future interest rate changes which might make it necessary for planners to discount the distant future at higher or lower rates than the immediate future.

The degree of uncertainty about future conditions also varies. When a manufacturer gets a contract for the sale of his product in larger quantities for the coming year than in the past, he will be more certain about his future sales than if he is producing to stock for uncertain future sales. Yet even contracts may be cancelled, as those holding wartime munitions contracts found out in 1945. On the buying side the uncertainty of rental costs may be removed by purchase of the property or, with slightly less confidence, by a long-term lease. Similarly, short-term borrowing by a series of notes, subject to interest rate changes at each maturity date, may be supplanted by refunding into long-term fixed interest obligations. Uncertainty on the buying side suggests that calculations should include the highest, or nearly the highest, conjectural cost. On the selling side, the lowest or nearly the lowest price or sales volume must generally be chosen in planning present inputs for future outputs. Allowance should be made also for possible delays or inefficiencies in the physical production process. The more conservative the production planner, the higher will he estimate possible costs and the lower will be his predictions about sales or efficiency. The ultimate decision in any particular case often depends as much upon the temperament of the person or persons making estimates and final judgments as it does upon verifiable facts.

7. Equilibrium of the Firm, a Summary and Evaluation.—In trying to increase profits (or prevent their being reduced) a firm may adopt any one or more of the alternatives which have been described. It may spend money on selling efforts or in improving its products. It may reduce its selling price and thus curtail its revenue until the hoped-for expansion of sales occurs. The production plan may be altered by present outlays for machinery, for managerial skill, or for a more productive group of workmen which may ultimately reduce cost by more than the proximate additional expense of getting them. Buying or renting a more favorable site might also be included in this group and represents a frequent managerial problem in retail trade. Still another alternative is spending money lobbying for favorable legislation or for laws curtailing competitors. For instance, independent stores have secured punitive taxes against chain stores in some states, and many industries have hired "legislative agents" to help them get tariff protection or subsidies. Regulatory commissions may be similarly courted or coerced.

The analytical problem for price and production economics is to represent interrelated variables by curves or simultaneous equations and show how geometry or calculus reveals the desired maximum or

minimum points. The conventional curves in two dimensions, whose intersection solves the problem, are now giving way to three-dimensional surfaces or to contour maps of substitution functions with tangency lines. None of these refinements should obscure two very important facts about abstract analysis. First, its schedules and curves are hypothetical and are intended to reveal principles, not quantitative answers to the businessman's problems. Second, the variables are so numerous in nearly every management situation that they could be expressed diagrammatically only by multidimensional surfaces which we cannot draw. Therefore, recourse must be had to the equations which link the variables, and in these equations the coefficients must be represented by letter symbols since their quantitative magnitude depends upon the unknown peculiarities of each possible situation.

The practical problem for businessmen is to choose the right combination of expenditures, but since there are so many possibilities, and results are so unpredictable, his judgment must often be based on "hunch" rather than on careful calculation. No formal solution can be given which is more helpful to the businessman or government official than the generalizations which could be made without having recourse to mathematical presentation. An exception might be made for certain cost functions where the relationship between variables is governed by physical laws and may be expected to repeat the experience of the past in the future. Even in this case knowledge of average or marginal cost of production at different outputs does not give the answer to a host of questions about price policy, selling expense, and future production planning.

Most actions of businessmen supplying goods or services for sale are designed to improve their economic position according to their information, reasoning, desire, or impulse. Economists speak of these activities in the direction of increasing profit as being supply changes promoting equilibrium of the firm. They sometimes imply that the ideas of businessmen about demand and cost schedules reflect actual objective facts and that, since these "data" are given, the correct action to maximize profit can be accurately determined by abstract analysis.

There can be little objection to treating the "equilibrium" of the firm and the "maximum profit position" of the firm as synonymous, but one must not be led by the use of such terms to think that every action of businessmen leads in the direction of equilibrium. Mistakes may be made because of inadequate information and poor judgment. Nor must we think that the economic conditions of demand and cost, which are chiefly responsible for concepts of demand and cost

schedules, remain constant for any appreciable length of time. They are changing continually. Even if subjective reflections of objective circumstances were completely accurate, human logic above reproach, and the profit motive supreme, any equilibrium of a given firm could last but a moment. Finally, the equilibrium formulae of the preceding pages are approximated only in part by the most careful business enterprisers and practically never with the use of academic nomenclature. For instance, most managers have no idea about their marginal costs, much less a point of equality with marginal revenue. As indicated in the earlier sections of this chapter and in frequent reference throughout the portions dealing with theoretical analysis, most business decisions on the supply side of the market (and the demand side, too) are based on a great variety and mixture of motives. Short-run or long-run profit maximization is only one motive, though usually the most important one.

Nevertheless, the concept of equilibrium of the firm looms large in recent economic theory and its basic principles should be understood by those who wish to keep abreast of developments in economic thought. It is also useful in understanding other economic concepts such as "normal price" and "normal profit," together with the issues of public policy which are related to these ideas in the framework of government regulation of competition and monopoly in a capitalistic system.

8. Normal Profit: Equilibrium of the Industry.—Corresponding to the expansion and contraction of output of the firm in its efforts to reach the equilibrium position of maximum profit is the entry or exit of firms engaged in the production of a given commodity. If it appears to some one that he can make a good profit in producing that commodity, he will be tempted to invest his idle capital in the materials, equipment, and labor required to do so, or he may divert his funds from other employments which he thinks less profitable. This new entrant will try to sell his output at the prevailing price, or slightly below it, or he may try to market a product of superior design and durability. If he is successful in selling his article, some or all of his customers will be people who formerly purchased from the firms originally in the field and these firms, therefore, will have to be content with a smaller volume of sales unless they are willing to reduce their asking price. In either case profit will be reduced. For instance, if a grocery store in a certain location seems to be making good profits, some one else is likely to open a store near-by, taking away some of the customers of the other store and perhaps stimulat-

ing price reductions.[1] The entry of firms into the automobile producing industry or into airplane manufacturing during recent decades furnishes another familiar example. New producers in a given field include, of course, either entirely new firms or old firms which add the new product to former lines or shift from one good to another.

On the other hand, firms may cease producing a certain good if it becomes less profitable than other goods they turn out, or if the firms themselves are driven into bankruptcy by falling demand. If a bankrupt firm is dissolved through the sale of its assets to others who do not intend to employ them in the same line of production, the supply of the product reaching the market will decline. The exit of producers in either of these two ways is generally ascribed to the presence of subnormal profits or of losses. In short, entry generally occurs when profits of industry are above "normal" and exit when profits are below "normal" (though see Section 8 in Chapter 9). An "industry" is in equilibrium with no expansion or contraction of supply either when each individual firm is making normal profit or, though profit rates of different firms vary, the average is "normal."

9. The Opportunity Cost Theory of Normal Profit.—The concept of normal profit has been given various definitions, none fully satisfactory. For the individual firm it may be thought of as that return to the owners of business enterprise which equals what they could get by hiring out their abilities, capital, or land to some one else instead of using these assets themselves. This is known as the "opportunity cost" theory of normal profits. The logic of the argument is that if business enterprise offers the opportunity to earn more residual profits than can be earned by contracting the sale of one's labor, capital, or land to others, people will be tempted to shift to self-employment of time and funds. Unfortunately, the argument does not hold so well when business earns less than "normal," for in many cases the capital equipment of the firm is so specialized that it cannot be sold to some one else except at a heavy loss. Hence, there is no opportunity to earn "the going rate" on the funds originally invested, but only on the shrunken value of stock shares or salvaged equipment. The best that can be done, when revenues still exceed operating expenses, is to accumulate the excess as a partial depreciation reserve and then invest it in a different line of production

[1] In some lines excessive entry leads to a general condition of operating at less than capacity. This keeps *ATC* above its possible minimum, and entry may lead to higher prices, not lower.

instead of replacing equipment that wears out. The opportunity cost approach, therefore, is more useful in explaining entry than in explaining exit.

For the group of firms comprising an industry, normal profit is often taken to mean the average rate of profit being reaped by other industries at the same time. This is also a variation on the opportunity cost theme. The argument is that new entrants into the field of enterprise, or those shifting from other lines of business, will tend to go chiefly to those fields where firms on the average are making higher profits than elsewhere. Low profits appear in industries which are experiencing a declining demand for their product or where in the past there has been excessive entry. "Cutthroat competition" is an adverse factor, while monopolistic agreements or practices are favorable to profits above normal. High profits are rarely due to excessive exit, but rather to insufficient entry for one reason or another.

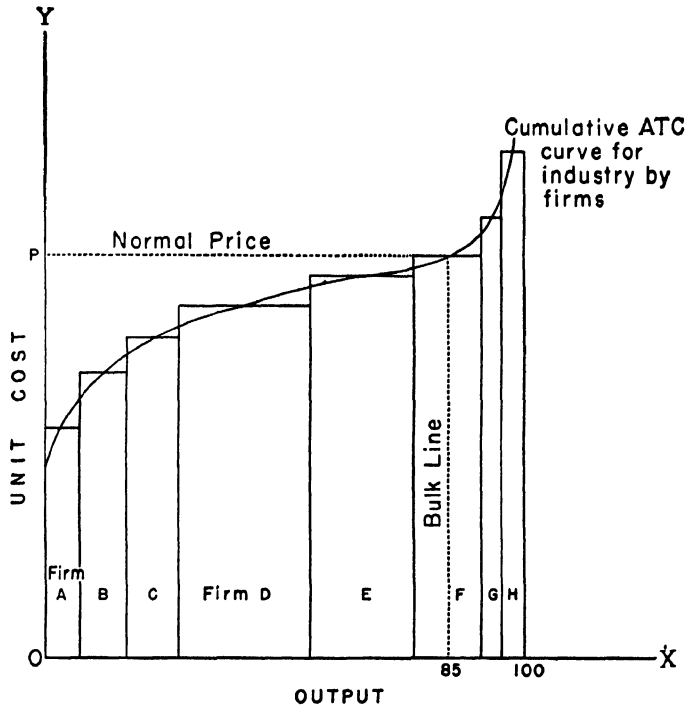
10. The Bulk Line Cost Theory of the Normal Profit Price.— Industries usually are made up of many firms earning profits individually at different rates, some high, some low. It is therefore somewhat difficult to describe a condition of "normal" profits for an industry as a whole. Even in prosperous times, when the majority of firms in an industry operate in the black, there are usually a few firms which are in the red. Several attempts have been made to define this state of an industry with precision, notably by the price-fixing agencies in the United States in 1917–1918 which advanced the concept of "bulk line cost" to divide the "bulk" of the producers who are making money from the few who are not. This line is said to be the price which is equal to or greater than the cost of production of about 85 per cent of the output of the industry. Fifteen per cent have a cost greater than this cost-price. There are many difficulties with this concept, but it is a useful one for the price-fixer. One difficulty is the accounting problem of determining what items should be included in cost. Another is trying to decide whether an 85:15 ratio should be applied to all industries at all stages of the business cycle and, if not, what alternative ratio is appropriate. A third is attempting to define the price when firms produce different yet similar products, like manufacturers of furniture or clothing.

The bulk-line-cost price approach to normal profits gives an appearance of greater precision than a simple reference to average profits, but it is far from being a good tool for analysis of conditions of entry and exit. In the first place it assumes that entry is stimulated by the observation that all, or nearly all, of the firms in an industry are mak-

ing a profit as opposed to a customary 85 per cent. A full theory of entry would have to describe many other motives and would have to admit that the 85 per cent concept is more influential among economists and price-fixers than among businessmen, most of whom have

FIGURE 44

BULK LINE COST AND NORMAL PRICE



never heard of it. In the second place, it ignores the cyclical fluctuations in the general level of profits which would make 85 per cent too low a figure for good times and too high for bad times. Thirdly, as pointed out in Section 8 of Chapter 9, entry is probably stimulated more often by abnormally high prices than by abnormally high profits, although the two usually go hand in hand. A careful entrepreneur is likely to compare probable selling prices in relation to his probable costs before making his decision to venture. On the other hand, an impulsive entrepreneur is more likely to look at the large profits of a few successful firms than to compute an average for the industry and compare it with the average profit in other industries. Finally, industries differ greatly as to (1) relative average profits in

good times (gold-mining versus copper), (2) relative average profits in bad times (public utilities versus steel), (3) relative ease of entry (retail merchandising versus shipbuilding), and (4) relative ease of exit without loss of capital (merchandising versus smelting).

11. Other Uses of the Concept of Normal Profit.—Because of these difficulties with the concept of normal profit, it would have been discarded long ago were it not for the usefulness of the idea in two connections. First, the competitive system of free enterprise is supposed to find one of its chief justifications in the alleged fact that the profit motive leads individuals to make rather quick adjustments of supply to demand, thus reducing excessive profits and transferring resources to their most urgent uses. Whenever the above-normal profits of a group of producers reveal that "demand exceeds supply" in that field, new entrants are supposed to be attracted until expanding supply forces profits back to normal by bringing prices down. This prevents a few from exploiting the many when they find themselves in a particularly fortunate situation. Competition is also supposed to force prices down when cost-reducing innovations bring large profits to those who first introduce them. However, sellers may choose to be content with a smaller fraction of the total market rather than to cut their prices when new competition appears. This is particularly common where the number of competing sellers is small (steel) or where they produce products which are slightly different though similar (harvesting machines). The argument that competition prevents the long continuance of greater than normal profits also runs afoul the facts which seem to reveal that obstacles to entry are so potent in some fields as to preserve high profits for many years running (aluminum, shoe machinery).

The second reason for keeping the concept of normal profit is that economists find it useful in equilibrium analysis and in that other abstraction, the "long run." Normal profit is considered essential to the long-run survival of each firm, to the stability of each industry and, therefore, to the stability of the economy as a whole. Abnormal profit is held to be characteristic of boom and depression stages of the business cycle and is one explanation of why these stages cannot be permanent. The concept of normal profit permits the development of many generalizations about price, the firm, the industry, and the economy which fit neatly into a coherent whole which is logically unassailable even if some of its premises are not entirely realistic. As emphasized in Chapter 1, the principles of economics are not intended to be absolute rules applicable to explain every case

Rather they are generalizations which are true most of the time when specified conditions are present. The relation between actual pricing methods and conventional economic theory will receive further elaboration in the sections which follow.

12. Normal Price Causes Normal Profits and Holds Supply Constant.—The concept of "normal price" is virtually synonymous with "long-run equilibrium price," since both depend upon the ideas of "normal profit" and "equilibrium for the industry." If the entire product of an industry is being sold at a given price and if at the same time there is no tendency for the output to expand or contract by action of existing firms or because of an unbalance between the rate of entry and exit, then this price is said to be the "normal price" for that product. An alternative definition would stipulate that it is the price which would maintain a constant rate of supply if the following things were to remain constant: (1) schedule demand, (2) prices of other products, (3) factor prices, and (4) technology. Although perhaps not in existence at any given time, it is the price which would be reached in the long run if enough time were allowed for adjustment of output by individual firms and for bringing exit and entry into offsetting balance.

Certain difficulties with this concept deserve attention. First, normal price thus defined cannot exist in actual business practice except accidentally and momentarily. Conditions change too rapidly; for instance, there would be a different "normal price" with each level of demand intensity and demand could certainly be expected to change several times within the period conceptually required for the adjustment of output by firms and for the restoration of a balance between the entry and exit of firms. Second, there can be no normal price for the products of an industry selling articles which differ from one another in some aspect of design, packaging, or advertised trade-mark (automobiles, radios, pencils). Since product differentiation is very widespread in modern times, the remaining field of homogeneous products where the idea of normal price might have applicability is quite small. Moreover, some of the few cases that remain reveal monopolistic practices which establish a practical stability of supply at the given level of demand, but the profit position of firms in the industry does not resemble the bulk-line-cost price picture described above.

Finally, it is very hard to define an "industry." One hardly knows where to begin and where to stop, since so many farmers, manufacturers, and merchandisers sell more than one product. For in-

stance, it would be hard to state just how much of the profit of General Motors should be considered in appraising the normal profit position of the automobile industry and how much should go to refrigerators, airplanes, diesel engines, batteries, and a host of other things including trucks. And should only automobile assemblers be included in the automobile industry? Or should all the parts companies be appraised, too? Many products compete for the consumer's dollar with objectively dissimilar goods. Thus it is a question of whether we should speak of "the canning industry" or one part of it such as "the vegetable canning industry," "the food preservation industry" (to include dried and frozen foods), or "the food industry as a whole." And what would be the limits of the latter? To turn to another field, is it proper to analyze "the radio industry," apart from the cinema, the legitimate stage, or all other forms of entertainment? If a broader concept is used, where would books and public lectures be included?

In the light of these and similar difficulties with the definition of a "commodity" or "industry" it seems best not to analyze the terms too closely, but instead to examine what the popular concepts can add to our knowledge of the working of the economic system. For formal analysis the terms probably should be abandoned entirely in favor of the actual maximizing or minimizing unit which is either (1) the firm, or (2) the private individual consumer, or (3) the factor source. This will get us by the logicians even though it still leaves us in trouble with the psychologists.

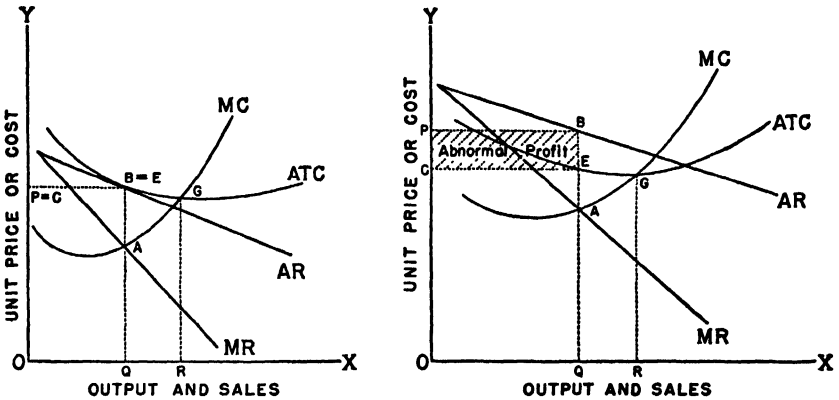
13. Normal Price When Normal Profits Are Included in Cost. —In an effort to preserve the concept of normality for price analysis under monopolistic conditions, a shift is often made to the opportunity-cost version of normal profits. Under this approach, costs as figured by the accountant are expanded to include a "normal profit" return to the owner for his otherwise uncompensated services as working manager or supplier of capital or land. The individual firm is then taken as the prototype of the industry. Its equilibrium price is equal to its *ATC* when these imputed "costs" are included. For the firm selling under monopolistic conditions, this means that when $MC = MR$, *ATC* must also equal *AR* (the price). This "normal profit" situation is shown by the following diagram, (A), Figure 45. Another diagram, (B), is given to show by contrast the more common situation under monopolistic conditions where difficulties of entry permit price and sales volume to be such that a profit is made in excess of normal.

The normal profits which are said to exist under the long-run equilibrium situation shown in Figure 45A are, by definition, equal to opportunity costs as explained in Section 8 of this chapter. By a

FIGURE 45

PROFITS UNDER MONOPOLISTIC CONDITIONS

(A) Normal Profit (B) Profit Above Normal



further inference these profits are considered to be the same as the average earnings of firms engaged in "competitive industry," a concept which will be analyzed in the next chapter.

Both diagrams of the monopoly revenue type shown in Figure 45 reveal that the minimum *ATC* point *G* would require an output *OR* greater than *OQ* which maximizes profit. This means that it is profitable for the monopolist to keep his selling price so high that sales and output are below that which would yield the lowest total unit cost.[2] This conclusion is confirmed by the common observation that, for a given stage of the business cycle, excess capacity is most extensive in those industries where there is the least price competition (cf. cement).

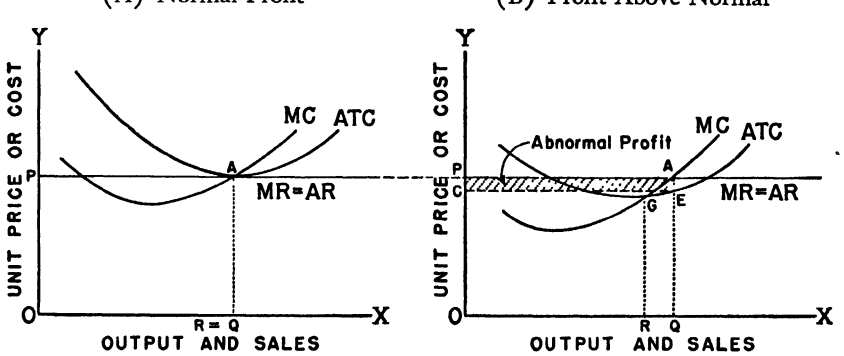
A competitive industry of the traditional type may be in equilibrium even though some firms make more profits than others. This is the bulk-line-cost-price idea of Section 10 of this chapter. In diagrams, normal profit must be included in *ATC* as a cost ("opportunity cost"). This is done in Figure 46 where two firms with dif-

[2] Exception: when demand is so great that *MR* intersects *MC* above *ATC* or at the place where they cross. This is very rare and if it should occur for any length of time would lead to an expansion of capacity. See Chapter 13, Section 7.

ferent costs are shown selling at the same price. The first firm is in equilibrium where $MC = MR$, even though the selling price, AR, equals total unit cost, ATC, because the latter contains normal profit. The second firm, B, makes normal profits plus the abnormal profit $PCEA$. The most profitable output OQ is *greater* than OR, which would have given the lowest possible ATC. Under monopolistic conditions OQ is *less* than OR, as shown in Figure 45.

FIGURE 46

PROFITS UNDER PURE COMPETITION



14. Summary of Economic Principles Dealing with Production Policies Which Influence Prices.—Since the concepts of normal profit and normal price involve the volume of production, they may be included in the summary of this chapter as determinants of the quantitative aspects of supply schedules of firms or industries. The other topics deal more with the buying policies of firms in relation to output than they do with selling and will be treated at greater length in the chapters on distribution, where the purchase of factors is under direct consideration. The following outline covers both viewpoints:

I. Total supply offered for sale is a function of:
   A. The output of individual firms, which is a function of:
      1. Managers' ideas about profit opportunities, i.e., about probable demand schedules in relation to cost schedules
      2. Availability of the necessary factors of production, especially capital, materials, and skills
      3. Presence and strength of monopolistic agreements
   B. Willingness to expand or contract the inventories of the individual firms

C. The number of producing firms, which is a function of:
   1. Past entry and exit, which is a function of:
      (a) Profit relative to other opportunities (normal profit)
      (b) Losses relative to financial resources and determination to carry on
      (c) Monopolistic restrictions upon freedom of entry
      (d) Technological or capital restrictions upon entry
      (e) Institutional stimuli to entry (subsidies, etc.)

II. When there are joint products, changes in the relative quantities produced and sold are a function of:
   A. Changes in the relative prices of the products
   B. Changes in relative rates of sale
   C. Willingness to allow changes in inventory
   D. Changes in the prices of factors, if use-ratios differ
   E. Changes in technology

III. When there are joint factors, changes in the relative quantities bought and used are a function of:
   A. Changes in technology which alter their relative efficiencies as substitutes or complements
   B. Changes in the relative prices of the factors
   C. Changes in expected price or rate of sale of single or joint products

IV. The time at which outputs are to be expected is a function of:
   A. The time of inputs
   B. Technological processes used
   C. Expectations regarding relative future prices of products and factors
   D. Interest or discount rates, and expectations of change in them

This summary goes beyond that of Chapter 9 both by adding more independent variables and by examining certain causes of the magnitude or change in these variables. Everything that from the viewpoint of this chapter is an independent variable is itself dependent on one or more other variables, only a few of which can be mentioned in passing. The list of these determinants is not intended to be complete, either in this chapter or in any other.

Chapter 11

A FUNCTIONAL APPROACH TO TYPES OF COMPETITION AND MONOPOLY

1. Difficulties with Substantive Classifications.—Much recent analysis in the field of monopoly and competition fails to distinguish clearly between substantive and functional connotations of terms.[1] For instance, a writer may start with a substantive definition of a monopolist as the sole seller of a certain commodity and then speak of the way in which that commodity competes with substitutes. The "monopolist" may reduce his price to strengthen his market position relative to these substitutes (cf. the price reductions in aluminum in recent years). On the other hand, the "pure competitor" of classic definition may never consciously reduce his asking price to sell more goods, but always sell "at the market." And the "oligopolist" may either engage in competitive price wars with his few rivals or collude with them in a monopolistic fashion to eliminate one or more forms of rivalry. Substantive and functional definitions are too often mixed. For logical consistency and clarity of thought they should be unscrambled wherever possible.

One solution to the problem is that of expanding the list of substantive determinants until so much is said about the product, the sellers, the buyers, etc., that alternative courses of action are narrowed down to a single one. This is the process of reasoning from market structure to market behavior. It leads to very extensive classifications of market situations by a series of combinations of possible opposites. One such list has been published containing 64 substantive classes and there is no good reason except lack of space why such a list might not be extended to 128, 256, or even more.[2] In terms of behavior patterns the list of only 64 situations is obviously redundant and hence was shortened by Bain to 14 by eliminat-

[1] A substantive definition is one which identifies by describing states of being. A functional definition describes acts of doing. The former is supposed to be something inherent; the latter is outward manifestation in behavior. The substantive does not change; the functional does.

[2] One class, for instance, is that in which there are the following substantive conditions: few sellers of a durable producer's good whose product variation is unimportant, yet it is differentiated in important degree and there are many buyers. See Joe S. Bain, "Market Classification in Modern Price Theory," *Quarterly Journal of Economics*, Vol. 56, August, 1942, p. 560.

ing duplicate and unimportant cases. Each of the 14 remaining situations is apparently intended to give rise to a distinct (and complex) pattern of market behavior.

A second solution is to begin with classes of market behavior and work backward as needed to the market structures from which they spring. This would classify different types of activity by buyers and sellers and reveal for each class one or more situations which are its logical causes. Or a list of combinations of forms of behavior might be developed comparable in length to the Bain-Mason list of combinations of substantive elements. Such a list might also be redundant in terms of the situations which supported the complex patterns of behavior, but this would not be sufficient reason for elimination and condensation. It is behavior which is our chief interest, behavior and its results. Hence the list of behavior patterns need be made only as long as necessary to include the cases which interest us. Or if the statistical measurement of significance is preferred, only those cases which occur with the greatest frequency need to be tabulated.

The writer has no desire to argue that the substantive-to-functional approach should be discarded, but only that the functional-to-substantive approach seems preferable in many expository summaries, and even perhaps as a background for empirical investigations.

Industry studies which give information regarding the relationship between price behavior and output variations on the one hand and seller-buyer-product-institutional circumstances on the other may be summarized with functional emphasis and functional classification just as easily as with substantive. It is only a question of which is preferable for the purpose in mind and which gives rise to the most useful vocabulary.

2. The Questions Asked in the Functional Approach.—The answers of economic theory will obviously differ with the questions asked. Hence it seems desirable to make the questions as explicit as possible. The ones which seem most appropriate in a functional approach to competition and monopoly are the following:

1. What are the different methods of price determination in the market?
2. What are the different forms of buyer behavior and seller behavior?
3. What are the causes of these forms of behavior?
4. What are the major patterns of "price behavior"?
5. What are the causes of these patterns?
6. How should these methods and forms of behavior be appraised in the light of social welfare?

Only the first two questions will be discussed in this chapter. The last four will be treated in Chapter 12. Many of the answers to these questions must be given in outline form to save time and space. For the same reason, many obvious inferences and explanations may also be omitted.

3. Methods of Price Determination.—The explanation of the actual market prices which exist at any given time is a short-run or market-time problem which must not be confused with the problem of price equilibrium in the long run. The first problem treats prices as effects of buyer and seller activities. The second deals with buyer and seller responses to price. In the traditional terminology the distinction is between schedule demand and supply as proximate causes of price on the one hand and price as one cause of market demand and supply adjustments on the other. The market behavior of individuals is our concern, not the elasticities of demand and supply. A brief outline of methods of price determination may be grouped under three headings as follows (the most common substantive equivalents are given in brackets to the left):

I. Prices set by individuals acting independently

A. By only one individual

1. Classified on basis of who sets the price

[*Monopoly*] (a) Price set by seller, accepted by buyer (because he cannot buy cheaper or on better terms)

[*Monopsony*] (b) Price set by buyer, accepted by seller (because he cannot sell higher or on better terms)

2. Classified on basis of motives of seller (or buyer)

[*Monopolistic competition*] (a) Considers reactions of those on other side of market (seller considers price elasticity of demand)

[*Oligopoly* or *oligopsony*] (b) Considers reactions of those on his side of market (seller considers cross-elasticity of demand and its effects on actions of rivals)

B. By two individuals

[*Bilateral monopoly*] 1. By both buyer and seller. If they have unavowed overlapping range, bargaining is likely. Subsequent agreement on price and acceptance by both.

C. By many individuals

[*Auction*] 1. By highest bidder among buyers when only one seller (or lowest bidder among sellers when one buyer)

[*Pure competition*] 2. By many sellers and by many buyers in some place where they or their representatives meet and learn of rival bids and offers. Totals on both sides at various

prices are compared until one price is found at which the totals are equal.
Subsequent acceptance by those buyers willing to pay it or more and by those sellers willing to take it or less.
Rejected by other buyers and sellers.

II. Prices set by individuals acting together
  A. Identity of viewpoint and interest of members of a group (borderline case, shades into the next)
  B. Price leadership of dominant member
[*Collusive monopoly*]
  C. Definite agreement among members on price to be asked or offered by each.
    1. Informal agreement, usually verbal
    2. Formal agreement, usually written

III. Prices set by the government or other external authority and accepted by both buyer and seller
  A. Specific price
  B. Fixed maximum or minimum price
  C. Formula for calculating price

Since nearly everyone operates on both the buying and selling sides of the market, the methods outlined permit many possible combinations. For instance, the first two methods under I–A give rise to four possible situations:

1. Firm sets price as seller, sets it as buyer.
2. Firm sets price as seller, accepts it as buyer.
3. Firm sets price as buyer, accepts it as seller.
4. Firm accepts price as buyer, accepts it as seller.

Furthermore, since most firms buy more than one thing and sell more than one product, there may be a great variety of buying and selling methods practiced by each firm. Smith may bargain with his organized workers, set his hiring price for other workers, purchase some materials from a list-price seller, purchase others at a price fixed by the government, etc. On the selling side there is likely to be more uniformity of practice if only because most firms (and individuals) sell a smaller variety of things than they buy. The large number of different types of buying and selling activity which may characterize a single firm constitute one good reason for avoiding the blanket appellations of "monopolist," "oligopolist," etc. An individual is what he does. He becomes something different when his actions change.

4. Ways of Competing Against Others.—Another very useful classification from the functional viewpoint is that which includes

ways in which sellers compete against other sellers and buyers against other buyers. This also includes methods of price determination, but views them from another angle, and it adds other methods of market activity not suggested in the first outline.

Firms compete when they contend against one another for sales to customers, for purchases from suppliers, or for business survival or growth. Competition as a form of action means struggle. Usually it is struggle with known rivals, such as other people buying or selling the same good. But sometimes the contest is with unidentified opponents, such as "sellers in general." [3] The "enemy" may also be governmental, the forces of nature, or one's own complacence. Minimum success in competition is a requisite to survival. Greater success permits growth in net worth or net income, or is evidenced by that growth. That which is often called monopolistic activity also influences survival or growth, but a discussion of its forms and implications must be deferred to a later section.

The major forms of competitive activity may be grouped in the following outline:

I. Direct methods of competition
  A. Marketing competition (against other sellers)
    1. Supply-price competition (usually against known rivals)
    2. Selling competition (against known or unknown rivals)
    3. Product competition (against known or unknown rivals)
  B. Purchasing competition (against other buyers)
    1. Demand-price competition (usually against known rivals)
    2. Buying competition (against known or unknown rivals)
    3. Specification competition (against known or unknown rivals)
  C. Predatory competition (against other buyers or sellers)
II. Indirect methods of competition
  A. Cost competition
  B. Production planning

It should be obvious that many of the forms of competition in this outline are forms of behavior often found in situations substantively called monopolistic or monopsonistic. This is clearly true of all the direct methods. The indirect methods are used by buyers

[3] Nearly all substantive "monopolists" compete in this sense even though they refrain from price competition with known rivals.

and sellers in both "monopolistic" and "competitive" situations. These arguments will be clarified by the sections which follow.

5. Direct Methods of Competition.—The direct methods of competition involve action in which the seller or buyer tries to help himself at the expense of others in a similar market position, or even directly to harm them. Marketing competition includes those forms of behavior which develop when the supply that a seller wants to sell is greater than the demand at the going price and terms of sale. The situation is sometimes called "a buyers' market," but it is really broader than that since a single seller may act competitively even when most of the other firms feel that there is no necessity to do so. Marketing competition is reduced in form and intensity by monopolistic activity and by the rare periods of "boom" conditions when most firms' existing capacity to produce is fully and profitably utilized.

The most widely recognized form of marketing competition is supply-price competition. Often called simply "price competition," this occurs when one firm tries to get additional sales by reducing its asking price(s). Several variations on this theme are possible. The objective may be to attract customers from known rivals or from unknown competitors.[4] Additional buying by present customers may be the goal. Or the price reductions may be defensive, not offensive. There is also the case where the price of X is reduced in hopes that customers (as in retail stores) may buy more of Y and Z whose prices remain unchanged.

Note should be taken of what supply-price competition is *not*. In the first place it is not the same as those price-raising techniques which are popularly called monopolistic, as will be explained in a later section. Price-cutting to increase sales does not lose its competitive character merely because some one expects this action to be followed later by a higher price. In the second place, price competition is very different from the market activity of persons selling under substantive conditions of "pure competition." No such seller has to reduce his asking price in order to sell more. He characteristically offers his output on the market for what it will bring. His choices involve only the amount offered and the time at which he makes his sale. Price declines are the result chiefly of reduced intensity of

[4] Sometimes a substantive distinction is made on this basis. Oligopolists are said to cut prices relative to *known* rivals whereas monopolists "set" their prices at the best point in their assumed demand schedules, whose elasticities are a function of the prices, qualities, and quantities of substitutes offered by *unknown* rivals. The difficulty of defending such a distinction is obvious.

demand. They may also occur because *many* sellers offer a larger amount for sale, or even because they diminish their asking prices. But prices under "pure competition" never fall merely because a *single* seller decides to reduce his asking price.

It is also desirable to recognize that any form of marketing competition may include activities which the business community or the injured firm may call "unfair." False claims may be made regarding the content of a product or the service it will render. Libelous statements may be made against rival firms or goods. Salesmen may coerce or bribe buyers in various ways incompatible with the mores of the group. In certain situations where there is an agreement or a tradition against supply-price competition, price-cutting itself may be called "unfair."

6. Purchasing Competition.—On the buying side the three corresponding methods of competition are less familiar because they occur only in those rare periods of capacity operation of the firms supplying a given industry. This has been characteristic of business in general only in time of war, but a similar "shortness of supply" occurs now and then in particular markets where demand becomes unexpectedly great or supply unexpectedly small. Natural calamities, strikes, shifts in fashion, and rapid changes in technology offer some illustrations.

Demand-price competition consists of raising the bid price to get scarce goods or services which would otherwise go to competitors. Occasionally the buying price will be fixed for one reason or another and the purchasing departments of firms will be able to get their desired supply of goods only by such methods of buying competition as offering prompt payment, absorbing freight, and agreeing to accept delayed shipment. Specification competition may occur parallel to product competition, as when a firm agrees to accept inferior goods in order to avoid getting nothing at all.

All of these forms of purchasing competition should be distinguished from those tactics used by buyers when supplies are not scarce. Under conditions of a "buyers' market," the purchasing department may bargain with suppliers to secure price reductions, may refuse to buy at all unless a fixed low price is met, may demand superior quality and service, etc. This type of activity is best associated with "monopsonistic" behavior and will be discussed below.

7. Predatory Competition.—In addition to the various forms of "unfair" competition that may occur in buying or selling, there are other things which firms may do to get the best of their rivals. These

include actions intended to weaken a competitor's ability or desire to continue in business by raising his costs of production or sale. Sabotage, acid, arson, and even mayhem have been used to inflict costly losses or to raise the price a competitor has to pay to get supplies of goods or services. Laws may be sought or applied which increase the rival's cost of doing business, such as tariffs and excises on imported goods, rigorous inspections at frequent intervals, etc. A firm may hire away the best executives of its competitors thus getting a double advantage (unless too high a salary has to be offered as inducement). The losing firm will have to use inferior men and the hiring firm will have a staff better able to introduce economies or increase sales. There are also various ways of stealing from rivals the trade secrets which have given them a special advantage. These include both direct theft, the employment of labor spies, and bribing employees to tell what they know. To injure a competitor by increasing his costs, a firm may stir up labor trouble at the rival's establishment. Or it may use its influence with those who supply funds or goods to cause them to charge the competitor more than they charge the favored buyer.

These forms of competition are somewhat different from those usually called "unfair" by trade associations or the Federal Trade Commission, although there is some overlapping. That is why they are grouped at this place under the heading "predatory." With the exception of tariffs, predatory competition is usually confined to situations in which sellers or buyers are few, because only in such cases can one individual see a clear gain to himself through loss to his rival. Under the "many" case of substantive pure competition in either buying or selling, spite or malice may give rise to such action, but it can hardly be called competitive in the sense which dominates this discussion.

8. Indirect Methods of Competition.—The indirect methods of competition are those which are not aimed at any particular rival or group of rivals, but which may often spell the difference between success or failure. Most important in this group are efforts to reduce cost by technological improvements. Practically all firms are continually on the quest for new methods which will reduce average unit costs. These include either better combinations of existing factors or the invention of new forms, both of machines and of materials, which will give the same output with less total cost for inputs.

Cost competition is a form of behavior so widespread that it is not limited to any one substantive type. It is found where sellers are sin-

gle, few, or many, where products are homogeneous or heterogeneous, durable or non-durable, etc. Successful cost competition may determine whether a firm is among the survivors when a decline occurs in schedule demand and is reflected in either falling prices, shrinking sales, or both. On the other hand, cost reductions may precede and make possible price reductions. If price cuts actually result, then cost competition may be considered one cause of price competition.

A second form of indirect competition is that of production planning. This involves timing the inputs and outputs of an enterprise so as to have them occur at those times which will bring the lowest costs and the largest revenues, i.e., the largest net profit. The best times for outputs depend on the business cycle in general and the product cycles in particular. For some products seasonal timing is important. For others the task is to secure the best adjustment of one's own efforts to those of rival producers or of potential buyers.

9. Forms of Monopolistic Behavior.—Monopolistic behavior is the opposite of competitive behavior in several ways. It includes action (1) to restrict certain forms of competition, (2) to restrain freedom of entry, and (3) to "exploit" buyers. When individuals compete, they make independent decisions which they expect will give them an advantage over their rivals. When they act monopolistically, they surrender some of this independence to a group or to a government agency. Competitive behavior is supposed to benefit consumers by giving them better goods at lower prices and better terms. Monopolistic behavior works in the opposite direction, particularly towards higher prices. The profits of competitive industry are proverbially low, while those of monopolies are high.

The anti-competitive and anti-buyer aspects of monopoly are interrelated. Sellers who do not "naturally" occupy a position where they can exploit buyers may seek such a vantage point. They do this by forming associations of rivals to reduce the struggle between them. They think to make greater profits by removing certain forms of competition which formerly benefited buyers. These associations may also be helpful in limiting the entry of new competitors.

The amount of freedom of action surrendered to the association usually varies with the tightness of the group. A trade association is less able to restrain competition among its members than is a holding company. Some labor unions exert more pressure on members to conform than do others. Mergers and amalgamations completely destroy the freedom of the combining units, but pools do not.

Associations may be voluntary or coerced. Firms may voluntarily agree to eliminate "cut-throat competition" among themselves. Or, paradoxically, a weak firm may be forced by such competition to sell out to a strong firm or to abide by its selling policies. Coercion may be applied in many ways, by denial of supply sources, by threats of a "price war," by cutting off bank credit, etc. Often that which appears to be "voluntary" association may contain elements of compulsion.

Governments may intervene to reduce competition as well as to enforce it. For many years the ICC has restrained rate-cutting by railroads. More recently we have had price-floors for agricultural products under the AAA. "Fair trade laws" also limit the degree of price-cutting that is legal. Sometimes industries seek to be treated like public utilities and thus be protected from the entry of competitors. In many states certain trades, like barbers, have obtained government support for guild pricing practices. They are given power to restrict entry and to set minimum prices which they then raise as rapidly as they can.

10. Anti-Competitive Behavior.—When firms get together in voluntary or compulsory association to reduce competition among themselves, they are seeking to reduce the number of competitors or their freedom of action. The ultimate objective in our capitalistic economy is usually, of course, to secure greater profit. However, an important intermediate objective of some business men is to reduce the number of occasions on which they must make major decisions of policy. To them decision making is painful. One major decision to cooperate may remove the need to make a whole series of other decisions which would be required if rivals were to act independently. Each cooperator surrenders part of his autonomy in order that others may be persuaded to surrender some of theirs.

There is set up a field of anti-competitive behavior which may include one or more of the following forms of cooperation among sellers. Price agreements are the most common.

FORMS OF ANTI-COMPETITIVE BEHAVIOR

I. Establishing uniform price policies
   A. Delivered price systems
   B. Open price systems
   C. Specific minimum prices, F.O.B.
II. Establishing uniform sales terms

III. Allocating markets
IV. Restricting quantity sold or produced
V. Standardizing quality

These terms are sufficiently descriptive to need no elaboration. Comments on uniform price policies were made in Chapter 9, Sections 11 and 12.

11. **Anti-Entry Behavior.**—Monopolistic behavior includes any action to increase the difficulty of entry of new competitors. Entry is opposed because new entrants will take away some of the sales of the older firms. Associations to prevent price-cutting and to reduce other forms of competition may be used to discourage entry. Their chief weapons are patent pools, control over supply sources, and other methods of making it difficult for a beginner to get started. Sometimes they do not act until a rival has begun operations. Then they engage in local price-cutting and other powerful forms of competition to squeeze out the newcomer. This is competitive behavior with monopolistic intent. Through control over government bureaus existing firms may prevent a potential entrant from getting the necessary licenses. Or they may cause certain regulations to be enforced against him which are overlooked in relation to existing firms. A newsboy may use his fists to keep another boy from muscling in on his street corner monopoly. Larger firms use more subtle pressures.

Entry may be difficult anyway. Scarce natural resources, large initial investment, rare personal skills, and limited franchises are obvious barriers. The mere fear of inability to break the hold which existing firms have on the market may deter possible entrants. Brand names with popular goodwill are hard to beat.

When obstacles to entry exist, whether natural or artificial, the protected firms often make supernormal profits. These serve as an inducement to entry. Hence monopolistic behavior may include efforts to conceal these profits, as by stock watering, excessive depreciation charges, etc. Supernormal profits depend upon the use of a selling price which is high relative to costs. At times when entry is especially feared, a temporary reduction in this price may serve as a deterrent. This problem will be discussed further in Section 13 below.

12. **Anti-Buyer Behavior in General.**—Monopolies have long been denounced for the ways in which they harm the consumer. The argument is usually that if there were more competition, the price would be lower, the product and associated services better. But such

denunciation and reasoning does not furnish a norm by which prices may be judged to be either "high" (monopolistic) or "low" (competitive). Nor does it tell us how severe price competition must be for the market to become free from all taint of monopoly. The personal bias of the polemist will obviously influence judgment. Is there any place where the economist may take a convincing stand?

At the outset he should avoid the substantive definition of a monopolist as one who is able to set his selling price. This definition would include all those who from time to time reduce prices in an effort to take customers away from other sellers and who are therefore described by most people as competing with one another. There is nothing to be gained by seeking to reverse popular connotations in this regard.

The functional approach offers a better, if not perfect, solution. Anti-buyer behavior may be divided into five parts as follows:

FORMS OF ANTI-BUYER BEHAVIOR

I. Raising prices when demand rises, but costs do not rise, or do not rise as much
II. Sustaining prices when demand falls (maintaining "rigid prices")
III. Sustaining prices when unit cost falls (appropriating the gains of "technological progress")
IV. Charging some buyers more than others ("price discrimination")
V. Charging a "high" price; making "supernormal" profit

The first four types will be analyzed briefly in the following section. The fifth is more difficult and will be deferred to Section 14.

13. Four Types of Anti-Buyer Behavior by Monopolies.—The first case of anti-buyer behavior is that in which a seller raises his asking price following an increase in demand. If there were intense price competition among sellers, no seller would raise his price unless driven to do so by rising costs. Therefore, for him to ask a higher price merely because demand has risen means that he is exploiting consumers. This monopolistic analysis must be contrasted with the conventional analysis which shows an increased demand raising prices in a "competitive" market because of the upward slope of the supply curve. That positive slope results from increasing marginal cost or increasing marginal reluctance to sell. It is not the same as the present case where there is no increase in cost, but merely a changed idea of "what the traffic will bear." Or the increase in cost is insignificant in relation to the increase in price.

The second case in the above outline refers to situations in which sellers maintain prices unchanged when demand falls, costs remaining

unchanged. Such sellers choose to keep prices constant rather than to adjust to the downward shift of their revenue curves. "Rigid prices" should be contrasted with what may be called "rigid quantities." In the latter case, a seller faced by declining demand tends to maintain his former rate of output and takes whatever lower price it will bring. Behavior of the first type is often called monopolistic and suggests a perfectly elastic (horizontal) supply curve. Behavior of the second type usually is associated with substantive competition and suggests a completely inelastic (vertical) supply curve. A comparison also should be made between the vertical supply curve of the extreme anti-rigid-prices school and the upward sloping supply curve of an industry in conventional analysis. Under the latter assumption some curtailment of supply is to be expected when schedule demand falls. High cost producers will be forced out and the marginally high cost production of those who remain will be eliminated as unprofitable. But this is clearly not the same as supply curtailment by those who engage in monopolistic pricing in the face of declining demand.

The third case of anti-buyer behavior is the least disputed of all. It occurs when costs decline because of technological progress or a drop in the price of one of the factors. If there is highly competitive behavior, the price of the finished product will quickly be cut and the total net revenue of the firm will tend to remain the same as before. Monopolistic behavior, on the other hand, will seek to keep prices up so as to keep for the firm most of the gain from the reduced cost. Prices will remain unchanged or will not be reduced as much as would be required if all the benefit from the cost reduction were to be passed on to the consumer. It is interesting to note at this point that the denouncers of monopoly will often admit the desirability of stimulating technological progress by temporary patent rights. This admission that monopolistic behavior is sometimes socially beneficial at the same time that it is anti-buyer in the short run opens the door to an extensive social critique of competition versus monopoly which is of great importance, but cannot be presented here.

A fourth type of monopolistic pricing which is often denounced as harmful to buyers is price discrimination. This was discussed at length in Chapter 9, Sections 10 to 14. It includes all forms of price differences quoted by a given seller which are not in proportion to differences in cost. Those who have to pay the higher prices may claim that they are being discriminated against in a way that would be impossible if competition were present. This is probably true, but competition cannot always be present, or if present might be harm-

ful, as with certain public utilities. High per ton mile railway rates, high utility rates for individual home-owners, etc., are often defended on the grounds that these rates are lower than they would be if the selling firms were unable to derive some revenue from business which would not be obtainable unless granted very low rates.

Not all price discrimination is anti-buyer and therefore monopolistic. Some may be pro-buyer and competitive. The difference may be seen by referring back to Figure 43. All prices charged above the equilibrium (uniform) price $OP = QB$ are anti-buyer. But sellers may also favor certain buyers by offering them goods at less than $OP = QB$. We usually call this "dumping" and recognize it as a competitive effort to win customers from other sellers. This case illustrates again the advantage of the functional approach to competition and monopoly. The substantive approach would bring both types of activity together since they both depend on the presence of "monopolistic" power to set selling prices.

14. Monopolies Charge "High" Prices and Make "Supernormal" Profits.—The fifth type of anti-buyer behavior is perhaps best known but it is very difficult to describe with precision and is least well suited to the functional approach. To appraise situations in which firms set prices "too high," monopolistic pricing must be defined in terms of its results in comparison with competitive pricing. This requires the assumption that there exists a normal competitive relationship between price and cost, e.g., that which yields normal profit. Monopolistic behavior then becomes that which sets prices which yield supernormal profit. This may sound like a very neat disposal of the whole problem, but it is really a deceptive trap unless the terms are very carefully defined, and even then it is tricky.

In the first place, normal profits must be expressed as a rate of return on net worth. This normal rate is not a fixed figure, but varies from industry to industry and from time to time. It is best defined in terms of the behavior or the desire to act which characterizes firms within and without the industry or situation. Normal profits do not attract sufficient entrants to cause an expansion of total production or productive capacity, but merely enough to offset exits. Normal prices in an industry are those which yield normal profits at the rate of sale which results. These are the prices presumptively set when there is "normal" competition among rival firms, although the preceding discussion clearly indicates the possibility that competition may meet with buyer approval even when it reduces profits for sellers to the vanishing point. There is a difference between the norm of

the economist talking in terms of the long run and that of the consumer appraising his short-run position as buyer.

The best that can be done along this line, therefore, is to define monopolistic behavior as that which sets prices so high that they yield supernormal profits according to somebody's criterion. If the judgment is rendered by people with money to invest who would like to earn similarly large profits, competitive behavior may sooner or later develop to force profits downward, and possibly prices too.[5] The economist is most likely to follow this approach.

On the other hand, a "substantive monopoly" may charge such low prices or experience such difficulties in getting buyers to purchase that it makes only a normal or even a subnormal profit. It should not then be denounced as a "monopoly," since from the viewpoint of results it is pricing "competitively."

Supernormal profits often depend upon hindrances to full production by restriction of output. If a single firm is involved, it will be able to maintain the "high price" which exploits buyers only if it limits its output to the amount it can sell at that price. The conventional analysis shows the maximum profit combination of price and output to be at a higher price and lower quantity than if the firm produced at minimum average total cost. If the latter point is further defined as the "competitive output," on the grounds that it will yield "normal profit," then the smaller output which yields a larger profit must be called something else, e.g., "monopolistic." [6] If several firms collude, and some or all of them have considerable excess capacity when they sell at the monopolistically established price, it is usually necessary for them to engage in further monopolistic action by agreeing to restrict output. Quotas are set so that the total amount offered on the market will not exceed that which buyers are expected to take at the fixed price. It may even be less if there are excess stocks to be moved.

A final comment about the supernormal profit approach to monopoly might be that such profits are hidden as quickly as possible. Assets are revalued upward, stock is watered, salaries are padded, and other techniques are used to prevent both buyer dissatisfaction and outsider jealousy. An economist must also be a good accountant if

---

[5] To simplify the discussion, comment upon obvious relations between price and quantity sold, type of product and cost of production, selling expense and total expense, etc. has been omitted in most cases.

[6] Supernormal profit for individual firms is possible, of course, without monopolistic behavior. For instance, the intra-marginal producers of a homogeneous product sold at a uniform price may make "efficiency profits" by devising ways to reduce cost. See Chapter 25, especially Section 7.

he would use this criterion of monopoly. Even then he would still face the very difficult problem of appropriate allowance for efficiency profits.

15. Summary.—Market prices are determined in many ways. These are most fruitfully classified as behavior types: by people acting independently, by people acting together, and by governments. When people act independently they compete against one another in marketing, product quality, purchasing, cost reduction, timing, etc. Monopolistic behavior includes collusion to set price, to control output, to allocate markets, etc. It may also involve action to discourage entry, and to exploit buyers.

The functional classification of types of competition and monopoly emphasizes what sellers and buyers do in the market. Its emphasis is institutional. Its generalizations are more varied than those of the substantive approach. It examines with semantic care the referents of words of wide usage in both popular and technical works on economics. Realism is the objective, rather than the solution of formal problems of logic. No institutional analysis can be completely realistic, because economic phenomena are infinitely varied in time and place.

Chapter 12

CAUSES AND EFFECTS OF COMPETITIVE AND MONOPOLISTIC BEHAVIOR

1. Causes of Different Types of Market Behavior.—In the preceding chapter on different types of market behavior, scattered references were made to the *causes* of action of one type or another. It is now appropriate to examine both causes and effects in more systematic fashion. In the first place, we assume that in a capitalistic society the major cause is the desire to maximize profit. As indicated in Section 3 of the preceding chapter, this goal may be sought by different methods of market pricing and behavior. Four of the most important are the following:

1. Independent price-setting and other forms of competition
2. Collusive group action to set prices and in other ways to restrict competition
3. Accepting the price (and terms) set by others; competing in other aspects of market or production behavior
4. Bargaining with buyers

In the second place, action in any of these four ways may be based upon a variety of conditions. None of them arises from a single, unique pattern. Finally, although any given condition may generally give rise to a certain pattern of behavior, the possibility of exceptions must not be overlooked. One entrepreneur may seek maximum profits in the short run, as by monopolistic pricing, whereas another may react to a similar situation by competitive pricing, believing that to be best in the long run.

Because of these varying possibilities, the classification of behavior in terms of substantive causes is difficult. It seems better to base generalizations directly upon types of behavior and take note of exceptions to causes instead of exceptions in action. Possible reasons for using each of the four methods listed above will be analyzed in the next four sections. After a couple of diagrammatic notes, the chapter will conclude with a discussion of temporal and spatial price patterns resulting from competition and monopoly.

2. Causes of Independent Price-Setting by Sellers.—An individual seller may choose to set his own price rather than to collude, to accept, or to bargain, if one of the following three major situations of fact and opinion is present.

I. The seller may perceive no close rival who might be influenced by the price set and therefore he feels no danger of retaliation. This assumption of low cross-elasticity of price between A and other products, B, C, D, etc., may be either a fact or an illusion. If it is true, it is due to the first two of the following causes or to the third:

   A. Product differentiation or market isolation ("a wide gap in the chain of substitutes"). This in turn is caused by:
      1. Product differentiation by a single producer, achieved by such things as:
         (a) A product design protected by patent, trade secret, time required to imitate, etc.
         (b) A trade-mark protected by copyright and bolstered by advertising, customer experience, etc.
         (c) The ownership of scarce natural resources
      2. Some type of merger or agreement among various producers of similar products
   B. A *relatively* low price for the product. (Buyer substitution is a function of relative prices as well as relative qualities. A seller may be a functional monopolist at *any* price range, but not a substantive monopolist. There is always some price above which cross-elasticity will become apparent for any good, no matter how differentiated.)
   C. The ignorance or indifference of rivals
      1. The price-cutting seller may be small in relation to other sellers.
      2. Price cuts may be secret.

II. The seller perceives rivals who will be influenced by his price, but acts independently nevertheless. (In this case there is appreciable cross-elasticity.)
   A. He may feel so much stronger than his rivals that
      1. He thinks they will tag along ("price leadership"), or
      2. He thinks he can beat them in a test of strength.
   B. He may think he can keep his price cuts secret.
   C. He may be a newcomer who must do something different to break into the field, such as price-cutting, special product differentiation, or ballyhoo.
   D. There may be general or particular overcapacity and price-cutting becomes an action of last resort.

III. The seller can postpone selling if there are no buyers at the price he sets.
   A. He may be a single seller of a differentiated product, such as a piece of real estate.
   B. He may be one out of many sellers of a homogeneous product who is willing to sell now if he gets his price or better, but otherwise prefers to wait. American wheat farmers often fall into this class.

For cross reference with the traditional approach one should note that substantive monopoly appears above under I-A, III-A, and possibly II-A-1, although most economists would probably put most of II under the heading of substantive oligopoly. Substantive pure competition should be associated with III-B, although it is rarely presented with that type of emphasis.

To complete the discussion of independent price-setting by sellers, a note might be added concerning the background of price discrimination. Three major causes or conditions are usually described (cf. Chapter 9, Sections 13 and 14).

1. There is no close rival. (This is usually interpreted to mean substantive monopoly, but it need not. When sellers formerly engaged in active competition with each other find that the demand at a fixed ceiling price exceeds what they can supply, they may engage in various types of "black market" price and customer discrimination. There is no merging, no collusion, no change in the number of sellers, but a different type of behavior emerges from the combination of scarcity and price ceilings.)
2. The total demand schedule can be subdivided, with different elasticities of demand in each subgroup.
3. The cost of keeping markets separate is not large relative to the gain from price discrimination.

3. Causes of Collusion.—Collusion occurs whenever a group of sellers believes that profits can be maximized better by concerted action in some respects than by rivalrous action. Functional monopoly is preferred to functional competition. Among the possible methods of securing noncompetitive behavior, collusion is adopted rather than outright combination whenever sellers do not want to surrender their right to compete in other respects. The fear of antitrust prosecution may be an influence on either side. It may discourage collusion, may cause it to be informal, or may stimulate complete merging of former rivals. A more detailed statement of the causes of collusion follows:

I. The product is homogeneous or is only moderately differentiated. (There is appreciable cross-elasticity among the products of different sellers.)

II. There is excess capacity in the sense that not all firms can produce at minimum average total cost and sell the total output without forcing the price down below that figure. This, in turn, may be caused by:
   A. A decline in schedule demand
   B. The entry of new firms, usually low-cost producers
   C. The expansion of capacity by existing firms
   D. In agriculture, a short-run glut may result from unusually favorable growing conditions

III. Belief that the "industry" demand schedule is inelastic above the prices prevailing prior to collusion.

IV. Government influence.
   A. To aid sellers thought to be harmed by too much price competition. Cf. AAA, NIRA, NLRA, etc.
   B. To promote the achievement of foreign policy objectives. Cf. permissive and compulsory international cartel activity

4. Causes of Price Acceptance.—Sellers accept the prices set by others under quite a variety of circumstances including the following:

I. The buyer may set the price and refuse to bargain.
   A. The seller may be unable to find any other buyer who will pay more, or
   B. The seller may deem it unwise for him to wait for the buyer to relent or for other buyers to appear.

II. The seller may fear imitation or retaliation if he acts independently.
   A. Strong rival sellers may cut prices as much or more.
   B. Bankers, union labor, suppliers of materials, and others from whom the seller buys and who have an interest in uniform action by all sellers, may coerce the one who breaks ranks.

III. The seller may see no possible gain from asking less.
   A. He can sell at the going price all that he possesses or can produce. This means that either:
      1. There is a buyer who will take all that the seller offers at a fixed price, or
      2. There are so many buyers and sellers that the given seller can supply only a very small part of the total.

IV. The seller may believe he would lose by asking more.
   A. He thinks that buyers would quickly turn to rival sellers who offer similar or identical goods at unchanged prices.
   B. He fears the penalties of a law imposing price ceilings.

These four cases of price acceptance by sellers include the situations frequently defined in traditional classifications. For instance, substantive monopsony is akin to I, substantive oligopoly resembles II-A, and substantive pure competition fits III-A-2 and IV-A. The last two categories were also illustrated in Section 2 under the heading of price-setting. It should therefore be clear that the traditional categories do not give rise to any type of behavior peculiar to them. This is true whether the substantive distinction be made upon either the number of sellers or product differentiation. One cannot reason backward from the function to say that price-setting is exclusively the activity of substantive monopolists, or that price accepting is sure proof of substantive pure competition. The merits of the functional approach are also supported by Section 3 on collusion and the following section on bargaining.

5. A Diagrammatic Note on Traditional Demand and Supply Curves for Different Types of Competition.—The traditional substantive approach to a "monopolistic" situation defines it as one in which an individual seller, or a group acting together, chooses the selling price which is expected to yield the greatest net profit. The price-setting is said to occur against a background of a sloping demand curve with AR above MR like those of Figures 33–36 or 45. The $MC = MR$ point, when found, leads to the price-quantity solution of the maximum profit problem. Though there is a tendency to establish such an optimum price, the actual behavior of "monopolistic" sellers is much more complex than mere price-choosing by formula. The preceding chapter has outlined and explained many different kinds of monopolistic behavior. The argument also shows that price-setting is as much a part of competition in the functional sense as it is of monopoly.

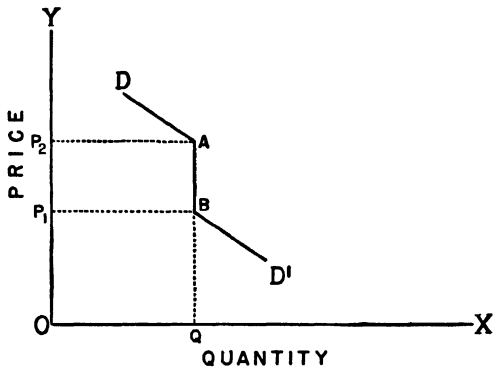
Three other substantive classifications are worthy of consideration. The first is pure competition, described as a situation in which there are so many sellers of the same type of product that no one of them can set the selling price independently of the others (cf. Section 4, III and IV). The traditional diagram has a horizontal demand curve ($AR = MR$) like that of Figures 30, 31, 32, and 46.

The second is pure monopoly. The description and diagrammatics of pure monopoly are not as uniform as those for pure competition. The most logical picture would seem to be that of a single seller, or group, able to raise the selling price without any loss of demand. This is the exact opposite of pure competition where any rise in an individual's selling price means complete loss of demand.

Diagrammatically it is also opposite. The demand curve is vertical instead of horizontal. Such a situation is highly unreal, but it is the true limiting case. If it ever did exist, the monopolist would raise his price to the point where demand began to fall off. He would revert to the sloping demand curve of "monopolistic competition" for any further price increases. This is shown in Figure 47. There is

FIGURE 47

PURE MONOPOLY FOR A LIMITED PRICE RANGE



"pure monopoly" for the firm selling at P_1, but after it raises its selling price to P_2, further increases encounter a sloping AR curve like that of Figures 33–36 and 45.

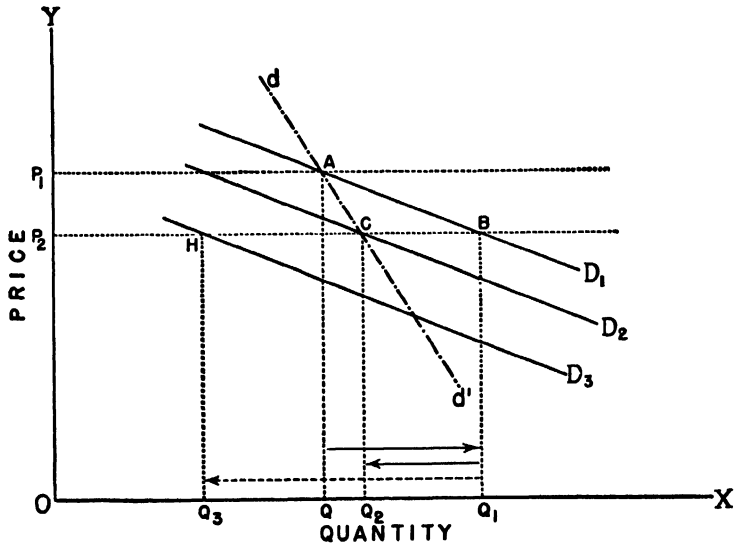
Another substantive classification of forms of competition which appears in most discussions of the subject is oligopoly. Here there are a few firms producing the same product, each conscious of the way in which its own actions may have a boomerang effect. Firm F, for instance, may think that a price cut would attract customers from its competitors. But if these competitors follow with a similar price cut, they will retain most or all of their customers. Firm F will gain something, but less than expected. There are at least three possibilities. Some buyers will be attracted by F's bold move in making the first price cut and will not return to former suppliers. Second, former customers of F may be induced to buy more than before. Third, a few new buyers may be attracted to the product by its now low price and Firm F will get its share. The latter two gains are shown in Figure 48 as those resulting from moving down the "industry demand curve," *dd'*, from *A* to *C*. If competitors had not followed suit, Firm F would have moved down its independent demand curve, D_1, from *A* to *B*. If the process is described in two steps,

sales of an oligopolist may rise first from OQ to OQ_1 and then drop to OQ_2.

In this traditional diagram there is no easy way to show the increment of customers taken from other firms. And if competitors

FIGURE 48

THE BOOMERANG EFFECT UNDER OLIGOPOLY



retaliate by cutting prices more than Firm F, demand may drop to D_3. The new sales quantity, OQ_3, will be less than OQ.[1] The boomerang effect in such cases really hurts. Even the sales increase from OQ to OQ_2 may be harmful if costs do not decrease as much as the prices drop.

These possible adverse effects of price-cutting often lead to acceptance of the going price as indicated in the preceding section.

6. Causes of Bargaining.—Sellers use bargaining to determine the price when they think that the demand is inelastic over a certain price range. In such circumstances a seller sets his initial price high and shades it as little as he can. The buyer, conversely, sets his first price low and comes up as little as possible.

This is often called a "bilateral monopoly" situation, but it is a

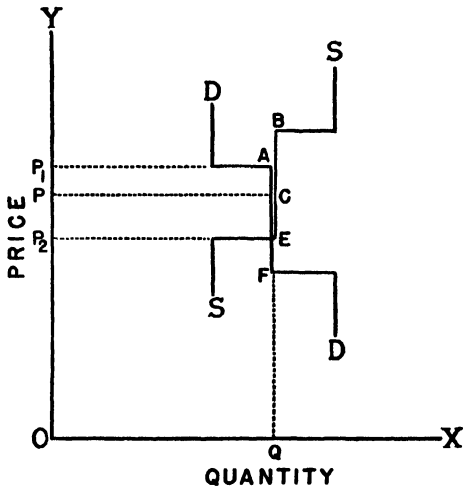
[1] A line connecting A and H would *not* now represent an "industry demand curve" like that between A and C. Competing firms would have different prices. This could happen if location or service differences made buyers prefer to deal with one seller rather than another. The full product homogeneity of traditional theory is rarely found in practice.

type of behavior peculiar to itself and can hardly be called anything other than bargaining or "higgling." In one sense the good involved may be considered different from other goods offered or sought. In another sense the bargaining merely occurs because either the buyer or the seller does not want to take the time and trouble to hunt acceptable substitutes. Any failure to reach a mutually acceptable compromise is evidence of belief in the availability of such substitutes.

7. A Diagrammatic Note on Bargaining.—A diagram like Figure 49 helps to clarify the concept of a bargaining range within which the price is determined by bargaining. Assume that DD shows the demand curve of a buyer who will buy quantity OQ at any price from QF up to QA. A seller in the same market will sell OQ at any price from QB down to QE. The bargaining range is AE where the two curves overlap. If the seller is the more clever bargainer of the two, the final price, P, will be closer to P_1 than to P_2. In formal theory the AF vertical segment of the demand curve would indicate pure monopoly for the seller for that price range. It is more often called

FIGURE 49

THE BARGAINING RANGE WHEN MONOPSONIST MEETS MONOPOLIST



merely a discontinuous demand curve. Similarly, the buyer would experience pure monopsony as he moved from B to E on the seller's supply curve.[2]

[2] The shape of both curves above points A and B and below points F and E may be either slanting, curved, or, as here, of the staircase type.

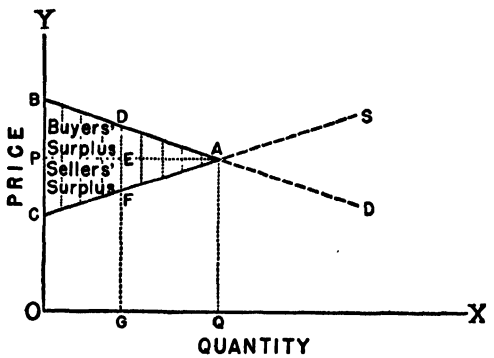
8. Bargaining and the Emergence of a Subjective "Surplus."— If the bargaining between monopsonist and monopolist results in a price such as P in Figure 49, both buyer and seller will feel that the price is a good one. The buyer may say to himself that he obtained the goods for less than he was willing to pay. He has "saved" the amount PP_1. This is often called a "buyer's surplus." The seller may also feel that he has gained PP_2, which is the amount obtained above his minimum price. This gain is known as a "seller's surplus."

Surpluses of this type also emerge from other market situations. The most common is that in which the market price P is set by the seller and buyers appear who are willing to pay more than P but do not have to do so. A housewife may go to market expecting to pay 70 cents for a pound of bacon. When she gets there, she finds that bacon is on sale that day for 65 cents. She has "saved" 5 cents. Similar illustrations on the other side of the market are not so common. But sometimes a producer may start producing a good with the expectation of selling it at, say, $1.00. When the good is ready for the market, he may find that it can be sold for $1.25, a gain of 25 cents.

There is a third possibility which has been discussed extensively pro and con. It is based upon continuous demand and supply curves such as those of Figure 50. If the demand curve represents a col-

FIGURE 50

BUYERS' AND SELLERS' SURPLUSES MAY EMERGE FROM CONTINUOUS DEMAND AND SUPPLY CURVES



lective cumulative schedule in which each buyer wants only one unit, such as one refrigerator, then a total buyers' surplus or consumers' surplus (BAP) can be computed in terms of dollars saved. If the supply curve is taken as a cumulative average cost curve for a group

of suppliers, a sellers' surplus or producers' surplus (PAC) can be shown.

When the demand curve is the cumulative total demand of buyers who want more than one unit, a buyers' surplus is more difficult to prove. Some argue that the principle of diminishing marginal utility indicates that the buyer would have paid a high price for the first unit, a slightly lower price for the second, and so on until the nth unit whose potential bid price is equal to the market price. The difference between the market price actually paid and that which would have been paid for each unit *if the goods had been bought one at a time at the maximum bid price for each successive one* may then be added to obtain a total consumer's surplus (BAP). The unrealism of this argument lies in the fact that the *total* quantity would not be bought at any higher unit price than that actually paid for the "last" or "marginal" unit. Therefore there is no consumer's surplus at all. Even if the diminishing marginal utility approach be accepted as describing the basis of the demand curve, there remains the problem of adding utilities to get a total in "utils." When different people are involved in a collective demand schedule, the task is impossible. Nevertheless, even if a buyer's surplus cannot be measured under these conditions, the concept is a useful one in certain economic problems, such as that of achieving equal sacrifice in distributing the tax burden.

On the selling side, the difficulty arises from the usual definition of the supply curve as a *marginal* cost curve. This is desirable and logical for small changes in the vicinity of the market price, but such a curve cannot be extended all the way back to C as shown in Figure 50. All the fixed costs would have to be included in computing the marginal cost of the first unit. This would cause C to be very high, far above B. The seller's "surplus" then becomes very difficult to calculate or to demonstrate on a diagram.

9. Parallel Classifications Are Possible on the Buying Side.—Almost all of the foregoing analysis has dealt with the selling side of the market. Similar classifications may be made also on the buying side. Functionally this requires description of various types of competition among buyers. Monopsonistic behavior replaces monopolistic. It includes anti-competitive efforts by collusion, anti-entry activities, and anti-seller devices. Individual buyers try to improve their buying techniques and to bargain for as low buying prices as possible. The substantive categories are also similar. Pure monopsony exists when the supply of the good being bought is in-

elastic for price cuts over a certain price range. Monopsonistic competition is revealed when buyers cannot get more goods except at higher prices. Under pure (buying) competition the supply curve as seen by the individual buyer is perfectly horizontal. The latter situation is very common because of the habit of sellers to quote fixed prices. It is modified by the practice of quoting quantity discounts.

Most firms operate monopsonistically in the purchase of some of their materials, equipment, or services. They operate competitively in "shopping around" to buy other things at lower prices than their rivals. They accept prices fixed by a seller only when they cannot bargain the price down or find a better price elsewhere. It is much less common for them to fix their buying prices than to fix their selling prices. Only very large buyers or well organized groups can do both.

10. Patterns of Price Behavior.—One of the main reasons for attempting classification in the field of competition and monopoly is the desire to explain various types of price behavior. We may also want to predict or to control price movements which we do not like. When the emphasis is upon price patterns, the classification may be either temporal or spatial.

The price changes which occur through time may be compared in terms of either their relative frequency or their relative amplitude. The prices of some commodities may change many times in a business day, like wheat on an active grain exchange. Other prices may not change for months or years at a time, like nickel candy bars, street car rides, gold at the mint, or new Fords. A price may be stable at one season of the year, but fluctuate considerably in others, like hothouse vegetables as compared with field grown. Or there may be long periods of stability interrupted by vigorous price wars, as in the case of retail gasoline in some areas.

So far as amplitude of price change is concerned, differences in range from high to low may be noted. Some prices do not depart very much from their norm of cost of production while others do. Durable goods and income-yielding property probably have the widest range of fluctuation. In other cases changes in the cost of production are significant, such as seasonal crops. Some price patterns reveal inflexibility upward, but considerable amplitude downward, while for other commodities the reverse is true.

Although the foregoing list of patterns of price behavior is far from complete, enough cases have been cited to indicate that there are many determinants of price patterns other than the degrees and

types of competitive behavior. With this qualification in mind, the competition-monopoly aspect of price stability may now be explored.

11. **Causes of Price Stability.**—Prices change most frequently when the exchange price is not set by single individuals, but by bargaining or by the balancing of equal quantities at various bid and offer prices in the market. A major reason for instability in the prices of staple goods traded on the exchanges is the frequent change in expectations. Bid and offer prices fluctuate with every rumor and every prediction of change. Even without the presence of professional speculators, exchange prices would change because producer-sellers, not to mention buyers, often do amateur speculating on their own part.

Prices tend to greater stability when individual sellers are able to determine the exchange price by setting their offering price. Although custom, contract, and demand inelasticity are also important causes of stability, monopolistic behavior must be given considerable weight. The functional monopolist is one who takes steps alone or with others to insulate the market from the usual forces causing price change. For various reasons he prefers price stability. Therefore, instead of changing his offering price with every minor fluctuation in demand or cost, he keeps it constant for weeks, months, or even years. He does not act like the substantive monopolist of traditional equilibrium theory whose asking price is always set so that marginal cost equals marginal revenue. Price stability tends to become a goal in itself, a rival to that of maximum profit. This is true, even though capitalistic apologists may argue that the individual entrepreneur is always seeking maximum profits "in the long run," and that *MC* and *MR* curves should be interpreted in that light. As indicated above in Chapter 11, Section 12, price stability is often the most prominent tactic of collusive monopolies. It is also common among collusive monopsonies.

Periods of stability interrupted by price wars generally indicate the presence of collusion among a few big producers plus opportunities for fairly easy entry by small producers. This type of price behavior will also be likely to occur if the major firms cannot quickly adjust supply to demand, if they cannot store raw materials or finished products in time of glut, and if there are no significant economies of scale.

Firms which cannot control the prices of their major raw materials are more apt to have fluctuating selling prices than those which are integrated vertically. The same is true for those who buy in a mar-

ket where prices fluctuate as compared with one which is monopolistically stabilized.

Even if all substantive monopolists sought to price their products according to the economist's formula, there would be differences in the patterns of price behavior. Demand changes and cost changes occur with different frequencies for different products. And the time-elasticity of supply varies considerably. Durable goods, for instance, reveal much greater amplitude of fluctuation on the down side than do nondurable goods. Therefore, it is wise to remember that substantive classifications of monopoly or competition in terms of frequency of amplitude of price change must be very loose, indeed. The ability to set exchange prices is only one of many variables that influence the degree of price stability.

Stable prices in the time sequence of an actual market must not be confused with stable prices in the equilibrium sense of economic analysis. Much of the traditional substantive approach is linked to the latter in logic, but has been transferred to the former by inference. For instance, the equilibrium price under "pure competition" is sometimes held to be more "stable" than the equilibrium price under "pure monopoly," whereas observation seems to show that the reverse is true. Part of the trouble undoubtedly lies in the vagueness with which the demand and supply schedules are often defined in each case, but even the sharpest substantive definitions give analytic tools inferior to those of the functional approach.

12. A Functional View of Geographic Price Patterns.—There are also many patterns of geographic price dispersal at any given time. Three types are quite common. One is the uniform, or postage stamp type of price. A second is the transfer differential pattern with differences closely proportional to transport costs from a central market or place of production. The third is the basing point system, where one or more cities constitute the places from which transportation costs are figured regardless of actual point of production. There is also a zone delivered system like parcel post rates where the average transportation cost, or something close to it, is considered in computing a delivered price for the entire zone. This is intermediate between the first and second types given above. Many non-systematic patterns also exist influenced by custom, differing costs of production, different degrees of competition, density of population, number of production points, etc.

Differences in geographic price patterns depend only in part upon the geographic dispersion of price-competing producers. If they

agree to act together in one way or another, the pattern will depend chiefly upon the type of freight charging or freight absorption scheme adopted. If they act independently, the price pattern depends upon relative costs of production, the distances between competitors, and their fighting strength. Price competition will be more severe and the price pattern more unpredictable if dumping practices are followed than if a uniform price is charged to all sellers F.O.B.

Here again a monopolist is what a monopolist does. If a firm chooses to engage in extensive price discrimination practices, it may be acting either competitively or monopolistically. When price discrimination involves cutting prices for certain sales *below* the average total cost (including necessary transportation costs and a normal profit), then this should be called competitive action even though its intent is to extinguish price-competition in the future. On the other hand, when a seller's price discrimination pattern involves only a series of price levels *upward* above average total cost, this is monopolistic pricing. This pattern is a method of exploiting buyers for the benefit of the seller, whereas the former type of price discrimination benefits certain buyers, at least for a time. (Cf. Chapter 11, Section 13, and Chapter 9, Sections 13 and 14.)

Neither intent nor future possibilities of opposite action should lead us to call competition "monopoly." We may condemn regional dumping as likely to produce anti-social results, but uniform pricing with competitive intent might also lead to the same outcome. In retail selling all buyers may be charged a below-cost price by a given firm when it promotes a "loss-leader." This is competition and it may or may not lead to the ousting of rivals who cannot stand long-continued losses. Consistency would dictate that he who proposes to ban regional dumping should also support "Fair Trade Laws." Each proposal condemns a particular discriminatory price competition which weak firms don't like and whose results economists may disapprove.

13. A Social Appraisal of Competitive and Monopolistic Behavior.—Since the main purpose of this chapter has been to suggest a functional approach to the concepts of competition and monopoly, it is not necessary to elaborate upon social implications. However, certain inferences may be made explicit. We should not condemn monopoly, substantively defined, without examining its behavior. Some of that behavior is competitive and some is monopolistic. Furthermore, not all competitive behavior is good nor all monopolistic behavior bad. It depends upon the results and upon the standards of

goodness which we use in evaluation. A few illustrations may be given.

Certain forms of competitive behavior seem to be anti-social according to generally accepted criteria. Most obvious is that which bankrupts newcomers or noncooperators in an effort to force them out of business and is followed by a subsequent price boost which exploits the buyers for the benefit of the firm that initiated the squeeze. But it is hard to draw the line between good competition of this type and bad. Sometimes only experience can prove that monopolistic behavior will be the ultimate result. And in the meantime consumers benefit.

When there is too much selling competition, we generally feel that resources are wasted which might have been put to more productive use. A welding of the competing advertisers into one large unit by a holding company or other technique might reduce this waste. But, of course, it might also lead to monopolistic activity of another kind which we would condemn.

The problem of excessive entry with resultant overcapacity and high unit costs is well known in some fields like retail groceries and gas stations. Behavior opposing entry has been called monopolistic, but it may succeed in preventing a waste of society's resources. Franchises and patents are governmentally sanctioned methods of restricting entry, presumably in the public interest. We restrain competitive behavior in this respect or in that and then hope that anti-social monopolistic behavior will not emerge. But it often does.

When competition becomes very keen, firms are sometimes led to exploit their workers or other suppliers, instead of their customers. The government has recognized this danger in the soft coal mines, for instance, and has introduced compulsory cartellization. But it is very hard to keep the balance. Competition between firms has been their protection in the past. Is government control sufficient for the future?

Another form of monopolistic behavior which has received the stamp of government approval is that collusion among workers known as collective bargaining. Here again arises the problem of distinguishing between the subsequent forms of activity. Some are accepted as good according to most criteria, some denounced as bad, and many are highly debatable. No appraisal will be attempted here, but the case is offered to support the argument that intelligent social control must go beyond the forms of monopoly to the types of behavior that emerge and must then distinguish among these the ones that are anti-social and should be controlled. Otherwise we may

throw out the good along with the bad. Our predilection for emotionally colored labels has often betrayed us.[3]

14. Summary.—The causes of market behavior patterns are found in part in the selling and buying conditions traditionally used in defining substantive monopoly, competition, and oligopoly. Pricing methods include price setting, price acceptance, and bargaining. These functional classes cut across the traditional categories, but seem more realistic and useful. Parallel classifications on the buying side exist for both functional and substantive viewpoints.

Buyers' and sellers' surpluses may be distinguished under certain assumptions regarding subjective maximum buying prices, subjective minimum selling prices, and objective costs. Temporal and spatial patterns of price behavior may be described. Monopolistic behavior promotes price stability through time. It also leads to geographic patterns of fixed price differentials. Monopolistic behavior generally is socially harmful, but there are frequent exceptions. Competitive behavior is usually beneficial, but not always.

Generalizations about methods of price determination should be made with care and applied with still more care.

[3] In the interest of brevity, comments upon other criteria for judging the merit of competitive versus monopolistic behavior have been omitted. These include effect upon technical progress, saving and investing, level of employment and output, quality of goods, incentives to government control, etc.

Chapter 13

TIME FACTORS IN INDIVIDUAL PRICE TRENDS: COST CHANGES

1. Introduction.—Chapters 11 and 12 have demonstrated the importance of the degree of price competition in the determination of the market price. We now return to the cost arguments of Chapter 8. But the approach is different. We want to know why prices *change*. The explanation might be sought in changes in the degree of price competition, in laws and court decisions, in sellers' expectations, and in other things. In this chapter, however, our attention will be centered upon cost changes as causes in periods of rising demand. The next chapter will examine some of the other possible causes of price change and will consider falling demand.

Specific questions to be answered in the next few sections include:

1. What is the concept of "time-elasticity of supply" and how does it differ from simple "price-elasticity"?
2. What three time periods of supply change are commonly described when demand rises?
   (a) What is "short-run" supply?
   (b) What are two ways of looking at "moderately long-run" supply curves?
   (c) What are "very long-run" supply schedules and how may they be used?
3. Is monopoly inevitable in the long run under conditions of decreasing cost?
4. What other cost changes influence price?
   (a) How summarize cost reasons for a downward price trend?
   (b) How summarize cost reasons for a rising price trend?

2. Time-Elasticity of Supply.—The definition of supply schedules and curves in terms of their setting in time involves three problems: (1) the difference in the times required for different commodities to evince the same price-elasticity of supply; (2) the different price-elasticities of supply for a given commodity in time periods of different length (usually increasing length); and (3) other cost

changes, i.e., those occurring independently of the variations in the schedule demand for the commodity which produce the price changes which serve as independent variables in the first two problems.

The concept of time-elasticity of supply is useful in explaining the amplitude and duration of individual price fluctuations. The term may be defined as the length of time that is required for a given percentage change in price (or rate of sales) to provoke the same or any other given percentage change in supply (or rate of production). Time-elasticity of supply is "long" or "short" only on a relative basis as the supply of one commodity is compared with another. For instance, for a commodity such as copper, a 10 per cent increase in price may cause a 5 per cent increase in supply within a month. On the other hand the quantity of new houses offered on the market each week may rise only 1 per cent within a month under the same price stimulus and a period of six months may be required for a 5 per cent expansion.

In several earlier sections of this work the concept of normal price was discussed, but little attention was paid to its time frame of reference.[1] This deficiency may now be corrected, at least in part, by examining the different time periods allowed a given producer or an industry to adjust supply to changes in price. When price is assumed to be the independent variable, then the change in quantity supplied, the dependent variable, ordinarily will be larger if a long time is permitted for readjustment of supply conditions than if only a short time is allowed. That is, the elasticity of supply is a function of time as well as of price change and therefore a three-dimensional diagram showing a supply surface would be more realistic than any single two-dimensional curve could be. However, the cumbersomeness of three-dimensional diagrams and three-variable tables suggests the construction of a series of two-variable schedules and curves, each for a different time period. The ones most commonly used are those for the short run, the moderately long run, and the very long run, although different authors use different terminology.

Certain difficulties should be acknowledged at the outset. *First,* the time periods will have different chronological lengths for different commodities, for different firms producing the same good, and even for the same firm producing the same good under different circumstances. This greatly complicates the problem of constructing a realistic schedule for accurate prediction of future relationships. It often leads to confusion through unconscious transference of time

[1] See, for instance, Sections 8 and 9 of Chapter 10.

durations from one situation to another, as from agriculture to manufacturing or retailing, or from cases where price competition prevails to cases where such competition does not exist. *Second,* the quantities of a supply schedule are really time *rates* of sale (strictly, of offering for sale). That is, they are quantities per unit of time. This time unit is always shorter, much shorter, than the time period for which the schedule has been constructed. Thus, in the short run, which may cover weeks or months according to the commodity under consideration, the supply quantity in a schedule usually shows the rate of sale per hour or per day. *Third,* the causal relationship does not always run from price to rate of sale. Under the widely prevailing conditions of monopolistic competition where the seller sets his selling price, changes in the rate of sale influence changes in price (and in the rate of production) rather than vice versa. *Fourth,* it is extremely difficult to give precise definitions of different supply-time periods, but the reason for making time distinctions can be made clear without this precision. Nor is it necessary that all writers use the same terms in designating the different intervals if we realize that the concepts are slightly artificial at best. The important thing to emphasize is the need to recognize and make allowance for elapsed time.

3. **Short-Run Supply Curves.**—The "short run" usually is defined as time long enough to permit increases or decreases in the rate of sale from a given stock of goods on hand which the seller does not have time to alter by purchase or production. It is sometimes called "market time" or simply "time period I." It is best illustrated by reference to agriculture, where seasonal changes or life cycles tend to separate the periods of selling from the periods of producing. In the temperate zone, crops usually are harvested in summer or in autumn. After the harvest period, farmers have a given fixed stock of produce on hand which they may sell at a higher or lower rate depending upon circumstances. One of the determinants is the price offered or obtainable, and its change from recent levels. A second is the urgency of sale, or the seller's need for cash to meet his obligations. A rise in price generally will induce more people to sell their stock than before. But it may cause some to think the price will continue to increase and therefore may lead to a speculative decision to hold back for still higher prices. Much farm produce is sold immediately upon harvesting for whatever it will bring. These sellers leave to speculators and those with storage facilities the problem of guessing which way the price will move in the future.

Those who set the price at which they will sell and then let the demand determine the quantity to be sold operate in a different kind of short run. Merchants, for instance, have customary order periods in which they purchase their inventories and decide upon the price which they think will be most profitable for them to ask when disposing of their goods on hand before the next order period. The price is based upon an expected rate of sale. If subsequent experience shows that sales are slower than expected, the price may be cut. If the rate of sales proves unexpectedly rapid, the price may be raised, but this is not considered good merchandising policy in most cases. Instead, the merchant seeks larger profit by increasing his rate of buying to keep pace with his rate of selling. He is content to make his gains through larger volume. Price increases might antagonize consumers and have a bad effect in the long run.

In the agricultural example which has been described in an earlier chapter as pure competition or cost competition, producers adjust their rates of sale to changes in the price. On the other hand, the merchandising example, which illustrates imperfect competition or price and service competition, shows producers adjusting their asking prices or their order rates to their rates of sale. In both cases the quantity offered at a given price is offered out of a fixed stock on hand and the cumulative totals offered by all persons selling a homogeneous good may be aggregated into a short-run supply curve or schedule. This curve may shift during the short time period whenever the attitudes of sellers change. The shift is basically the same whether it seems to be to the right or to the left as with quantity changes, or upward or downward as with price changes. A *de*crease in the price asked for a given quantity constitutes an *in*crease in schedule supply just as truly as an increase in the quantity offered at a given price.

Past cost of production or cost of purchase tends to be unimportant in setting reservation prices in these short-run cases for one or more of the following reasons: (1) the cost is not known (as often in farming), (2) speculation is a stronger motive than the quest for normal business profit, (3) the seller must accept the prices buyers offer even if they are below cost, (4) there is urgent need for prompt sale (because of need for cash or because the seller cannot store goods profitably), (5) sellers do not expect to be selling the same sort of goods in the same place in the future (random sellers, or cases where the replacement merchandise will differ from the stock on hand), (6) replacement cost is higher or lower than past cost.

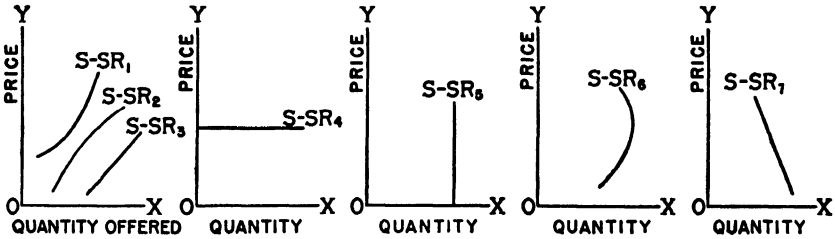
Because of the many different situations involved, there is no one typical short-run supply curve. In cases where a uniform price is

set by the seller, the curve is a horizontal straight line which may move up or down if the price is changed subsequently because of unexpected changes in the rate of sale. In those cases where the price is determined by the bids and offers of many sellers (pure competition), the collective supply curve usually slopes upward to the right, but any individual supply curve may be either positive in slope, perpendicular, negatively inclined throughout or concave to the price axis, i.e., partly negative and partly positive, as shown in Figure 51.[2]

FIGURE 51

TYPES OF SHORT-RUN SUPPLY CURVES

(Most common at left, least common at right)



4. Moderately Long-Run Supply Curves.—The moderately long-run time period usually is defined as long enough to permit expansion of production by the use of additional variable factors but not long enough to permit an increase in fixed factors. Sometimes it is called "long run," "time period II," "intermediate," or even "short run" (in cases where time period I is called "market time"). These alternative definitions are not at all precise if only because there is no hard and fast line dividing variable and fixed factors. However, this vagueness makes little difference if the main purpose is kept in mind, which is to show how an increase or decrease in production may be influenced by the relationship between price and cost.

If we make the same assumptions that were used in Section 10 of Chapter 8, that marginal and average costs can be determined, are known by the seller, and are used by him to regulate his output, then his supply curve will approximate his marginal cost curve in this time period. According to the formal analysis of selling under conditions of pure competition, he will maximize his profit by keeping $MC = MR = AR =$ Price. Since marginal cost rises with an expansion

[2] The negative slope which appears in the last two supply curves is decidedly unusual. However, it seems to occur in the market for some types of labor as wage rates rise. Case 5 may occur in stock markets when falling prices bring out stop-loss orders.

in quantity produced, the cost figures in the marginal cost schedule or on an *MC* curve may be read as equivalent to price figures, so that the most profitable output can be read directly from that schedule or curve.

On the other hand, if we abandon the assumptions of pure competition, and shift our attention to the problem of a price-setter, there is no simple relationship between price and marginal cost. Maximum profit comes at that output which equates *MC* and *MR*, but *MR* is not equal to *AR* = Price. Furthermore, under these conditions of monopolistic competition the increase in demand, as explained above, is not immediately reflected in an increase in price, but rather in an increase in the rate of sales, or sales orders, at the price currently quoted by the seller. (See Figure 54.)

In either case, whether price setting or price acceptance is the policy of the seller, he ordinarily will increase output with an increase in demand. The speed with which this can be done depends upon how quickly he can do two things: (1) procure the necessary additional variable factors of production and (2) combine them in the production process to create that which buyers want, whether it be services, products of a certain form, or goods at a certain time or place. For small increases in output, say 5 to 10 per cent, the factor problem usually is not difficult, although there are exceptions. Extra materials generally can be furnished upon demand from suppliers' inventories. The time required is that of securing delivery and depends chiefly on the distance the goods must move. Occasionally transportation bottlenecks may impede delivery, or problems of financing the transaction may create delay.

In wartime, the chief problem of supply expansion often is that of securing the necessary priority or allocation order from a government agency. In peacetime, sellers of production goods sometimes put their customers under quotas and refuse to sell them more than a fixed amount, so that if a certain buyer wishes to expand his output he may have to spend some time getting his quota raised, finding alternative sources of supply, or substituting other raw materials. The restrictive practices of international cartels might be cited for illustration, and collusive monopolies in this country have also restricted sales, though perhaps less openly. When Seller B does not have enough merchandise in stock to meet the increased demands of Buyer A who wishes to expand output, B must expand his own production before A can be supplied (unless the buyer can purchase elsewhere). This requires more time, but introduces no new problems in supply-elasticity, so we can return to the first firm, A, and continue with the

analysis by considering the problems he may meet when trying to expand output.

Additional labor, for instance, usually can be obtained in either of two ways: (1) by hiring new workers or (2) by inducing old employees to work overtime. The latter is the quicker of the two methods though often the more expensive. If no additional workers can be recruited locally, delay may be caused by having to bring in workers from more distant localities. If they are hired away from rival firms, the cost rises sharply, since an increase in wages usually must be shared with veteran employees too. Sometimes closed shop agreements with closed unions impede expansion of output. The long apprenticeship period required by certain unions furnishes another instance where the supply of a factor cannot be increased rapidly.

Enough has been said to indicate that the responsiveness of supply to an increase in demand has different time dimensions under different circumstances even when the problem is merely one of securing additional variable factors of production. Usually the quicker an expansion is effected, the more it will cost the expanding firm and therefore the larger must be the expected profit inducement. The increase in price or rate of sale needed to bring forth a certain increase in rate

TIME-ELASTICITY OF SUPPLY WHEN PRICE INCREASES

	Base Price and Increase	Quantity Produced After	
		One Month	Two Months
Original	50¢	1,000	1,000 Normal rate
Increased to	60¢	1,200	1,400
Increased to	70¢	1,500	1,800

TIME-ELASTICITY OF SUPPLY WHEN RATE OF SALES INCREASES

	Base Rate of Sales and Increase (Price is 50¢)	Quantity Produced After	
		One Month	Two Months
Normal	1,000	1,000	1,000 No inventory change
Increased to	1,200	1,100	1,200 Inventory down, then constant
Increased to	1,500	1,250	1,400 Inventory down even after two months

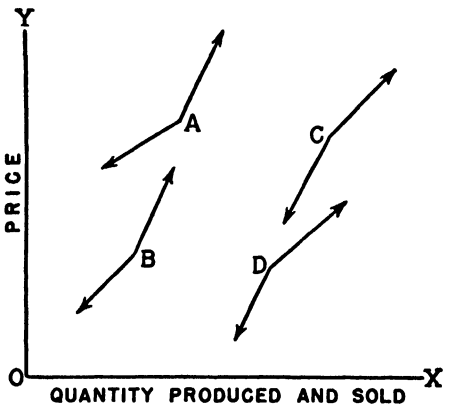
of production cannot be predicted without a close scrutiny of the particular firm or industry being analyzed and of the economic environment in which that industry operates. In the "moderately long-run" time period under consideration hypothetical supply schedules should specify in some way the length of time presumed to elapse between the causative change in demand and the resultant change in supply.

These two hypothetical supply schedules show only two of many possible situations. The second schedule assumes that the price remains constant at 50 cents throughout, which would be likely to happen only if unit costs did not rise appreciably (as when 1,500 units are not more than capacity operation and unit factor prices do not increase with the firm's necessary expansion in demand for them). Neither schedule looks far enough ahead to answer questions about probable entry of competitors and possible "overproduction" which, particularly in the first case, might force the price down in some future month not shown here.

5. Directional Supply-Elasticity.—Finally, we should note that these are one-directional schedules; they tell what might happen to supply if the demand for the product were to rise. That is, if the price having once risen to 70 cents, should decline to 60 cents and then to 50 cents, there is little likelihood that the supply would shrink at the same time rate at which it expanded. It usually takes much less time to lay off workers than to recruit them, to decrease orders for materials than to increase them. Exceptions may be found in the case of long-term contracts which cannot be cancelled quickly. And if new firms enter an industry under the lure of large profits, they leave it much more slowly when prices and profits fall.

FIGURE 52

MODERATELY LONG-RUN SUPPLY CURVES MOVING FROM GIVEN PRICE IN GIVEN TIME



Therefore, the above schedules cannot be drawn into curves unless arrows are attached to indicate the direction of change in the inde-

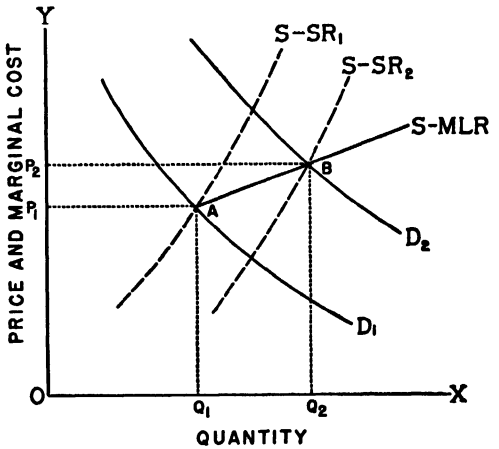
pendent variable (price or rate of sales). If such curves are to be used at all in graphic explanation of how supply fluctuates, they must all start from an assumed given price and move either up or down from it. This will give a "broken back" curve something like those in the accompanying diagram, but there will have to be a different curve for each hypothetical starting price (or rate of sale). Curves *A* and *B* of Figure 52 are probably the most common type for manufactured raw materials, i.e., less elastic above starting price than below it, while curves *C* and *D* reveal agricultural commodities (if the time period is extended to at least one year).

6. Two Types of Moderately Long-Run Supply Curves: Industry Marginal Cost Curve or Successive Price Points.—The moderately long-run supply curve may be of two types. The first is the collective marginal cost curve explained at the beginning of Section 4.

FIGURE 53

MODERATELY LONG-RUN SUPPLY CURVE

I. The Equilibrium Type Derived from the Marginal Costs of an Industry or a Representative Firm



The second is the line connecting two short-run price-quantity points. The first shows what would happen if the market price were always the "equilibrium price" under pure competition as described in Chapter 7, Section 11. The second is more realistic. It may be drawn for any type of market supply, whether hypothetical pure competition or any one of the many kinds of monopolistic competition. It should be distinguished, however, from a line connecting two price-quantity situations in the short-run time period or in the very long

run. That is, the quantity increase from the first price to the second comes from new production, not from stock as in the short run, nor from increased capacity as in the very long run.

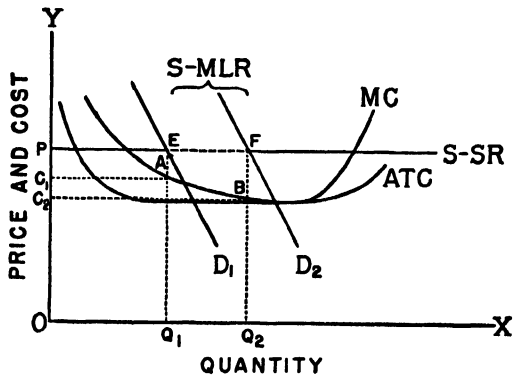
The *MLR* supply curve based on marginal cost will slope upward to the right. This is true whether it depicts the collective marginal costs of a group of firms or Marshall's "representative firm." One may say also that this type of curve will not be as steep as an *SR* curve for the same industry. The *MLR* supply curve is more elastic than the *SR* curves (Figure 53).

The second type of *MLR* supply curve may slope upward, downward, or move horizontally. The most common situation is that of a seller offering his goods at a fixed price. His *SR* supply curve is horizontal at that price as in Figure 54. His *MLR* supply curve may be identical with it for moderately increases in the rate of sales. This is the case with most manufacturers and retailers. In addition to

FIGURE 54

MODERATELY LONG-RUN SUPPLY CURVE

II. Derived from Successive Short-Run Points When Both *SR* and *MLR* Prices Are Constant



their fixed price policy which underlies this argument, most of them have constant or nearly constant marginal costs. The following diagram illustrates the situation.

Figure 54 may be interpreted as follows. Demand increases from D_1 to D_2. (Both of these demand curves are *AR* curves. No *MR* curve is shown because there is no need to find a hypothetical maximum profit price, as was the case in Figure 33.) Sales at price P rise from Q_1 to Q_2, determined by points E and F. The output is adjusted to the rate of sales. Therefore, average total cost declines

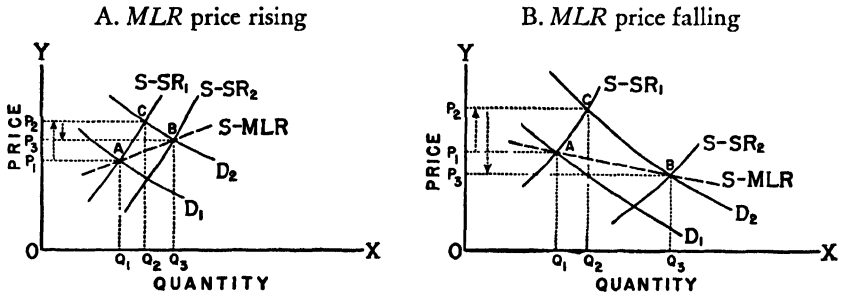
from C_1 to C_2 determined by points A and B. Marginal cost may be ignored since it does not constitute a supply curve nor here influence price.

A second possibility considers that the price does change as output expands. This is most common with goods which are relatively homogeneous, such as agricultural products and industrial raw materials. The price may rise or fall, as shown in the two parts of Figure 55. One must note carefully that Figure 55A is not the same as

FIGURE 55

MODERATELY LONG-RUN SUPPLY CURVE

III. Derived from Successive Short-Run Points When *S-SR* Rising



other figures which it resembles. In Figure 53, for instance, the *MLR* supply curve is drawn as a solid line to indicate that it determines both points A and B. In Figure 55A, the *SR* supply curves are decisive and the *MLR* curve is dotted to indicate that it is derived from A and B. Similarly, Figure 55 should not be confused with very long-run supply curves of increasing or decreasing cost. The latter could be drawn in the same way, but the line connecting A and B would be given in advance as equilibrium points on a curve of increasing or decreasing average total cost for a representative firm whose capacity was increased. This will be explained below in Sections 7 and 8.

A diagram like Figure 55 may also be used to explain why prices may rise in the short run under the impact of increased demand, but fall in the long run. The short-run price rise from P_1 to P_2 stimulates output expansion in the moderately long run. This causes a new short-run supply curve to appear at $S\text{-}SR_2$. This curve intersects D_2 at B instead of C and the price falls accordingly to P_3. The degree to which P_3 falls below P_2 may be slight, as in 55A, or large

as in 55B. An intermediate position is also possible in which P_3 is below P_2 but is equal to P_1 (not shown in Figure 55).

7. Very Long-Run Supply Curves of Representative Firms.— In certain situations a third period of supply time may be distinguished in relation to capacity changes of representative firms. This is called variously "very long run," "time period III," or simply "long run," and is defined as time long enough to increase the quantity of the "fixed factors" employed, i.e., the factors which are considered as fixed in the "moderately long run." All factors become variable, strictly speaking. The vagueness of this definition is apparent, and the borderline between the two long-run time periods and the two groups of variables is extremely hazy. However, it is reasonable to make a general distinction between direct labor and materials on the one hand and machines, floor space, land, or executive personnel on the other.

The usual reason for additional investment in the latter is that the intensity of demand is great enough to push output beyond "capacity," i.e., past the point of lowest average total unit cost (see Chapter 9, Section 13). Further expansion may be profitable even in the face of rising costs if the price exceeds average cost at that point, but before long the existing facilities will be overtaxed and marginal costs will rise sharply. If output is driven that far, costs can be reduced by increasing some of the "fixed factors." The farmer will buy or rent more land; the manufacturer will get more machinery or floor space; the middleman will get more storage or display room. In each case the effect will be to restore more efficient relationships between the factors of production and thus to lower the unit cost of production. This is the significant result, and the fact that such additions usually take a longer time than is required to get more labor and materials is merely incidental. When a firm is operating at less than capacity and an increase in demand causes it to expand output toward but not beyond capacity, there never will be any "very long-run" situation no matter how long one may wait for supply adjustment. Where capacity is reached and exceeded, the changes described as occurring in the "very long run" follow those associated with the "moderately long run."

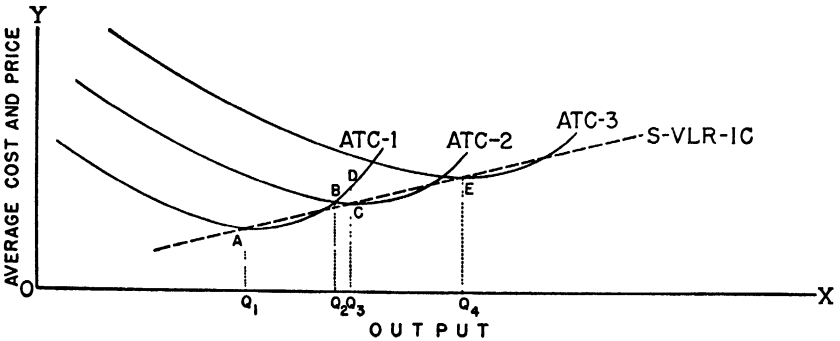
The time required for making additions to fixed capital depends upon the nature of those additions and the supply conditions prevailing in the market. Orders for additional machinery may be met from supplier's stocks on hand or the machinery may have to be built, particularly if it is of special design. New floor space may be rented

or purchased if available, or constructed if not, provided that construction materials and labor can be obtained. Land may also be purchased or rented. Sometimes it is logical to think of writing a contract with a new executive or research expert as resembling the addition of a "fixed factor" or making a "very long-run" investment. Such men are not always readily obtainable.

It should also be noted that the motives and occasions for securing additional fixed factors are not confined to demand increases which

FIGURE 56

VERY LONG-RUN SUPPLY CURVE OF INCREASING COST DERIVED FROM NADIRS OF SUCCESSIVE *ATC* CURVES



already have raised output beyond capacity. Purchases may be made in anticipation of such an eventuality either because it is thought impending in the near future or merely because there happens to occur a "good opportunity" to obtain land, or a building, or an executive. Sometimes the expansion will be initiated as a method of increasing sales, rather than as a result of such increases. This occurs when the management of a firm decides to seek lower costs through production on a larger scale, with the ultimate objective of expanding sales through the lower prices made possible by the reduced costs. Such is the traditional history of the Ford Motor Company in the days of the "Model T."

The degree to which costs may be lowered by an expansion of fixed assets will of course vary from firm to firm and industry to industry. Three possibilities are frequently discussed in economic literature: higher, lower, or constant average total costs. All *comparisons are made at successive capacity operating rates,* i.e., at the lowest possible average total cost under the larger capacity figure as compared with the smaller. These have been described by different phrases such

as "increasing cost," "decreasing cost," and "constant cost," or "diminishing returns," "increasing returns," and "constant returns." In the case of "increasing cost" or "diminishing returns," the advantage of expansion lies in the fact that production with the smaller capacity has been pushed so far past the point of lowest average total cost that the unit costs prevailing before expansion are higher than those prevailing afterwards, even though the *ATC* at capacity in the second situation is higher than in the first. This usually is shown by a diagram of successive *ATC* curves such as in Figure 56.

When (in the moderately long-run time period) production expands beyond the capacity operation quantity Q_1 of the original *ATC*-1 curve, unit costs begin to rise. When they exceed the cost (Q_2B) shown at output Q_2, it becomes profitable to expand facilities (in the very long run) so that a new unit cost curve is obtained like *ATC*-2.[3] The capacity operation point, *C*, on this curve has a higher *ATC* than the capacity operation point, *A*, on *ATC*-1. But Q_3C is much less than the cost, Q_3D, which would have been incurred if the expanded use of factors had been limited to those associated with the moderately long run instead of the expanded (durable) facilities usually associated with very long-run changes. Similar sequences of *ATC* curves might be used to derive *S-VLR* (Supply—Very Long Run) curves of constant cost and of decreasing cost (see Figure 57 below and also Figure 55A above).

A few cautions must be given about the nature and usefulness of very long-run supply curves. In the first place, these curves are not reversible; they must be read always from left to right and never backward. A contracting demand rarely will cause disposal of portions of the fixed plant, and even in the occasional cases where this does happen, the contraction will follow a different pattern from the expansion. In the second place, the successive *ATC* curves must be considered either prototypes of all the firms in the field (to have them all the same size would be most logical) or composite curves for the industry as a whole. This assumption also requires that "normal profits" be included as a cost of production and, of course, that there be no restraint on competition among the firms involved. Thirdly, the *S-VLR* curves differ fundamentally from other supply curves in that the independent variable is not price, but rather is the scale of capacity production which itself changes because of the postulated increase in schedule demand. Enlarging output capacity brings either

[3] The further assumption of free entry of competitors is necessary, because otherwise rising costs could be offset by increased prices.

rising, falling, or constant points of minimum average total cost and these nadirs in turn establish the trend of normal price. Fourthly, a rising *S-VLR-IC* curve is not the same as a rising *S-MLR* curve. The rising costs of the latter usually are due to a more intensive use of a given fixed factor. Inferior grades of the factor may be used in some firms and superior grades in others, but in each case the quantity used must be held constant. The rising costs of the very long-run supply curves, on the other hand, are due to rising minimum *ATC* points on successive moderately long-run supply curves when the scale of production is increased by expanding a fixed factor as well as the variable factors.

Finally, if a very long time actually does elapse between the successive changes in the production capacity of typical firms, then *other changes* are likely to occur which also will affect costs. For instance, soil may become exhausted or eroded, mineral deposits may become depleted, stands of virgin timber may be cut, etc. The cost of labor or industrial materials may rise so that the chief factor causing higher *ATC* points is a rise in the general price level, not a change in scale of production. On the other hand, improvements in technology may so reduce the cost of production as to offset all these forces and the price of goods in "increasing cost" industries may decline historically. This argument about the influence of historical change also applies in situations where decreasing costs otherwise would exist. Expanding the scale of production reduces unit costs, but this reduction is more than offset by the rise in such costs as those enumerated in the preceding paragraph. As a result the historical trend in prices is upwards, not downwards. The long-term trend of product prices in an industry is never sufficient evidence to establish with certainty whether that industry produces under increasing, decreasing, or constant costs. These classic categories belong to a static world of arbitrary assumptions, not to the dynamic world of real life where other cost changes may occur to offset them.

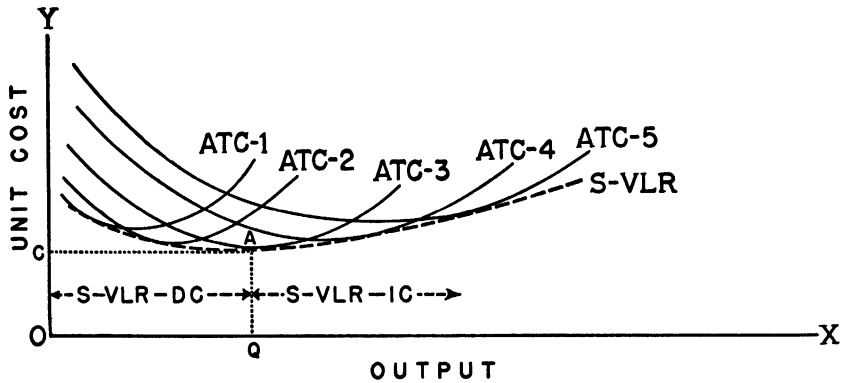
8. Very Long-Run Supply Curves as Planning Curves.—A curve comparable to the very long-run supply curve of decreasing cost has been used for other purposes than historical analysis or prediction. Some have used it for a "planning curve" to indicate how an enterpriser might reach a decision in calculating how big a plant to build. Obviously he would want to build the one which would yield him the most profit. If he contemplates producing a standardized product under conditions of pure competition, which does not happen

very often in modern times, he will want the lowest possible *ATC*. This can be determined if he knows the *ATC* curves of successive plant capacities as shown for a hypothetical instance in Figure 57 where *ATC*-3 is the *MLR* curve with the lowest nadir.

The decline in the *S-VLR* curve up to output *Q* may be attributed to the economies of large-scale production, while beyond *Q* the diseconomies of bureaucratic management might increase sufficiently to

FIGURE 57

PLANNING THE OPTIMUM-SIZED PLANT BY USING A VERY LONG-RUN SUPPLY CURVE BASED ON A SUCCESSION OF *ATC* CURVES



overcome additional production economies and therefore the stage of increasing costs would be reached. Each successive *ATC* curve of the moderately long-run period begins at a higher point because of the increased investment in plant capacity.

A diagram of the kind shown in Figure 57 has an extremely limited applicability at best. Most enterprisers lack sufficient knowledge of probable costs to do more than guess at a sequence of *ATC* curves like those drawn above. The actual capacity of the plant often is frozen by the size of one important fixed factor which happens to be available in a certain amount, such as a factory site or a power unit.[4]

[4] If fixed factors could be obtained freely in any desired magnitude, the *ATC* curves of the *MLR* all would be so close together that their nadirs would merge and become identical with the *ATC* curve of the *VLR*. The latter curve, if then considered a supply curve in the very long run, would indicate that all factors were variable and that there was no lumpiness in any of them, i.e., that there were no "fixed" factors, but all were perfectly divisible. A curve could be drawn which would be marginal to such an *ATC* curve, but it would have little practical value, since some factors always are lumpy in actual experience and the most profitable output rate is calculated upon the basis of more or less intensive use of these factors, i.e., upon the *MC* curve of the moderately long run.

Furthermore, practically all manufactured products are sold under one or another type of monopolistic competition, not pure competition as implied in the above analysis. That is, the quantity sold is limited on the one hand by the price fixed by the seller, by his trade association, or by the leader he is following, and on the other hand by the efforts of his own sales force. Hence, the most desirable capacity of the planned production unit will be determined more by the estimated sales volume than by the estimated costs. This is also true for much planning in the field of wholesaling or retailing. Finally, most planning of optimum capacity is based upon questions of changes in present facilities. It does not start from scratch, and the disposal of existing plant often will be an important factor in the final decision.

9. The Apparent Inevitability of Monopoly Under Conditions of Decreasing Cost.—A second use of *S-VLR-DC* curves is in the problem of monopoly. Economists have argued that if the manager of a given firm knows that expanding plant capacity will enable the firm to reduce unit costs, he will be induced to undertake such expansion (assuming he can get the requisite capital funds). This will permit the firm to cut prices and to undersell competitors who do not expand as rapidly.[5] If the total quantity which buyers will purchase at a price equal to or slightly greater than *OC* in Figure 57 is less than *OQ*, then the first firm can achieve a substantive monopoly. Such is the case with most public utilities. If, however, buyers want considerably more than *OQ* at such prices, the expanding firm will experience increasing costs and firms with smaller capacity can meet its prices by equalling its costs. This explains why expanding demand in many industries has been followed by an increase in the number of firms rather than in the expansion of existing firms. The small newcomers can compete successfully with the larger old firms.

Several other obstacles to the achievement of monopoly through the economies of scale may be mentioned. One is the rise in shipping costs as additional customers are sought at greater distances and across possible tariff walls or other politico-economic barriers. There is also the ever-present possibility that the expanding firm will choose to be content with less than the whole market and will utilize its reduced costs as a method of expanding unit profit rather than cutting prices. Or the competing firms might get together at some stage in

[5] Note the inversion of cause and effect in this approach. Cost changes cause price changes instead of vice versa; and price-competition is antithetical to the cost-competition (pure competition) situation assumed to underlie the first *S-VLR* curves.

the game and agree upon price maintenance or a division of sales territory.

10. Summary of Possible Cost Reasons for the Downward Price Trend of a Commodity.—It generally is argued that the historical trend downward in the prices of ce ain commodities (such as automobiles) can be explained best in terms of the long-run decreasing costs of representative firms, as described in Section 7. Actually, there are many possible reasons for decreasing cost. A check list such as the following may be useful.

1. In the *moderately long run,* when the major factors are held fixed and there is an expanding demand for A:
   (a) Increasing physical returns (declining AVC) may occur as the use of more of the variable factors makes a more efficient combination of the factors.
   (b) The total cost of the fixed factors will be spread over an increasing number of units (declining AFC).
   (This group of causes ordinarily would not be relevant, since the declining ATC which is the joint result of these two is rarely the basis for decisions to decrease prices.)
2. In the *very long run,* when all factors are variable and there is an expanding demand for A:
   (a) Economies of large scale production may occur because of:
      (1) "Internal economies." These include all those reductions in cost which are made possible by increased division of labor, use of more and better machines, reduced waste, quantity purchase discounts, cheaper selling costs, etc.
      (2) "External economies." These include particularly those reductions in the cost of goods bought which result from their being produced in larger volume by a given industry to meet the expanding demands of the buying firm (or industry). Or satellite supply and service industries may develop.

   (These causes are the most significant ones underlying the downward sloping portion of the $S\text{-}VLR$ curves. They are usually less important for price trends over a period of several decades than No. 3 below.)
3. Efficiency rises because of inventions which reduce costs by reducing the total necessary quantity of labor, capital, equipment, or land.

(a) Improved tools or machines for the harnessing of mechanical power.
(b) New products, such as chemicals, more productive than those formerly available.
(c) New processes or ways of combining factors.

(Inventions have no necessary time pattern. They may occur at any time and be introduced at any time. If they represent inexpensive changes in the variable factors, they will be substituted in the *MLR*. If the inventions affect fixed equipment and are expensive, time must be allowed for accumulating the necessary funds (unless they can be borrowed). This may involve waiting until existing equipment is worn out and depreciation funds collected, which would indicate the appropriateness of association with the *VLR*. Again it should be stressed that *MLR* and *VLR*, although useful concepts, are incapable of precise definition.)

4. Decreased cost of buying factors of production, caused by:
   (a) All of the above possible reasons for the decreased cost of producing any good.
   (b) Other reasons for reduced factor prices than those included thus far in this cost-side approach (see Section 5, Chapter 16).
   (c) Increasing supply of factors arising from causes other than increased demand for them, such as extensive immigration forcing down the wage rate.
   (d) Declining demand for factors, such as that resulting from cyclical recession or a secular decrease in the price level.
5. Fortuitous improvement occurs in the quality of free goods essential to production. (Limited chiefly to agriculture where weather conditions are vitally important.)

In the above outline, points 1 and 2 assume that the initial cause of the decline in cost is an increasing demand for a certain product. The cost reductions mentioned under points 3, 4, and 5 require no such assumption, although certain demand changes are implicit in 4-d. These other possibilities complicate the problem of considering an *S-VLR* curve as one existing through a period of time. Any one or more of them may exert their influence to change the price-quantity relationship of such a curve before sufficient time has elapsed for small-scale fixed factors to be replaced by large-scale factors. To avoid this possibility by having recourse to *ceteris paribus* is to run the danger of forgetting forces of extreme importance in explaining history or predicting future developments. There is also the danger of confusing the production unit with the administrative unit, as in

business combinations which may bring together two or more units producing the same thing so as to increase total output without affecting the scale of production in the usual sense.

11. Summary of Possible Cost Reasons for the Upward Price Trend of a Commodity.—The increase in the price of certain articles over a period of years also can be explained in part by a cost-side approach. Many items in the following outline of causes of increasing cost, historically considered, parallel those for decreasing cost, but there are significant omissions and additions.

1. In the *moderately long run;* expanding demand for A:
   (a) Decreasing physical returns as output is expanded past the point of diminishing average returns.

   (Because of the inability to expand certain fixed factors, such as land sites, and the difficulties of entry, the *MLR* period may be prolonged considerably both in quantitative and temporal extent.)
2. In the *very long run;* expanding demand for A:
   (a) Internal diseconomies such as inflexibility, lack of employee interest, complex rule-books, etc.
   (b) External diseconomies resulting from the internal diseconomies forced on others by expansion of A.

   (This group of causes is not as significant as under decreasing cost, and is generally less important than point 1, above.)
3. Efficiency falls, because
   (a) Workers' morale declines.
   (b) Machines wear out.
   (c) Management becomes less capable or energetic.
4. Increasing cost of factors used in making A:
   (a) Diseconomies of suppliers when production expands under stimulus of other demand than that from makers of A.
   (b) Increased total demand for factors rigidly fixed in supply.
   (c) Decreasing supply of certain factors, such as exhaustible natural resources.
   (d) Increased bargaining strength of suppliers, such as union labor; or more competition among buyers.
   (e) Other non-cost reasons for increasing prices, particularly cyclical booms and secular uptrends in prices, tariff changes, or property taxes.
5. Fortuitous decline occurs in the quality of free goods (poor weather) or of certain raw materials forcing substitution of inferior ingredients, as in wartime.

12. Summary of Economic Principles Governing Cost Changes Which Influence Price.—When time is allowed for supply and cost to become adjusted to those changes in price which result from an increase in schedule demand, the basic price determinants alter. Present price becomes a long-run determinant of future price. The first change is a rise in market price. This permits an increase in marginal cost in the moderately long run. This may be followed by an expansion in the production capacity of most of the firms in the industry so that the nadirs of their average total cost curves change. Whether these nadirs rise, fall, or remain constant depends upon the balance between the economies and diseconomies, both internal and external, of increasing plant capacity. The argument is presented best through using the concept of a representative firm and presupposing pure competition. Very long-run cost and supply curves must be read from left to right; they are not reversible. The same is true, although less significantly, for moderately long-run supply curves.

During the time which must elapse for the foregoing adjustments to occur, other cost-influencing changes are almost inevitable. Therefore, the upward or downward trend of individual prices should be described as a function also of such things as changed prices of the factors used, inventions, etc. Some variables push prices upward, some downwards, and the net result may obscure the fact that opposing forces were at work. Occasionally the forces are balanced so evenly that constant cost may occur for a while and the price may remain relatively stable.

Chapter 14

INDIVIDUAL PRICE FLUCTUATIONS

1. Introduction.—The previous chapter considered chiefly long-run trends in prices, particularly as they are based upon changes in cost. This involved an explanation of the manner in which increases in schedule demand alter the average total cost curve in the very long run. Intermediate price changes also were considered, but the major emphasis was upon the trend of normal prices in the face of rising demand under conditions of pure competition. The present chapter is not so narrow in its focus. It includes a discussion of both short-run and long-run price changes in individual prices under conditions of declining demand. Institutional changes and monopolistic situations are considered. Both upward and downward deviations from normal price are analyzed and their implications examined. But the basic question remains the same, "How can individual price fluctuations be explained in the economy in which we live?" The answer again is approached chiefly from the supply side.

Specific questions follow:

1. What is the effect of falling demand upon price?
   - A. When sellers can adjust output but must accept price?
   - B. When sellers can adjust both output and price?
   - C. When firm exit occurs?
2. Why do current prices of commodities deviate from normal prices?
   - A. When prices rise?
   - B. When prices fall?
   - C. When supply changes suddenly?
3. What diagrams show the different supply-reaction situations?

2. Effects of Falling Demand When Sellers Can Adjust Output but Must Accept Price.—Most of the discussion of the time-elasticity of supply in Chapter 13 was based upon the assumption that the initiating force was an increase in schedule demand manifest in a heightened price. But what are the probable developments if demand falls, and why? The answers can be given best by considering separately the supply conditions likely to exist under price acceptance and those under price adjustment.

Assume first that the manager of the individual producing unit has to accept the going market price and that he has been operating at his most profitable output under that price. Then he will make more profit by reducing output as the price falls than by continuing to produce at the former rate. If he knows his cost schedules, he can calculate the most profitable output for any price by finding the rate at which marginal cost equals market price. This is logically true with either a moderately long-run or a very long-run *MC* curve. But this apparently simple solution in formal analysis requires careful examination and qualification.

The marginal cost curve that prevails in a period of contraction may be based on certain variables which do not appear at the same stage in the expansion of production or change at a different rate. Instead of being concerned with additional costs, the manager has to calculate "subtractional" costs. Some costs must be incurred anyway, others may be reduced or eliminated by decreasing output. These two groups usually are called fixed costs and variable costs, but the content of each category will differ with the direction of the change in output, not to mention also the magnitude of the change and the time allowed for adjustments to occur. (Refer also to Chapter 13, Section 7.)

If output is to be reduced, obviously there will be reductions in outlays for labor and materials. There is, however, the possibility that certain labor contracts cannot be broken and wages must be paid even if the employee is not working. Such costs are not subtractional unless the demand remains shrunken until after the contracts expire. Somewhat the same situation is faced when managers are reluctant to drop skilled and experienced workers who have no contract claim, but whose replacement would be very difficult when expansion again became desirable. Materials may be purchased on contracts which still have weeks or months to run and in addition sizable inventories may have been accumulated at the plant. In the latter case the only saving from not using materials already purchased would be the salvage value of the materials when sold to some one else. A third type of subtractional cost would be the maintenance and upkeep on plant and equipment which is saved when these are not paid. Fourth, wear and tear depreciates durable capital goods even when maintenance and repair expenditures are made as they become necessary. This is physical depreciation as opposed to obsolescence. The latter will occur whether the machinery is being used or not. Fifth, in some cases a royalty payment may be avoided or reduced by curtailing operations.

When all of these subtractional costs which are appropriate in a given case are deducted from total cost at the going rate, a schedule of total costs at lesser rates may be constructed and a directional marginal cost schedule or curve may be derived therefrom for the moderately long-run period. This schedule is the one which should guide a manager who calculates his costs accurately and carefully so that he may quickly adjust his output to changes in demand as shown by variations in the price. A manager of this type probably would be hard to find, particularly in agriculture, where conditions of pure competition are approached most often. In addition to the difficulty of rapid and precise adjustment in output to the figure which the schedules or curves indicate would maximize profit, there is also the problem of price expectations discussed above in Section 6 of Chapter 10. A mere fall in price usually is not enough to provoke a changed output, but rather an expectation of a future price low enough for a time sufficiently long to warrant the trouble and inconvenience of seeking to save money by curtailing production.

Other conditions which may retard or discourage output reduction in the face of lower prices include (1) the possibility that falling profits may stimulate inventive efforts and produce successful cost reductions which were not forthcoming when prompted by the mere hope for gain. (2) Reduced prices may force out marginal firms which cannot stand losses or which prefer to shift to some other line. This eases the pressure on the remaining firms by reducing total supply and thus lessens the price decline. (3) If the owner-manager can substitute more of his own labor for that formerly hired, he can reduce out-of-pocket cost without reducing output, although this possibility is significant only in small enterprises. (4) Some small-scale producers also have tried to offset falling prices by increasing their own efforts so as to get *greater* total output, hoping that total revenue will not diminish even though prices have fallen. (5) If the owner-manager has no better alternative use of his time and resources (very often true in farming), he may continue to plant as much or make as much as before, in the hope that the future will not be as black as present prospects may indicate.

3. Effects of Falling Demand When Sellers Can Adjust Both Output and Price.—When a seller sets his own price and demand falls subsequently, he will experience a decline in his rate of sales. He ordinarily will curtail production as a result, rather than to continue and to accumulate inventory, but his decision again will be

affected by some of the exceptions noted in the preceding section such as long-time contracts for the purchase of labor or materials, large raw material stock piles, ignorance of precise costs, and future sales expectations.

The basic questions are, "Will he cut his price? And by what guide?" The answer must be that those who set their own prices generally do not change them as often in responses to changes in demand as price acceptors change their output. The $MC = MR$ equilibrium formula may be applied if data are known for any given demand and cost schedules, but when demand schedules shift, prices usually are not altered until demand has changed considerably. This is particularly true for reduced demand where the diminished output which results is often produced at higher unit costs than before and cost data have to be understood pretty well before prices will be cut "to reduce losses." They are more apt to be increased. As a matter of fact, there is usually great uncertainty about the exact elasticity of demand schedules, particularly new ones, and the tendency among price setters is to assume greater inelasticity than actually exists. Price setting is always something of an experiment, and the temperament of the manager may have more to do with price policy in the face of declining demand than any analysis of maximum profit positions.

In the case where the seller fixes his price in concert with other sellers, additional deterrents to price-cutting are present when sales diminish. If the decline in the sales of one firm is the result of its failure to meet sales competition by other firms, the weaker firm obviously will not succeed in getting the others to agree to a general cut in prices. It may break away from the group and try price-cutting independently, but is not likely to do so until all other avenues of relief have been exhausted. The fear of retaliatory action and ostracism also may prove an important deterrent. If, on the other hand, the entire group is suffering a decline in demand for its product, the price is more apt to break when some major firm seeks to get the jump on its competitors, but even so price agreements usually increase price rigidity and delay price-cutting.

Since one purpose of price setting is to avoid those fluctuations in price which result from changing demand when goods are offered for sale regardless of price, there is no necessary connection between demand variations and market price. Therefore, the comments about the best relation between demand, sales, and cost, which received attention in the analysis of the effects of shrinking demand under conditions of pure competition, are largely irrelevant here. However, a

distinction should be made between those situations in which there is active price competition and those in which there is not. In the former case a decline in demand often is followed rather quickly by a price cut, but not in the latter case where tacit or implicit price agreements are the rule.

Another factor which influences the rapidity with which monopolistic price adjustments are made in response to a fall in demand is the margin of price above average total cost when that total includes all explicit and imputed outlays except those to the owner for his services of labor, land, or capital. If this margin is large at any given volume of sales, a considerable drop in sales may occur without causing the increased overhead costs per unit to rise sufficiently to wipe out the profit margin entirely. The immediate loss from price reductions plus the difficulties of raising prices in the future usually discourages even those price cuts which can be demonstrated to be beneficial. The history of many large firms during the 1929–1939 depression reveals numerous instances of price rigidity in the face of considerable declines in sales. Most of these firms made money throughout the depression or in all except one or two years, thus indicating the probability that prices had been set so as to yield a very large margin of profit in normal times.[1]

4. Effects of Falling Demand on Price Through Firm Exit.— Under pure competition the entry or exit of a single firm cannot influence the price, but if enough firms are driven out of producing commodity A by a fall in the demand for it, an appreciable reduction will occur in schedule supply via exits and the price may rise again. Or the rate of price decline may be retarded, assuming that demand continues to fall, as in the case of goods rendered technologically obsolete by a new invention (cf. natural camphor giving way to synthetic). Under monopolistic competition the exit of a single firm may or may not affect the price, depending for instance upon the degree of price competition that exists, the number of firms in the competing group, the cost curves of the survivors, and the competitive distance between the differentiated good produced by the withdrawing firm and the substitutes offered by rivals (cf. electric refrigerators displacing iceboxes).

Even though the impact of falling demand upon price via firm exit is not subject to generalization in any sweeping fashion, the relation between demand and firm exit is worthy of review. Since falling de-

[1] The depression also fostered economies which undoubtedly helped to maintain profits by holding down the rise in *ATC* when sales declined, but these are not sufficient explanation.

mand implies a decline in either price or sales, it also may be linked with those reductions which come through price competition, thus combining the analysis of conditions under the two general types: pure competition and monopolistic competition. Firm exit may be either voluntary or involuntary. It may involve either temporary or permanent cessation of production. Voluntary exit may occur for a number of different reasons including prospective bankruptcy, discouragement, physical incapacity of owner-manager, or visions of greener pastures elsewhere. Involuntary exit is the result of inability to meet demands of creditors, who ask the court to intervene and appoint a receiver to protect their interests. This may lead to reorganization of the firm with a scaling-down of fixed charges or to a dissolution of the enterprise and sale of its assets, usually in segments. Only in the latter case does genuine exit exist in the sense that a curtailment of supply is an almost inevitable outcome.

A firm facing a declining demand or engaging in price competition may continue in business temporarily even though its selling price does not cover its average total costs. It will postpone paying any costs that it can, such as by reducing the outlay on maintenance or stalling off demands of creditors. Individual circumstances will determine the payments which can be postponed most easily, but bondholders, landlords, and even tax-collectors are frequent victims. The firm will, of course, do its best to introduce cost-reducing innovations wherever possible, even to the extent of bargaining more severely with workers and with others from whom it buys goods and services.

It is sometimes said superficially that if the price falls below the average variable cost,[2] the firm will find it unprofitable to continue operations, but this does not constitute a satisfactory theory of firm exit. In addition to the numerous possible reasons for exit which already have been given, there should now be added the absence of financial resources sufficient to sustain operations during a period of producing at a loss. It matters little whether the losses are measured by a price which is lower than *ATC* or lower than *AVC;* the latter is merely an extreme case of the former. The important thing is the size of the losses in relation to liquid assets or the firm's ability to borrow. The amount of loss is also a function of the total sales made at prices below total cost, i.e., of the length of time during which such sales are made. It is often the prospect of a change for the better

[2] This would include all costs which could be avoided by closing down divided by the total output at the expected rate of sales, both costs and sales being figured with reference to an identical time period such as a week or a month.

which keeps a firm going, or despair about future losses which causes one to quit.

Finally, a firm may shut down on what it hopes is a temporary basis in order to reduce its losses while waiting for that upturn in demand which would make operations profitable again. Such action may be induced by a combination of operating losses, short-run pessimism combined with long-run optimism, and a minimum of selling costs. The last point refers to the ease with which sales outlets and customer acceptance could be regained after a shutdown. The selling problem also confronted many firms after the war, during which they were forced to shut down or convert from one product to another because of a shortage of essential materials for customary peacetime production.

One might add that often when a manufacturer or other primary producer shuts down, the flow of goods to the ultimate consumer is not stopped immediately. Other producers may increase their output of an identical or similar good. Even if this does not occur, the inventories of finished products which the manufacturer possesses or the stock in the hands of wholesalers or retailers may still be sold to supply consumer needs for weeks or months thereafter. The time required to drain these lower reservoirs in the stream flow of goods often delays the time considerably when the flow from the primary source becomes necessary again.

5. Upward Deviations from Normal Price.—Even though expansion and contraction plus entry and exit of business firms tend to adjust supply so that transactions prices generally do not deviate very much from hypothetical normal prices, minor deviations exist most of the time and major departures are frequent in some lines. Rising demand may spread overhead costs and lower *ATC* further below the asking price than it would have been at the output contemplated when that price was established. Occasionally monopolists take advantage of buyers' needs and raise their asking prices even though production costs are falling, or at least have not risen proportionately. (Cf. Chapter 11, Section 13.) They may be able to do this successfully because at the same time demand may have become more inelastic or may have shifted upward.

In other cases of rising profit margins there is no artificial obstacle to entry and expansion, but merely the fact that technically these require time. Shifts in schedule demand occur more rapidly than supply can be readjusted. Competitive bidding may force the price up above cost, or competitive offering may push it down below cost.

There is a short "time-shift ability" of demand upward and a long "time-elasticity" of supply. As was explained in Chapter 13, different industries require different lengths of time to adjust their output to the most profitable rate following any given percentage change in prices. If this price change has been away from normal price, the supply adjustment usually will be such as to cause price to move in the opposite direction, i.e., back towards normal price. The adjusting expansion or contraction in supply may occur in either the moderately long run as explained in Sections 4 to 6 of Chapter 13 or in the very long run as described in Section 7. No absolute measure of time-elasticity of supply is possible, but the conditions which tend to make it long or short deserve comment.

In agriculture, there have been more violent swings in the price of crops, like deciduous fruits, which take a long time to mature after planting, than in the prices of grain and other annuals, although the figures are difficult to interpret because of variations in the weather and the incidence of plant diseases or harmful insects. The time-elasticity of supply expansion is also longer for certain animals like beef cattle than for others like rabbits. In mining, certain deposits may be exploited more quickly than others because of the nature of the mineral deposit, and therefore the price need not go so high nor stay up so long to bring about the supply expansion which will tend to pull it down again. In manufacturing, the more expensive goods are usually those which require the most labor to produce and therefore take the most time. Objects in this class include heavy machinery, factories, railroads, steamships, office buildings, bridges, and canals. In the field of consumer goods the best illustration is homes. The long time technically required to expand the output of some goods also tends to discourage starting to expand, because of the inevitable uncertainty about whether the price will remain high until the goods from the expansion program reach the market.[8] The result is further to increase the amplitude and duration of price upswings for goods having a long time-elasticity of supply.

6. Downward Deviations from Normal Price: The Influence of Durability.—Current prices also frequently fall below normal. How far they fall and how long this downward deviation occurs depends chiefly on the article's durability. Ease of repair and of transfer to other uses are additional determinants.

[8] The significance of expectations in production planning also has been discussed above, particularly in Chapter 10, Section 6.

Most of the goods whose expansion of supply takes a relatively long time also are durable in the sense that they take a relatively long time to wear out in use. As compared to food which is a "single-use" good, those other products are "repeated-use" goods. They have a long time-elasticity of service supply when demand is falling as well as when it is rising, but the time required for a given percentage change in supply following a given percentage change in price is likely to be different going downward from that going upward. Since the demand for a repeated-use good is for the series of services which it can render, the supply of the good is really the supply of a series of such services. If extra care and maintenance can prolong the useful life of a machine, building, or other durable good, they increase the supply of the desired services and tend to offset the contraction which may result from the curtailment or the end of new production. As a result a decline in the demand for a durable good does not bring a quick shrinkage in the total supply of services from such goods even though the output of new goods drops rapidly.

It is further true that most durable production goods are sold under monopolistic conditions with a policy of price rigidity which sustains prices even though demand falls. Therefore, one cannot turn to the machine tool industry, for instance, to find an illustration of the slowness of supply quantity to shrink in the face of falling prices. But there is one durable goods industry where price competition and flexible prices prevail and can be used as an example even though the results are somewhat oblique. The construction industry has as its products homes, apartments, office buildings, and factories, which usually are sold as a unit together with the land on which they are erected. Therefore, even though it is the land which actually does most of the fluctuating in value, the declining demand for structures sometimes depresses the price of existing buildings to the place where new construction virtually stops. But the supply of (the services of) buildings does not diminish very much since they are so durable and there is no appreciable offset on the supply side to counteract the decline in demand and send the price back up again. In the terminology of this section, there is a long time-elasticity of supply contraction.

The extreme case of durability is that of land itself, particularly building sites, which have virtually perpetual life. Urban land offers the best illustration of extreme fluctuation in price because of long time-elasticity of supply both downward and upward. But it does not have a long-run normal price because it has no cost of production by which such a price may be established. (Cf. Chapter 10, Sec-

tion 13.) A partial substitute, and source of expansion on the upswing, is the cost of development which in some cases is considerable, as for draining, irrigating, levelling, filling, or creating means of ingress and egress via highways or railroads. Unless it has been inherited, land in its various forms probably has had a cost of acquisition or development for any given owner, and fluctuations sometimes are measured from such costs rather than from the normal price used for man-made goods.

A further impetus to extreme downswings in the case of durable goods may be found in their effect upon demand. It is in bad times that the demand for new machines, building, etc., falls off and it is also in bad times that owners try to make their existing equipment last longer. Hence replacements tend to be bunched in the upswing or prosperous period of the business cycle and diminish rapidly during the downswing. This irregular purchasing accentuates the fluctuations in price and/or sales which would occur anyway because the size and durability of many production goods cause long time-elasticity of supply both upward and downward, particularly the latter.

7. Unexpected Supply Shifts May Cause Price Fluctuation.— If we reverse the causation picture and consider cases in which the deviation in price results from an initial shift in supply plus a sluggish adjustment of demand, we find the major illustrations in agriculture. Here the vagaries of the weather may upset the best of production plans, whereas consumption habits are slow to change. Furthermore, farmers are notoriously given to ill-calculated expansions of planting following good price years and are only somewhat less thoughtless in contractions after poor ones. They do not clearly foresee the total effect of the combined judgments of thousands of competing individuals most of whom change in the same direction at the same time. Even with uniformly productive weather, alternate periods of high and low prices are therefore likely and are revealed in the history of some farm products such as meat animals and orchard fruits. Some manufacturers also have been guilty of similar practices, chiefly on the expansion side, but in recent years these mistakes in production planning have been curbed both by the diminishing number of competitors in some lines and by the emergence of strong trade associations able to gather needed information. The prevalence of price-fixing by manufacturers and merchandisers obscures the results of miscalculation in those fields while government intervention in support of farm prices is rapidly becoming an impediment to supply

contraction in agriculture, even though benefits are paid farmers for not planting.

8. A Summary of Supply-Side Causes of and Restraints on Individual Price Fluctuations.—Some deviation from normal price is more common than otherwise. The foregoing sections have sought to explain why divergencies from normal are not, in some cases, quickly overcome. Detailed explanations of fundamental shifts in demand were given in Chapter 2 and need only be summarized here by saying that they spring from changes in desire or purchasing power which either are induced by the producers of the commodity (as by advertising) or occur independently because of external forces. Analysis of cyclical fluctuations in prices and costs must be deferred to another chapter, where further comment will be made on individual price fluctuations under the question of why certain prices fluctuate in a pattern which differs from the general trend, but this is not the same as the current problem of deviations from normal price.

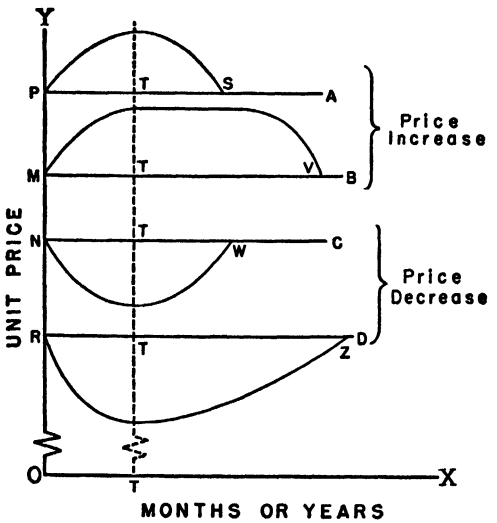
A few special cases deserve some brief concluding comment. First, when goods are made to order, their prices ordinarily do not depart very far from average total unit cost (including normal profit). However, if they take a long time to make (bridges and large buildings), unforeseen changes in the supply of labor or materials may develop or unforeseen production difficulties occur so that the cost may become much higher than was expected when the original contract was drawn. Second, the normal prices of secondhand goods can be only rough approximations based on cost new (or cost of reproduction) less the customary depreciation and plus cost of renovation. The possibility of self-use limits the price-depressing effect of declining market demand, but there is no ceiling except through the availability of substitutes or the presence of government price control. Third, bargaining is very important in determining the prices of articles which are heterogeneous and those whose normal price cannot be determined readily or to which the seller is indifferent. This includes land, buildings, secondhand goods, scrap, and articles, offered by speculative or necessitous sellers. In all such bargaining particular transactions prices may deviate widely from the average selling price or from a calculated norm of cost-less-depreciation. Ignorance of worth, impatience, or inability to discover better alternative bids or offers may all play a part.

9. Diagrammatics of the Time-Elasticity of Supply.—The general idea of variations in time-elasticity of supply among different commodities can be expressed with some simplification in a two-

dimensional diagram in which the horizontal axis represents elapsed time in weeks, months, or years, while the vertical axis represents unit price of the commodity being considered. In Figure 58 there are four comparisons, two on the up side and two on the down side. Each curve is drawn in relation to its own hypothetical normal price

FIGURE 58

TIME-ELASTICITY OF SUPPLY IN FOUR SITUATIONS



under the assumed conditions of constant cost in the very long run. (There is not intended to be any significance in the relative height of these very long-run horizontal cost curves.) In each pair of cases there is assumed to be some uniform price stimulus which is expected to continue and to which supply adjustments are therefore made. There is the further somewhat artificial assumption that all four prices reach their peak or nadir after the same interval of time. The upward time-elasticity of supply of commodity A is shorter than that of B (TS is less than TV). The flat top to B's curve is intended to suggest that there may be some delay before the new goods reach the market, but when that happens there will be a rather sharp drop to the old price. Speculative action as in farm products might hasten the decline. The decrease in price on the down side for C and D are drawn so as to be greater than the increases for A and B because supply contraction is generally slower than supply expansion. The time-elasticity of supply adjustment for D is made greater than for

C (*TZ* is greater than *TW*) to indicate a hypothetically greater durability of D.

Curves like those of A and C or B and D might follow one another in time sequence for a single commodity E, particularly if the cause of the price drop was supply overexpansion following the first price increase. In that case an initial price change caused by a change in demand might cause a series of "ripples" or oscillations around the normal price with diminishing amplitude until stability was achieved again. This result would occur, however, only if every other force affecting demand or supply were to remain constant while producers were figuring out the best output by a process of trial and error. A state of affairs of this kind is improbable in a dynamic world, and the argument about disappearing ripples is intended merely to state a tendency, not to describe a probability. (See Figure 59.)

Under some circumstances there would be no dampening effect, or even an antidampening effect so that there would be an increase in the amplitude of the ripples. Technically, the ripples will diminish if the demand schedule is more elastic than the *MLR* supply schedule. They will increase if the reverse is true, if producers learn nothing from experience, and if they allow the successive expansions and contractions to occur over the widening range provoked by the original *MLR* supply curve. These latter conditions seem improbable, but the price history of potatoes, corn, and cabbage in this country seems to indicate that many American farmers learn very little from experience. Their good fortune in certain years has been the result, not of wise planting and good weather, but of war's demands and nature's adversities elsewhere.

Diagrammatically, the price-quantity oscillations give rise to a chronological curve which spirals inward or outward. Against the background of a demand curve and an *MLR* supply curve the spiral looks like a cobweb. Hence the oscillation argument is sometimes called the cobweb theorem. It is an interesting oddity of infrequent importance except that the reasoning may help statisticians to derive static curves of demand and *MLR* supply (see Chapter 15).

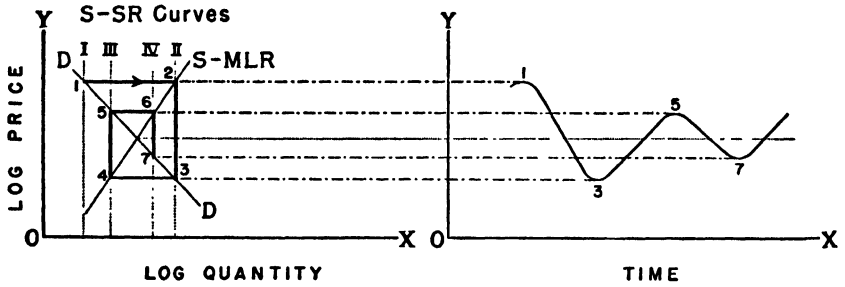
10. Summary of Economic Principles Dealing with Individual Price Fluctuations.—The major problems of individual price fluctuations may now be summarized from the arguments of this chapter and earlier ones. Price fluctuations may be considered as either changes from any prior price, or deviations from a "normal price." The latter has been the theme of Chapter 14. Demand aspects are slighted in the following outline.

FIGURE 59

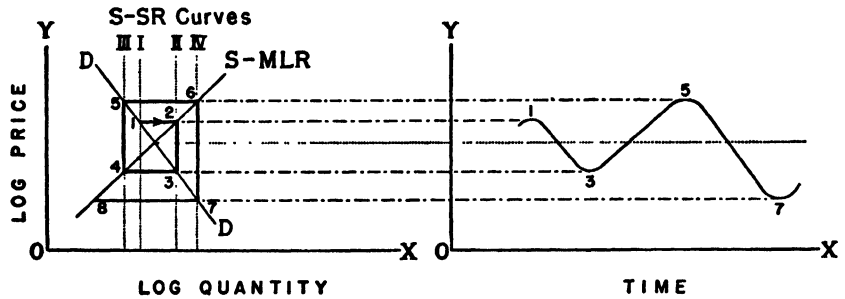
THE COBWEB THEOREM

STATIC CURVES (cobweb) — HISTORICAL CURVE (smoothed ripple)

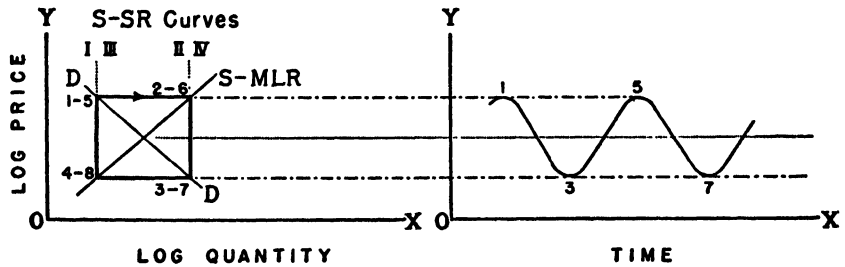
A. Demand More Elastic than *MLR* Supply



B. Demand Less Elastic than *MLR* Supply



C. Demand of the Same Elasticity as Supply



NOTE.—If price is established at point 1, the supply forthcoming in the next production period will be that shown by point 2 on $S\text{-}SR_{II}$, but this can be sold only at the price indicated by point 3, which in turn will evoke the smaller supply shown by $S\text{-}SR_{III}$ and establish the price at point 5, etc. Logarithmic axes are used so that the slopes of the straight-line curves will indicate their elasticities.

I. The concept of "normal price" involves:
  A. Two assumptions (except as noted):
    1. Entry and exit are not restricted by the government nor by collusive agreement.
    2. Expansion and contraction also are subject only to the gain-seeking decisions of the individuals involved.
  B. Two approaches:
    1. Demand: the "normal price" stimulates no change in demand intensity.
    2. Supply: the "normal price" stimulates no change in the rate of production and offering for sale. This requires the simultaneous presence of two equalities for each seller: $MC = MR$ and $ATC = AR$ (or Price).
      (a) The AR and MR schedules depend upon the magnitude and elasticity of demand. The latter is particularly important when competing firms produce differentiated products and set their own prices.
      (b) The ATC and MC schedules for individual firms depend upon
        (1) The prices of the factors, and
        (2) The efficiency of use of the factors, especially in relation to technology and the quantity of fixed factors. The latter influences the "capacity" of the representative firm, but is dependent also upon demand as explained in Chapter 13.

II. Deviations from normal price are a function of:
  A. The initiating changes in demand or supply.
    1. Their amplitude.
    2. Their duration.
  B. The time required for offsetting changes in demand or supply.
    1. For a supply change upward.
      (a) Reproducibility of product (time required to produce more of it if factors are available).
      (b) Availability of needed factors of production.
      (c) Ease of overcoming collusive or governmental restraints to expanded production.
      (d) Ease of entry of new firms.
    2. For a supply change downward.
      (a) Time required for completion of irreversible production processes already started (cf. crops).
      (b) Durability of products already sold and in use (also affects demand).

(c) Length of contracts for the purchase of factors.
(d) Ability to continue operations at a loss (liquid assets, borrowing capacity, etc.).
(e) Ease of exit.
    (1) Liquidating value of fixed assets.
    (2) Ease of shifting fixed assets or labor skills to alternative occupations.

3. For demand changes.
    (a) Alteration of desire patterns.
    (b) Changes in buying habits.
    (c) Changes in the availability or in the prices of other goods.
        (1) Substitute goods.
        (2) Complementary goods.

The major loose ends of this summary which must await further discussion are two: (1) the prices of the factors of production, and (2) changes in the total volume of spending by business, consumers, and government. The first of these topics will be treated in Part II on the prices of services. The second will be discussed incidentally in the chapters on interest and profits. A more complete treatment must be sought in books treating changes in the general price level rather than in individual prices.

Chapter 15

THE QUANTITATIVE APPROACH TO PRICE CHANGE

1. Principles of Quantitative Economics.—The principles of economics expounded in the preceding chapters have stated that certain demand and supply forces exert an influence upon price. If allowed to act independently, they tend to force the price upwards or downwards according to whether the relationships are direct or inverse. If opposing forces are present at the same time, and that is usually the case, they tend to offset one another so that the outcome is in doubt. This uncertainty, however, does not invalidate the principles of economics because, like all scientific laws, each is based upon the premise of *ceteris paribus* (see Chapter 1). Although this is a valid defense, the pure scientist in the modern world has difficulty in retaining his detachment. The economist in particular often is asked to explain the workings of very complex problems, to predict the future course of prices or employment, or what to do about a distressing situation, such as the low incomes of farmers.

To give an adequate answer to questions of this type, the economist must develop principles which explain what happens in the intricate maze of economic life. He must first go beyond simple single-variable functions. This has been the objective of earlier chapters which have shown that economic changes usually result from the action of several variables. The second step is to shift from a qualitative to a quantitative approach. Instead of being content with principles which answer questions about the *direction* of change in an effect which follows a given cause, the economist also must seek to explain *how much* the change will be.

The specific questions answered in this chapter include the following:

1. How can economics be made a science of price prediction?
2. What are the major independent variables causing changes in the price of a staple farm product like wool? How are their "elasticities" computed and related?
3. How do quantitative economists allow for shifts in schedule demand? the "time trend" factor.

4. What are the major difficulties in obtaining quantitative formulae for explaining price change?
5. What pitfalls should be avoided in applying these formulae to predict price change?

2. The Prediction of Economic Effects or Changes Is Desirable, but It Is Difficult When There Are Many Variables.— Both big business and big government plan for the future. If they cannot control it, they at least want to be able to adapt themselves to it. They face practical questions such as how much the domestic sales of coffee will decline if the price is raised five cents per pound when a subsidy ends or Brazil imposes an export duty. Or if our government should negotiate a 50 per cent reduction in the tariff on flaxseed from Argentina, how much would imports increase? How much of a drop in the interest rate on realty mortgages is necessary to stimulate the building of an extra 200,000 new homes next year? What is the best rate structure for an electric power company entering a certain backward rural area? These are problems of the type which confront the practicing economist. Each one is unique and must be studied separately with careful examination of parallel situations in the past. But the more economic principles, both qualitative and quantitative, the analyst knows, the quicker and the better will be his answers.

Complete accuracy of prediction is never possible because unforeseen circumstances usually intrude during the time interval which separates the prediction and the occurrence. However, the shorter the time perspective, the narrower the market, and the fewer the variables involved, the smaller will be the margin of probable error. The forecaster is aided by the force of habit and routine in human responses. If he is dealing with large groups of people, the unusual cases tend to offset one another so that the accuracy of predictions may be increased.

Economic statistics of production, prices, income, sales, foreign trade, bank deposits, etc. reflect the reactions of large groups of individuals over a period of time. Relationships between these time series may be discovered by the use of reasoning, mathematics, and statistical methods. Experience or "common sense" may indicate which data best reveal the demand or supply forces which are most likely to influence certain other data. That is, the dependent variable is affected by one or more independent variables, and if the latter can be discovered by observation and reasoning, much time can be saved from the laborious work of statistical correlation by concentrating first on those data which seem obviously related. The others

can be ignored unless the calculated curve proves to be a poor fit when compared with the curve of the original data, thus indicating that something of importance has been overlooked. This approach may indicate also the type of relationship between the variables, whether direct or inverse, simple multiples or squares, continuous or intermittent.[1]

3. **Dynamic Determinants of the Price of Wool.**—A recent study of the forces determining the price of fine wool in the United States furnishes a good illustration.[2] Elementary reasoning would suggest that the changes in the price of wool must be influenced on the demand side by changes in consumer income and on the supply side by changes in domestic production, domestic stocks, foreign production, foreign stocks, and the foreign exchange rate (since most of our wool is imported over a high tariff wall). Over a period of time, one would expect to find also some price-depressing effect from the trend toward substitution of rayon for woolen fabrics. Each supply and demand factor will have a different quantitative effect upon the price. All of these variables may be summarized by the verbalized equation:

$$\text{Price} = \left(\begin{matrix}\text{Consumer}\\ \textit{I}\text{ncome}\end{matrix}\right)^{a} \times \left(\begin{matrix}\textit{D}\text{omestic}\\ \text{Production}\end{matrix}\right)^{b} \times \left(\begin{matrix}\textit{F}\text{oreign}\\ \text{Production}\end{matrix}\right)^{c} \times$$

$$\left(\begin{matrix}\text{Domestic}\\ \text{plus}\\ \text{Foreign}\\ \textit{S}\text{tocks}\end{matrix}\right)^{d} \times \left(\begin{matrix}\text{Foreign}\\ \textit{E}\text{xchange}\\ \text{Gradient}\end{matrix}\right)^{e} \times \left(\begin{matrix}\textit{T}\text{ime}\\ \text{Trend}\end{matrix}\right)$$

The small letters *a, b,* etc. represent the elasticities of price in terms of the factor in the parentheses; that is, the percentage change in price that will result from a given 1 per cent change in each factor. Time trend has no elasticity coefficient since it is the sum of all the remaining factors which operate slowly and gradually through time.

The equation may be expressed in the functional notation of mathematics by using the italicized letters to indicate the factors involved:

$$P = f(I, D, F, S, E, T)$$

---

[1] In simple mathematical terms, *a priori* reasoning may suggest not only the terms of the function, but its form, whether it is linear, curvilinear, or repeating, whether in two dimensions, a surface in three dimensions, or some undrawable figure in a multidimensional space.

[2] Charles F. Roos, "The Dynamics of Commodity Prices" in *Studies in Mathematical Economics and Econometrics in Memory of Henry Schultz,* edited by O. Lange, F. McIntyre, and T. O. Yntema, Chicago: University of Chicago Press, 1942, pp. 283–285.

By various statistical devices one may compare the time series of each of these factors and their quantitative effect upon price may be determined as an elasticity figure. This usually requires that the data be averaged or summated for successive yearly periods and that changes be expressed in terms of the difference between successive years. The probable form of the equation is decided by *a priori* reasoning and its constants are obtained by various correlation techniques. The final equation is then applied to the data to obtain calculated "should-be" prices which are compared with the actual prices in order to determine "correctness of fit" (most easily seen by plotting on the same graph the curves of both calculated and actual prices). If the fit is not good, the trouble may lie in either some factor which has been overlooked or in a time lag which has not been properly calculated.

The exact equation obtained by Dr. Roos in the wool price study is not given in his article, but he does indicate the elasticities as follows:

a (Consumer Income)	=	+ 0.88
b (Domestic Production)	=	− 0.48
c (Foreign Production)	=	− 0.03
d (Domestic + Foreign Stocks)	=	− 0.04
e (Foreign Exchange Gradient)	=	+ 1.38

This may be read to mean that a 1.00 per cent change in consumer income, for instance, will mean a change of 0.88 per cent in the price of wool in the same direction. Changes in foreign exchange rates influenced the domestic price of wool more than anything else. Changes in consumer income (that of individuals in the lower bracket) were second in importance, and changes in domestic production were third. The other two factors were not very important in the period studied, 1923–1939. The signs preceding the elasticities indicate that an increase in consumer income or foreign exchange rates increases the price, while an increase in production or stocks reduces the price. If these elasticities continue to operate in the future, they should be helpful in estimating what would be the effect on wool prices if the government were to change the tariff rate, control domestic production or sales, or manipulate foreign exchange rates. Since no time lags are indicated in the analysis, the future price cannot be predicted by examining current data. An economic or political prediction about all of the independent variables must be made first, and after that an estimate can be made of how much these expected changes will affect the price of wool.

In other analyses of time series relationships, time lags have been discovered. Where they do occur, the emergence of phenomenon X

indicates that it soon will be followed by phenomenon Y, provided no unusual circumstance intervenes. Thus, a rise in the price of corn usually is followed by a rise in the price of hogs, or an increase in payrolls precedes a rise in the reported volume of retail sales. Sometimes historical data reveal characteristic fluctuations, such as the seasonal variations in the price of eggs or the volume of money in circulation. If these fluctuations tend to have a definite pattern of amplitude and a regular periodicity, fairly precise quantitative and temporal predictions are possible, but these conditions are rarely fulfilled. Thousands of men are still trying to "beat the stock market" by guessing better than the average guesser which way the prices of certain stocks will move, how soon, and for how long.

4. Quantitative Studies of the Elasticity of Demand.—Although our attention is now directed chiefly to the complex causes of price fluctuations, simpler studies of the same quantitative type should be noted, especially the measurement of the elasticity of demand for various products. In these, price is made the independent variable and the quantity demanded is made dependent upon it. Sometimes another independent variable is added in the form of an index of consumer income. All other variables, such as changes in consumer tastes, are assumed to change so gradually and so slowly that they may be represented as a time trend, t.[3]

A famous study of sugar prices was made by Henry Schultz. (See both his *Statistical Laws of Demand and Supply with Special Application to Sugar,* 1928, and Chapter VI of *The Theory and Measurement of Demand,* 1938.) In the latter work, he derives the per capita demand formula:

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

to express the influence of changes in price, y, upon the demand for sugar, x, in the period from 1875 to 1895. The origin of the time trend, t, is taken as 1885. The elasticity of demand in this function averaged — 0.37. That is, on the average a 1 per cent change in the price brought an opposite 0.37 per cent change in the quantity demanded. The elasticity at high prices was greater than at low prices and for equal prices it decreased numerically each successive year of the twenty-year period in his study.[4]

Another fruitful result of Schultz's approach is its emphasis upon the shifting of demand curves as indicated by the time trend factor, t.

[3] Cf. the distinction between market demand and schedule demand as described in Chapter 2, Sections 1 and 2.

[4] The price data used in this study were corrected for changes in the price level.

This quantitative estimate of the trend of demand is of vital importance in public programs for the support of prices or the curtailment of production. It is not enough to know how much a rise in price will reduce demand (the elasticity coefficient), but we should know also the long-term trend of demand in order to discover probable future needs for greater or lesser curtailment of production. In some of these studies of agricultural commodities, consideration also must be given to the effects of changes in the prices of both substitute goods and complementary goods. Sugar and coffee, for instance, are complementary to each other, while beef and pork are substitutes for one another. In each class, the prices of both goods are interrelated. Statistical studies reveal the quantitative correlation between such items and help in forecasting and in controlling their price movements.

Most of the commodities studied by Schultz have a fairly stable demand but a fluctuating supply, because they are food or fiber staples of wide consumption whose output is subject to the vagaries of the weather and crop hazards. Their sellers are numerous and they cannot fix the selling price except when they get the government to do it for them. Hence the quantitative studies of price-elasticity of demand are highly valuable to agricultural policy planners. The following table summarizes some Schultz's findings: [5]

TWO MEASUREMENTS OF PER CAPITA DEMAND FOR CERTAIN AGRICULTURAL PRODUCTS, UNITED STATES, 1915–1929 (EXCEPT 1917–1921)

Product	Price-Elasticity of Demand	Time Trend (Per Cent Shift per Year)
Rye	− 2.96	− 7.66
Buckwheat	− 0.90	− 3.20
Oats * (ex. 1917–1923)	− 0.60	+ 0.54 †
Barley	− 0.53	+ 0.69 †
Corn	− 0.48	− 2.13
Hay *	− 0.46	+ 2.24
Sugar	− 0.34	+ 0.90
Potatoes	− 0.32	− 0.15 †
Cotton (1914–1929)	− 0.12	− 0.75
Wheat	− 0.08	− 0.21

* Consumption per animal.

† Not statistically significant, since the "standard error" exceeds the figure shown. This comparison is a statistical device for testing the probable precision of the averages or coefficients calculated from samples.

[5] See *The Theory and Measurement of Demand*, Chicago: University of Chicago Press, 1938, Chapters 6–15.

The above commodities have been ranked in order of diminishing price-elasticity of demand. Rye is the only one which had an elastic demand, and its experience was exceptional in these years. For all the major crops the demand was less than unity and closely approached complete inelasticity in the case of cotton and wheat. In all such cases a reduction in the size of the crop will increase the total return to the farmer-producers. This explains in part the reasoning underlying the programs for crop reduction and restriction in this country during the 1930's. The time trends are significant also. Demand was declining during the period for six out of ten crops and very rapidly for rye and buckwheat. The positive trend for oats, barley, and potatoes is not certain, since in each case the derived trend percentage is less than its standard error. Hay is the only crop showing a definite increase and that is on a per animal basis, so that it might be offset by a decline in the number of animals.

5. Difficulties in the Quantitative Measurement of Economic Relationships.—One should not get the impression from the above description of the use of statistics to strengthen deductive theory that all economic questions can be answered by their use. On the contrary, the usefulness and reliability of mathematical equations involving numerical coefficients, constants, and exponents in prediction and control are definitely limited. In the first place, there is much room for personal choice in the selection of the data to be analyzed—whether to use the data gathered by one agency or another, or those related to one level of transactions or another. Sometimes less reliable data are chosen in place of more trustworthy data because the former may possess advantages of continuity or completeness. Second, when certain time series are incomplete or have changed their bases of compilation or publication during the time interval, interpolation or extrapolation must be employed, with the errors inevitably attendant upon the use of these devices.

The third difficulty already has been mentioned: the problem of choosing the most appropriate equation and method of statistical analysis. For instance, Schultz gives six different answers to the elasticity of demand for sugar from 1915 to 1929 depending upon the equation and method used. They run from -0.31 to -0.40, which is not a seriously wide spread, but when the probable error is subtracted from the former figure and added to the latter, the range becomes -0.08 to -0.52, which is so wide as to cause hesitation in trying to predict the probable change in demand which would follow a given change in price. The fourth difficulty is an analytical one:

how to decide from the historical evidence whether a given regression represents a demand curve of the short-run type or a supply curve in the long-run group. If general observation of the motives and actions of producers indicates that their output may be expected to fluctuate more often than the intensity of demand, we may infer that the data should be analyzed to reveal a demand curve. If demand fluctuates more rapidly, then it would be more logical to develop a price-quantity supply curve. Certain statistical methods also have been developed to aid in making the distinction as to which curve should be sought in given historical data.

Fifth, it is important to recognize that the derived curve is essentially a picture of an average situation during the period analyzed and that this average might have been different if the period had been selected so as to begin or end at a different time. For instance, according to Schultz, the demand shift for wheat was only — 0.21 per cent per year during the years 1921–1929, but if 1921 to 1934 be taken for analysis instead, the time trend would be — 0.50 per cent, more than twice as large, and the price elasticity of demand also would jump from — 0.08 to — 0.19.

6. Problems in the Application of Price Prediction Formulae. —In addition to the difficulties of obtaining a formula which correctly states the relation between independent and dependent variables, there is also the problem of knowing when, where, and how to apply these formulae. Their usefulness in prediction depends upon the future repetition of past patterns of events. These include not only the economic routine of demand and supply, but also the highly unpredictable political developments, "acts of God," inventions, court decisions, etc. Errors may arise both from failure to realize the exact situation which the basic data represent and from misapplication of the findings. But errors are not limited to the faulty application of inductively derived conclusions. If anything, they are more apt to be the mistakes of those who seek to apply deductive theorems.

One problem of application lies in the nature of the data. We are prone to think of a demand curve as applying to the purchases of ultimate consumers, but the buyers whose purchases are reflected in most of the statistical data of prices and quantities are middlemen, manufacturers, or even outright speculators. The derived demand curves picture their reactions to price change, and because of the greater speculative element in all such buying the probable error in prediction exceeds that which probably would hold for routine buying by consumers. A statistical demand curve often is pictured as

having uniform elasticity throughout and also as being reversible, but we know from other evidence that the elasticity is usually different at high prices and at low ones, and that buyers react differently to a price increase and to a price drop. Similar situations exist on the supply side. Farmers in particular tend to be guided as much by the recent trend of prices as by the actual level prevailing at the time of planting. Furthermore, the weather plays such an important part in the outcome of planting decisions that if possible the supply responses of farmers probably should be measured in terms of acreage planted instead of bushels or pounds of crops harvested.

Statistical investigations of the price behavior of particular items must be limited chiefly to standardized products bought and sold by large numbers of people. They include farm products, minerals, and other industrial raw materials. If prices are subject to monopolistic control, they tend to be more stable, but any changes that do occur are less predictable by outside analysts. If the government intervenes with price-support or price-ceiling programs or introduces special devices to stimulate or curb production, predictions must be revised to take these into account. Past experience, nevertheless, will help in judging the necessary extent of these efforts to achieve the desired goals and also in overcoming the difficulties that will be encountered in striving to reach them.

7. Summary.—This chapter is an introduction to the problems of measurement of economic forces. Two approaches may be used. The first asks what are the forces which *cause* the price of a commodity to change over a period of years and how strong is each of them. The second asks what are the many *effects* which follow from a given price change and what are their magnitudes. Only the first has been sketched here. It must be mastered if economics is to be a science of prediction and not merely a series of generalizations regarding tendencies.

Most principles of quantitative economics cannot be developed and learned in advance. They must be formulated in terms of the statistical correlations bearing upon the problem at hand. Statistical sources, however, can be discovered and appraised. Principles of statistical analysis and advanced mathematics are fundamental tools. Therefore, the practicing economist has much to learn after becoming familiar with basic qualitative principles if he wishes to undertake the analysis of specific problems in quantitative terms.

PART II

THE PRICES OF SERVICES

Chapter 16

GENERAL PROBLEMS OF SERVICE PRICE THEORY

1. The Prices of Services Resemble the Prices of Commodities. —The second part of this volume deals with the prices of services. These prices have four well-known names: wages, rent, interest, and profits. The first three are paid by business firms or individuals to other individuals. Profits are received by individuals, but are paid by no one, unless by "society" for the services rendered by business units. Our basic problem is to explain the prices of these services and to examine their effects. Many of the principles of commodity price theory apply to service prices, too. But markets are different and certain behavior patterns are peculiar to buyers and sellers of services. That is why the two types of prices are usually examined separately.

Service price theory is often called distribution theory. One reason is found in the fact that wages, rent, interest, and dividends are usually paid by a business unit out of its income.[1] They seem to represent the division or "distribution" of that income. Another reason is that most individuals receive their portion of the national income in one of these four ways. Almost the entire income of the country is "distributed" to people through the prices paid for factor services which the people sell. Exceptions may be noted for charity, social insurance payments, pensions, graft, and bribes. The latter two are interesting in that when they are paid to people who do things that others want, they are akin to wages, but sometimes they are paid for *not* doing things and then are clearly in a class by themselves.

Distribution theory also may include the economic problem of the relative sizes of the aggregate shares of the national income going to each major factor group. For instance, economists study the question of why labor as a whole gets more than two-thirds of the national income in this country, and why it gets a larger share at one time than another. In a society which depends upon individual re-

[1] The difference between profits and dividends will be clarified in a later chapter. For the present the terms will be considered virtually synonymous.

sponse to profit prospects for the guidance of production, the size of the profit share may be very important. Expectations of percentage return per dollar invested are the prime consideration, however, not the percentage of national income going to owners of firms as a whole.

The specific questions to be considered in this introductory chapter include the following:

1. Why is service price theory also called distribution theory?
2. What are the four main classes of service payments and how may they be subdivided? Classes of services?
3. How may "use prices" be distinguished from "ownership prices"?
4. What are the general characteristics of demand and supply in factor markets?
   (a) Peculiar subjective and objective considerations?
   (b) Joint demand, derived demand, cross elasticity, etc.?

2. There Are Many Subgroups in Each Factor Class.—Although economists usually divide factors into four main groups and call their *payments* wages, rent, interest, and profits, there are many subgroups in each class. Price analysis that holds for one subgroup must be changed somewhat to describe another. Labor wage problems, for instance, are different for the salaried worker and the day laborer. Organized musicians sell their services in a different manner from casual farm laborers. Similarly, in talking about rents, one should distinguish many classes of rental markets. The rental problems of parking lot owners differ from those of companies owning automobiles for hire. Homes are rented differently from office buildings, and farms from factories. Dollars look much the same when borrowed, but lenders differ in their willingness to accept certain types of risk. No two business units are completely alike in revenue and cost experience. The profit price that society pays for their services is uniquely determined in each case.

Despite these differences, cases may be grouped according to the elements which they have in common. Generalizations may be made about the phenomena which are found in small groups, or in larger groups which include these small groups. Theories which describe and explain collective bargaining, for instance, have certain common denominators with theories about the wages of unorganized labor. These elements in common may then be combined by a process of abstraction into a general wage theory. The important thing is to remember the level of abstraction for which a theory is designed. No theory is a complete map of reality. It is merely an outline sketch

to indicate the picture whose details may be sought or ignored according to the purpose of the moment.

3. The Factors of Production as Services and as Sources.—The *services* whose prices are to be analyzed in this part of the book are often called "factors of production." They are used in agriculture, manufacturing, trade, and all other productive enterprises. But they are strangely without distinct names. The services of human beings are sometimes called "labor," sometimes "work," and sometimes just "services." The word *labor* often is used also to describe a group of laborers, as when a manager speaks of "bargaining with labor." When one rents the services of land, he may seek its fertility, its minerals, its ability to support buildings, its sunshine, etc. There is no single term which describes the variety of services which the same piece of land might render. So, too, with capital goods such as buildings, railroads, and machinery. They shelter, aid the movement of goods, improve the efficiency of labor, etc. The services of capital funds are sometimes described by the alternative term *purchasing power*. But funds also are desired for the service they perform in furnishing "liquidity" and in facilitating "hoarding." A similar difficulty arises with the services of a business unit. These are usually called "production," but the profit reward which goes to the owners is not for production, a gross return, but rather for success in achieving a positive *net* return in the act of production.

Even though the services or factors have no distinct names, their prices are almost always called wages, rent, interest, and dividends (or profits). These prices are paid to the owners of the *sources* of the services. Rent is paid to the owners of tangible property, interest to the owners of funds, and dividends to the owners of incorporated business units. These three are forms of "property income." If a man's body were considered to be wealth which he owned and let others use, wages could be called "property income" too.

4. Use-Prices and Ownership-Prices.—Payments for services may be called "use-prices" and payments for sources called "ownership-prices." Distribution theory is largely concerned with the "use-prices" of factors of production which are obtained on a temporary basis. Wages, rent, or interest are paid on some type of time contract for services. In each case there is an express or implied promise to return the factor source substantially intact. In renting real property, an exception is made for "normal wear and tear," a vague and often litigious phrase. When borrowed dollars are returned as dollars in some form of legal tender or acceptable draft or check, the

debt is usually considered to have been fully discharged even though the purchasing power of the returned dollar may be much less than that of the borrowed dollar. Workers likewise may be worn out by their jobs, may contract occupational diseases, or may be injured in accidents. In recent years this impairment of the factor source has been made an increasing responsibility of the employer. Workmen's accident compensation laws have been followed by old age assistance to which employers must contribute. Sickness insurance, especially for occupational diseases, also is now beginning to appear on statute books in this country. It has long been common abroad.

Sometimes the source of the factor service is purchased outright. An "ownership-price" is paid which closely resembles the "transactions price" paid for commodities whose services are desired for direct consumption. It gives the buyer the right to use the services of the good without any stated time limit. For instance, the services of land, buildings, and machinery may be obtained by paying a rent-price for temporary use or an ownership-price for permanent use. A business unit also may be rented, as in the leasing of railroads, or bought, as through the purchase of stock shares. Now that slavery has become illegal, men and women cannot be bought outright like other animals or inanimate objects. Their services must be purchased on a time basis. Currency may be bought outright, as in the foreign exchange market. More often it is borrowed. The price paid for the use of funds for a period of time is called "interest." [2]

Another type of use-price is that offered by society when it employs profit prospects to entice or induce people to use factors in production. Society does not obtain permanent rights to factor sources committed to productive effort. Therefore, profits cannot be called an ownership-price. But profits also differ from most use-prices. There is no contract requiring that they be paid for services rendered. Nor is any definite time period involved like that for wage, rent, or interest payments. There is merely the implicit promise that if the income from the productive use of the enticed factors exceeds the outlays, the difference may be kept by the owner of the factor sources as his "profit" for the interval.

5. Subjective Considerations in the Use-Prices of Factors.— Subjective considerations dominate the selling side in factor markets. This is the opposite of commodity price markets where the selling side is chiefly objective and where it is the consumer buyer whose

[2] A distinction will be made later between the services of the funds in the hands of the buyer and the services which the owner of the funds performs when he allows the buyer to use them.

motives are usually subjective. The factor seller is usually a private individual and the buyer a business unit. The individual is influenced by *subjective* considerations particularly when it is his labor which he is selling. He considers what he must give up, his leisure, his time which he might use for himself, the pain and fatigue of exertion, etc. If he owns many pieces of real property which he rents, he may consider his being a landlord a business. He may adopt the objective profit-and-loss attitudes of the manager of a business unit. But if he has only one or two pieces of property for rent, subjective considerations of effort, of things given up, of use-opportunities foregone, or of the condition of his property when returned, may weigh heavily in his calculations of the price he will try to obtain for its services. A similar line of reasoning seems applicable to large and small lenders and to owner-managers of small business units as compared with wealthy individuals owning many shares of stock.

On the opposite side of the market a similar contrast may be discerned. The buyers of factor services are usually the managers of business units who are motivated by *objective* comparisons between income and expenditure, between costs and revenue. Of course there are exceptions to the generalization. Some services are purchased by consumers for their direct use such as the so-called personal services of barbers, physicians, taxicab drivers, and tailors. The services of property are often sought for residence purposes, either directly through renting a house or apartment or indirectly by borrowing to buy a home, automobile, vacation, etc.

6. Derived Demand for Factors.—Since the demand for factors of production comes chiefly from business firms, it has certain peculiarities. In the first place, the intensity of that demand depends in part upon the demand for the product of those firms. Therefore, factor demand tends to rise and fall as the manager changes his ideas about the future rise or fall in the demand for his product. In the second place, any single factor is only one of many which the firm needs to conduct its business. Factor demand is joint demand and is therefore a function of the prices of other factors and the technical possibility of substituting one for another. If many good substitutes exist at fairly comparable prices, then the price-elasticity of demand for a given factor will be great. The cross-elasticity of demand will also be large. A rise in the price of one will cause much substitution of others for it, as in the case of labor and machinery (or the funds to buy it).

The factors which cost the most in a joint demand situation will have the largest influence on the selling price of the product or on the ability of the firm to continue in production. Hence their elasticity of demand will be greater both for price increases and for price decreases than will that of factors whose cost is only a small fraction of the total. The relative importance of factors will differ of course from product to product so that generalizations of wide applicability are difficult. It is at least a 95 per cent generalization, however, that wages constitute a larger part of total costs than does interest. Therefore, a 10 per cent increase in wage rates is likely to reduce the demand for labor more than a 10 per cent increase in interest rates will reduce the demand for funds. The demand for labor is more elastic than the demand for funds.

If the factor demand problem is approached from the narrower focus of the individual firm, or even the industry, the truth of another generalization becomes apparent. The more inelastic the demand for a product, the more inelastic the demand for the factors used in making that product. If unions ask higher pay, the employer is most likely to accede when the product demand is inelastic. If he has to raise the price to pay higher wages, his sales and therefore his employment will not decline very much. Bituminous coal is a good example of a derived demand that is inelastic, at least in good times. The factor demand by makers of commercial pastry products is probably much more elastic.

7. General Aspects of Demand and Supply in Factor Markets. —Service price theory seeks to explain the prices of the factors, their relative magnitudes, their changes, and their effects. Like all other prices, service prices are the result of the interaction of demand and supply. But factor markets are even more heterogeneous than commodity markets. Institutional variations must be examined in detail before comprehensive generalizations can be attempted. At this point only a few are possible.

There is some price-elasticity in the demand of individual buyers. This is based upon diminishing marginal productivity from use of inputs, decreasing marginal revenue from sale of outputs, or both. There is also some price-elasticity of supply by individual sellers, derived from increasing marginal costs, whether objective or subjective. Because of their cumulative character, collective demand and collective supply almost inevitably follow the usual laws. Factor heterogeneity makes for considerable bargaining. And of course there are many cases in which sellers are able to set and to obtain their offering prices. Price setting by the buyer is less common.

There is much cross-elasticity of demand for factors, although it does not appear so quickly as in consumer demand for commodities. The case of rising wage rates and their stimulus to the introduction of labor-saving machinery is well known, but the substitution takes time. There is also considerable income-elasticity of demand. Rising business incomes usually promote increased demand for factors. In both these types of demand-elasticity, technological factors influence the speed and degree of response. Subsequent chapters will examine each of these points in detail.

The elasticity of factor supply may be generalized under the same three headings. Some factors, like mineral-bearing lands, are quite limited in supply and therefore evince very little supply-elasticity of the price variety. Others can be reproduced and usually obey the law of increasing marginal cost. The longer the time allowed for supply expansion, the greater the apparent supply-elasticity in response to the rising factor price. Long-term contracts for loans and leases reduce the immediate cross-elasticity of supply. Ultimately, shifts in employment do occur at the initiative of the seller. Income-elasticity of supply is of the usual positive type in the case of loanable funds, but with labor supply a negative relationship often appears. For durable goods of fixed quantity, the income-elasticity like the price-elasticity of supply is zero, at least temporarily.

Factor markets may be classified in the traditional way as revealing either "pure competition," "monopoly," etc. A more realistic approach would stress the importance of custom in some cases, bargaining in others, legal minima and maxima, etc. The different types of collusive action by sellers should be detailed and not grouped under the artificial headings of "oligopoly," "duopoly," and monopolistic competition.

8. Summary.—The foregoing arguments and observations show the need for dividing price analysis into many parts. Service price theory is one main division. It has four traditional parts: wage theory, rent theory, interest theory, and profit theory. Each of these must be subdivided still further by type of factor and type of market if we are to get a more realistic picture of the causes and effects of prices in our economy. Even then, particular problem situations in any subdivided field will differ from one another.

Business firms pay use prices for the services of factors of production. Sometimes they buy outright the service source by paying an ownership price. The demand for factors is derived from the demand for their products. Joint demand situations influence both the price-elasticity and the cross-elasticity of demand for factors.

Service prices are best explained in terms of types of market behavior. Knowledge of the details of market situations involving labor, land, etc., helps to build generalizations about common aspects of service price determination. This is the inductive approach to theory. Knowledge of general principles of service pricing helps one to analyze specific problems. This is deductive reasoning. Economics has need of both. The following chapters, therefore, will consider both the institutional (inductive) and the theoretical (deductive) approaches.

Chapter 17

SPECIFIC WAGE RATES: AN INSTITUTIONAL APPROACH

1. Problems in Wage Theory, a Preview.—The problems of wage theory may be divided into at least four parts. The first would explain specific wage rates by examining the interplay of forces determining the demand for and the supply of labor in particular situations. This may be called the institutional approach and will be examined in the present chapter. The second might examine the more general problems of equilibrium situations under certain assumptions. On the demand side this includes the equilibrium of the firm in its demand for labor. On the supply side there is the analysis of equilibrium wage rates for different labor groups. This is often called "wage theory" *per se,* but it is really only an *a priori* approach to wage problems.

A third part would involve a broader approach and consider the differences in *real* wages from region to region and from one time interval to another. Finally, there is the relation between money wage rates and the aggregate demand for labor in a given economy. This is one aspect of the problem of unemployment. For convenience, each of these four parts of the theory of wages will be treated in a separate chapter, even though they are closely related.

2. Topics to Be Studied by the Institutional Approach.—Money wage rates are prices and may be analyzed by the approach used for commodity prices. Like the prices of goods, the prices of labor services are determined in many ways. The wages paid and received in particular situations are determined by the conditions of demand and supply that prevail. Bargaining is much more common than in the market for goods. Therefore we must examine carefully the customs, laws, and types of economic organization which influence bargaining. These are called institutional forces. They help to set the range of asking and offering prices in the minds of the bargainers. They also explain the relative bargaining power of the buyer and seller.

Because of the great variety of these institutional situations, generalizations are difficult, but they must be sought. Only thus can the study of wage rates be made useful for prediction, adaptation, or control in connection with the labor problems of everyday life. The following list of questions treated in this chapter indicates one way to classify situations. Another approach will be found in the concluding summary. Both should be studied carefully.

1. What is the nature of labor, its relation to laborers?
2. How are particular wage rates determined?
   (a) When set by employers?
   (b) When set by workers?
   (c) When bargaining occurs?
   (d) When set by law?
3. How and when are wage rates raised?
   (a) By the initiative of the employer?
   (Concepts of "labor float" and "time-elasticity of supply")
   (b) By the initiative of the worker?
   (Concepts of "time-elasticity of demand" and "elasticity of substitution")
   (c) Alternatives to increased pay?
4. How and when are wage rates reduced?
   (a) By the initiative of the employer?
   (b) By the initiative of the worker?
   (c) Alternatives to reduced rates of pay?

3. Labor and the Worker; the Factor Service and Its Source.—Labor is the mental or physical energy expended by a human being for something other than the pleasure of expending that energy. When a person puts forth such energy he is said to work and is usually called a worker. Highly paid workers often coin more high-sounding terms to describe themselves and their labor such as "executives" and "professional services," but the fundamental similarity remains. The economist is chiefly concerned with labor performed for others for pay in money or in goods. Occasionally he finds it useful to examine the circumstances surrounding labor performed for oneself, either to produce something for future use or to obtain immediate satisfaction. There is little to be gained by quibbling about where the line should be drawn between labor which is "work" and that which is "play."

Labor, the factor service, is inseparable from the worker, the factor source. That is why labor as a factor of production is often considered to be the worker himself rather than the effort he puts forth. More important is the fact that the factor source is in-

separable from the owner of that source. For the service to be rendered, the owner of the source must be present. This is not true in the case of real estate, personal property, or funds, where all that is required is the presence of the factor source.

Because labor is a service performed by human beings, its supply is limited by things which people dislike, such as the fatigue of exertion, boredom, and the fear of injury or illness from work. Because labor may also be performed for himself or the time used in more pleasant ways, the worker is a demander as well as a supplier of labor. The owners of other factor sources may sometimes occupy a similar dual position on both sides of the market, but less frequently. The owner of rental property can rarely use it for himself, and even the owner of loanable funds usually considers those funds desirable for their income-producing potential and does not seriously consider the possibility of spending them. Both may let their property lie idle waiting a more opportune time or better terms of sale, but they do not "consume" the services of the factor source during the interim nearly as often as does the worker. The distinction can also be made in terms of the much larger number of possible uses of a person's time than of property time. Real and personal property are more highly specialized in their functions than are human beings. An unemployed machinist may not have any opportunity to use his skills at home, but he can probably think of many things he would enjoy doing with his time.

On the other hand, a man looking for work usually feels the need for pay more than the need for leisure or time to work for himself. An unemployed worker usually has no other source of income than his labor. By contrast people who own rentable or loanable property which is temporarily idle usually have other property income and often a supplementary work income. The savings of workers are usually small, but most workers now can get temporary and limited help from unemployment insurance funds. The result is that workers may be pictured as weighing alternative uses of their time when they try to decide whether to take a given job opportunity or not. They may also appraise other aspects of the job such as hours, conditions of work, etc. The point to be emphasized here is the extensive subjective element in decisions regarding the sale of labor. The decision of a worker selling labor resembles that of a consumer buying commodities when the latter weighs the satisfaction expected from the good purchased against the satisfaction of keeping his money with its possibilities of future use for a variety of other goods. The subjective valuations of both worker and consumer must be contrasted

with the objective monetary valuations of the business firm on the buying side of the factor market and on the selling side of the commodity market.

There are exceptions, of course. In some cases labor is demanded for immediate personal consumption, such as the services of doctors, barbers, teachers, and entertainers. The consumer is the employer and the worker is selling labor to meet a direct and not a derived demand. On the other hand, labor is sometimes sold like a commodity as in cases where the employers hire unskilled farm laborers and sell their labor in gangs under contract to large buyers. Or a person like a professional baseball player may make a long-term contract to deliver services to some buyer who is privileged to sell that contract to some one else.

There are thousands of different services which laborers can and do render for others. Each of these different types of labor may also be performed at a multitude of different places. Hence it is possible for many different wage prices to exist at any given time and there is no single price nor any common denominator for that which we call labor. The economic analyst can study the height and movement of any particular wage price he wishes. In a general study like this he must confine himself to the quest for explanations of wage rates in the labor market situations which occur most frequently or which have the most significant effects. These include hourly, daily, weekly, or other time-period wage rates, whether called wages, salary, stipend, or other synonym. It also covers wages paid per unit of service performed, such as factory piece rates, selling commissions, and professional fees. The name is not important.

4. Particular Money Wage Rates May Be Set: I. By an Employer.—The specific wage rates which exist at any particular time and place are the results of agreements between employers and workers. The most common situation is probably one in which employers offer to pay and workers agree to accept the going wage for the particular type of service involved. It is customary for large employers to have a wage policy of paying certain definite rates for the various types of work. When one man quits or dies, another is hired to take his place at the same wage if he possesses at least as much skill. If not, the replacement is hired in a lower category until he acquires the skill needed to qualify him for the higher pay classification.[1] Small employers tend to follow the same procedure,

[1] This does not mean to imply that wages are determined by skill alone, nor that all employers follow this type of employment policy. Bargaining is very important, as will be argued below.

although more by approximate rule-of-thumb technique than by the schedules of a personnel department.

In recent years a technique of job evaluation has been developed by certain large employers. An expert analyzes each type of work to determine the demands which it makes on the worker. For instance, jobs vary in the amount of skill and experience they require and each one is rated at some number on a scale of 1 to 10. Job A may rate 9 points on the skill scale, while job B may be evaluated at only 6 points. Similar ratings are made for the amount of judgment required, the physical energy output of the worker, the risk of accident, etc. The total rating for a job might then be $9 + 6 + 3 + 4 + \ldots = 35$, or some such figure. A system of weighing the components may also be used, giving more importance to one aspect of a job than to another. A comparison of the totals reveals the relative demands of various jobs upon the worker. All those in the 35-40 group, for instance, should therefore be paid more than those in the 25-30 group. Wage scales are often set by employers in cooperation with workers, using such rating systems as a basis. Whenever any new job is created, it is evaluated by the method used for other jobs, and the appropriate wage is thus determined.

5. Particular Money Wage Rates May Be Set: II. By Bargaining.—Particular money wage rates may be determined also by bargaining between the worker and his employer. Unless the worker is an executive or the possessor of some very rare manual skill, he is not likely to do much bargaining as an individual. He will either accept or reject the employer's offer. But in a few cases there will be higgling, offers and counter proposals, until agreement is reached or negotiations are broken off. The principles of wage determination under such circumstances are the same as those for any bargained price. If the potential bids and offers overlap, an agreement may be reached. The precise figure will depend upon the circumstances. In general, relatively great strength and skill in bargaining lead to a wage close to the favorable end of the overlap. For the employer, this is the lower end; for the worker, it is the upper. This, however, is only an elementary statement. We must still examine many things which influence (1) the bid and offer ranges and (2) the relative strength of the contending parties and their skill in bargaining.

When bargaining fails to produce agreement, the contending parties may agree to submit the dispute to arbitration. This is most commonly done when organized labor is involved, not individual workers. The arbitrator (or board of arbitrators) listens to the

arguments and then recommends a certain wage. Arbitrators use various formulae in reaching their decisions, depending on the case at hand. Sometimes they decide upon the basis of the apparent relative power of the disputants and then hunt for a formula which yields the same result but appears to be "impartial" and "scientific."

6. Particular Money Wage Rates May Be Set: III. By the (Organized) Workers.—In a few cases labor-unions become so strong that they can virtually dictate the money wage rate which employers must pay. They control all the available supply of a given type of labor, such as musicians. They announce the rate of pay for each service their members might perform. Employers must pay this wage or go without. No individual worker is allowed to work for less than the "union scale," under penalty of expulsion from the union. If an employer tries to hire nonunion labor, union workers may conduct a strike or boycott against him.

This is the familiar "monopoly" problem. The union behaves monopolistically in trying to get as high a price as possible for the labor involved. Wages tend to be high relative to the wages paid other workers not so tightly organized to reduce supply-price competition. Yet there is a limit to the height to which union wages can go, namely, the cost of acceptable substitutes. These substitutes may be machinery, land, or other types of labor which may be used in producing the commodity or service involved. Their prices and their efficiencies in production influence the price- (wage-) elasticity of demand for the labor concerned. These problems and that of the time required to make the substitution will be discussed at greater length below. At this point we need add only the comment that sometimes the commodity involved may be "priced out of the market" by rising labor costs. Some consumers may turn to other goods which can be produced with another type of labor, presumably cheaper than the high-wage group. Prefabricated houses are a good illustration. Or the substitute good may be produced by the same labor in a cheaper way, such as phonograph record music as compared with live orchestra music.

Employers differ in their ability to bargain with strong unions. Small employers may pay union rates without argument. Larger firms or groups of firms may engage in long and protracted negotiations at the time of contract renewals. In such cases relative bargaining strengths and skills return to the picture.

The actual wage rate demands of unions are determined in various ways. Prior rates are the starting point. Changes (usually up-

ward) spring from changes in the cost of living, the level of employer profits, the volume of unemployment, the wages received in similar occupations, etc. Pay differentials in union scales are rarely worked out with the precision of job-evaluation analysis, although sometimes union men join in the analysis itself. Customary differences tend to be retained and justified on the basis of what has been received in the past.

7. Particular Money Wage Rates May Be Set: IV. By the Government.—At times the government intervenes to set particular money wage rates. This is usually done by establishing maximum or minimum rates. Occasionally there are regulations which require the payment of the "going wage." This often means the union scale in that community. Governmental units usually have civil service scales for their own employees. In times of generally rising wage rates, the government scale may be lower than that for private business. Inflexibility is one reason. Taxpayer resistance is another. Often there are also certain advantages in working for the government, such as greater certainty of employment in the future, retirement pensions, longer paid vacations, regular seniority promotions, etc. For many people these advantages offset lower pay.

Legal minimum wage rates have been established in many states, and industries affecting interstate commerce are now covered by federal legislation. Since these rates are usually higher than customary minima—otherwise the laws would be futile—they supersede custom and make bargaining very unlikely. In wartime legal maximum rates may also be established by government action as fixed dollars-and-cents amounts, as a certain percentage of the rates paid on a given base date, or whatever the rates were before ("Whatever was is right"). Again bargaining disappears because employers are so eager to keep their workers on the job that they are willing to pay all that the law allows.

In either case, legal maxima or legal minima, classification of workers is often used to evade the law. Apprentice beginners are sometimes paid less than the minimum on the grounds that they are just learning the business, and then are fired when they get experience. Evasive methods of reclassification for men at the top are easy. When it is possible to change jobs from one firm to another, workers can hunt the highest ceilings.

8. Three Methods of Attracting Additional Labor.—When employers need more labor, or a new firm enters the labor market, there are three distinct ways in which this labor may be obtained.

The first is to offer the going wage and make one's wants known. Unions may be asked to furnish more men.[2] Newspaper and billboard advertising may be used. Employment exchanges may be asked to send applicants from their lists of unemployed persons. Other devices include streetcorner recruiting and the establishment of special employment offices downtown or in other cities.

A second method is to offer various nonwage advantages, hoping to attract workers from other jobs which pay the same wage rate but lack these advantages. During the second world war, employers were prevented by law from raising wages to attract workers from other jobs. Hence they resorted to other methods. Some served hot lunches free, or at prices far below cost. Systems of rapid and automatic promotion were introduced. Lavatories, lockers, and showers were modernized. "Wage incentive" schemes offered higher pay rates for increased output. Vacations with pay, retirement pensions, and sickness allowances also proved attractive. At other times, workers have been attracted to certain jobs rather than others because of the greater probability of regular employment, of more work weeks per year or per decade.

A third way of getting additional workers when sales expand is through increased wages. This method will be explained in Section 10, but first a cautionary comment is in order. Firms competing in the labor market usually are quite reluctant to engage in wage-rate competition among themselves for fear of spoiling the market. Each knows that if it takes the lead in raising its wage offers, one or more adverse developments are likely to follow. First, its existing employees probably will demand higher wages to equal those of persons newly hired. Subterfuges of classification and new job descriptions often fail to carry conviction about an alleged lack of discrimination. If the wage increase is confined to a lower bracket, those in higher-paid groups whose ranks are not being increased will often demand raises so that "customary differentials" will be maintained. Second, other firms may raise their wage-rate offers quickly so that the first firm gets no special advantage in the labor market and finds itself faced with a higher wage cost without any gain. Third, the wage increase which is easy to initiate when workers are needed may be virtually impossible to abandon after the recruitment campaign is passed or in later periods when employers would rather reduce wages than to lay off workers.

---

[2] Sometimes merely signing a union contract, previously resisted, solves an employer's labor problems.

9. Overtime Work as an Alternative to Hiring Additional Workers.—Because of these dangers and also because of the cost of hiring and training new workers, many employers deem it preferable to persuade their existing workers to work longer than the usual time per day or week. The extra cost of overtime pay, usually time-and-a-half, is thought to be less than the cost and disadvantages of expanding the labor force. This method has the distinct advantage that it can be instituted quickly and discontinued easily when no longer needed. Nor does it disturb the basic wage structure.

On the other hand, there are several disadvantages. The last hours of a long day are not as efficient as earlier hours. Workers accustomed to a five-day week do not work as well on an extra sixth day. Therefore the labor cost per unit of product is usually more than 50 per cent higher, even though the pay rate per hour is only time-and-a-half.

Workers differ in their willingness to work overtime. Some welcome the additional pay. Employers may advertise the overtime pay opportunities as a method of attracting workers of this type. Others do not want to spend more hours away from their homes, families, rest, or recreation. When salaried workers are not paid overtime, they have another reason for not being enthusiastic about long hours. Occasional overtime work may be acceptable, but not frequent nor protracted subtractions from leisure time. Wise employers take all these possibilities into account when deciding which methods to use for recruiting new workers. Worker-management councils are very helpful in ascertaining worker attitudes in such situations.

10. Wage Raises Are a Last Resort.—If an employer cannot attract the additional labor supply he needs by any of the foregoing methods, he may raise his wage bids above those formerly paid. By offering more than other employers of similar labor, he hopes to attract workers from them to him. But he does not want to pay any more than he has to. This involves some guessing on his part. He estimates how much of a wage increase is needed to get the number of workers he wants at the time he wants them. The employer must take into consideration not only the probable subjective response of prospective workers, but also certain time factors. It takes time to spread the news. It takes time for workers to make up their minds and to shift from one job or one community to another. It takes time to interview applicants and to train them for their work. And during all this time, rival employers may take defensive steps by increasing either wage or nonwage inducements to workers to remain.

All these factors differ from case to case. Useful generalizations are difficult, but a few may be made.

11. Wage Increases Vary with the Time-Elasticity of the Supply of Labor.—The quicker an employer wants additional labor, the higher he must bid for it. This is an application of the principle of the time-elasticity of supply explained in Chapter 14, Section 2. In that chapter the concept was defined as the length of time required for a given percentage price increase to call forth a given increment in supply. In the labor market a different statement of the same principle appears more useful. The time-elasticity of the supply of labor may be measured in terms of the height of the wage increase required to attract a given amount of labor within a given time.

This necessary wage increase will differ with the type of labor, the type of employer, the phase of the business cycle, etc. It is small (the *time*-elasticity is short) when there is a large float in the labor market area of the individual firm. If the float is large enough, advertising or nonwage incentives may suffice and the necessary wage increase may be zero. But if the local float is small, many of the needed workers must be enticed away from other firms or induced to migrate from other regions where they have been unemployed. This requires a larger wage increase (shows a longer time-elasticity). When labor organizations are present, they may help get the desired labor if the employer will negotiate a new contract at higher rates of pay. This makes the time-elasticity *for that wage increase* shorter than it would otherwise have been. Put in another way, the necessary wage increase is lower than would have been required without the union's help. On the other hand, a wage increase without a union contract or without certain terms demanded by the unions may prove very ineffective. The dissatisfied unions may strike, picket, or do other things to prevent the employer from getting labor at all or to make labor recruitment very expensive.

12. Wage Increases Are Limited by the Intensity and Elasticity of the Demand for the Labor Product.—A second useful generalization is that wage increases are limited by the intensity and the elasticity of the demand for the product. The greater the recent demand for the product, the more labor will be wanted and the higher the wage that can be paid. If the enterpriser thinks that his profit opportunity will not last long, he must hasten his labor recruitment. Or if he has a contract with a penalty clause for delays, as on certain construction jobs, he must work fast to prevent losing money. The

higher the wage that an employer must pay to get extra labor within the desired time, the more irksome is that wage increase.

The elasticity of the demand for the labor *product* is also important. If the employer believes that he can raise the selling price of his product to offset the additional labor cost, he will be more willing to raise his wage bid. This is rarely true in the situation being considered here where the employer takes the initiative. It happens much more often when the wage increase is being demanded by organized workers from a group of employers who produce a similar product. This case will be analyzed in a later section where attention shifts to wage increases initiated by workers.

13. Wage Increases Are Limited by the Production Function.— Sometimes employers consider their production function in deciding how much they can afford to raise wages to attract or to keep workers.[8] If the variable cost in terms of labor input remains constant, the employer is not bothered with this problem. But when additional output involves decreasing labor efficiency, employers must consider the problem of rising labor cost per unit of output. This is true whether or not wage increases are made per unit of labor. But it is even more important when additional labor is sought at higher wages. There are few institutional aspects to the problem, however, and its consideration is best deferred to the next chapter.

14. When Should Individual Workers Demand Higher Pay?— In most cases wage increases are not initiated by the employer, but by the worker. The problems concerned with this approach occupy the next few sections of this chapter. At first we shall consider the viewpoint of the individual worker and later that of labor-unions. Most wage increases initiated from the labor side spring from the bargaining of organized workers, but the general principles introduced in this section will prove very useful in the analysis of the group situation.

Individual workers usually think they are paid less than they are worth. They have only a very hazy idea of just how much they are worth, but it is generally something more than they are currently getting. Workman Joe Doakes is not immune to the materialistic standards of a capitalistic society. Like his boss, he wants more than he is getting. He knows that in the competitive arena one must struggle with constant self-assertiveness if he would get ahead. His ego rises to the occasion. His desires are stimulated by motion

[8] The production function is the relation of output to unit costs.

pictures which exalt material success as the highest good, by radio dramas, by magazine fiction, by advertising. Joe may come under the spell of Marxist oratory. He may accept the dogma of capitalist exploitation of the workers. Even without fully understanding the doctrine of "surplus value," he may look at the profit his firm is making and say that its size proves he is being underpaid. He may not attempt to use logical reasoning at all. The boss may be paying higher wages to Jim Blount than to Joe Doakes who considers himself just as good a man. So Joe asks the boss for a raise.

In many cases what employers know is not much greater. They know profits are rising or are falling and they think that they do or do not need to keep Joe in a good humor. The threat of his quitting may or may not cause concern. In general, employers are more willing to grant raises when they are prospering than when the tide is turning against them. In small firms they are also influenced by such unpredictable circumstances as whether they happen to be feeling good when the request is received, and the manner in which they are approached.

Because of the importance of these subjective forces in the decisions of both workers and employers, it is very difficult to describe with precision an equilibrium wage rate for a given situation. The worker must face the prospect that if he insists upon a certain pay scale and if the employer is unwilling to grant it, the job may be ended. Hence the worker seeking a raise must have some scale of values which leads him to prefer the risk of unemployment to continued acquiescence in the wage he is getting. It may be that he knows of alternative positions elsewhere or merely that he prefers to work for himself or to be idle. Decisions regarding such matters are often reached without careful weighing of alternatives.

In very general terms one may state, on the supply side, that when wage increases are being considered wage rates tend toward an unstable balance between the marginal desire for more income and the marginal desire for assurance of job security. The individual worker may be pictured as achieving maximum satisfaction for the moment either by accepting what he has of both wages and employment prospects or by demanding higher pay and weakening his certainty (?) of continuing at work. Perhaps the negative approach is better. He minimizes his dissatisfactions with his pay and his feeling of insecurity by accepting the status quo or by making a bid for a change. Once the demand for more pay is made, subjective and objective states will change. He may win the increase or he may be denied. He may become more contented or more dis-

contented. Worker attitudes are in a continuous state of flux. Their extreme instability should be contrasted with the attitudes of employers, whose measurements of pro and con are stabilized somewhat by the objective figures of income and outgo on the accounting records of the firm.

15. Organized Labor Should Also Count the Cost of Wage Demands.—When workers are organized, the situation is more complex. Some one speaks for the workers as a group, but he also speaks for himself. A true leader influences the views of his followers at the same time that he reflects them. A union official cannot often go contrary to the expressed wishes of a majority of his followers, but there is usually much opportunity to use his own discretion. Many a democratic organization is dominated by a vocal and aggressive minority. Rivalries may exist between elected officials. Personal considerations and objectives may influence a leader's interpretation of the wishes of his group. Generalizations about union wage demands are therefore more difficult to make than those which apply to individual workers. But they should be sought wherever enough similarities can be seen to warrant the belief that a rule may hold more than 50 per cent of the time.

For instance, if workers want more pay, should they risk a strike to get it? The decision may be made by leaders under blanket authority given them, or there may be a specific strike vote by the union members. They will be influenced by their answers to many questions. Is it likely that an agreement can be reached without having to strike? Would a strike eventually win the wages demanded? Or a lesser figure? How long would the strike last and what would it cost in lost wages? Would other unions be likely to contribute strike relief funds to the workers? If the wage increase is won, is the employer likely to discharge some workers and substitute others or install machinery instead? How much would this cost the organized group of workers?

The economist can offer some general precepts to guide the decision, but he should be under no illusions about their being used in actual strike situations. Emotion then becomes stronger than logic, or rationalization makes desires seem to be logical reasoning. There is also much difference between cases. For instance, if the demand for the services of the laborers asking higher pay is elastic, then a pay increase per worker would bring an income decrease for the labor group as a whole. Union leadership should consider this possibility before taking action. On the other hand, if the labor demand

is inelastic, union hiring halls or some other device may be used to spread the work among the entire membership so that individual and group alike benefit. Or restrictions upon entry into the union may gradually reduce its size as older members drop out. This procedure is limited, of course, to highly skilled craft unions and could not be applied to large industrial unions.

The real contribution of the impartial economist is to say to both sides that the cost of a strike or lockout is greater for each than they are apt to think when tempers get frayed and patience is almost gone. It is illogical to argue against strikes on the ground that the workers usually lose more during their period of unemployment than they gain by higher wages. Statistics could not be obtained to prove any such generalization. It is not sufficient to cite individual cases which seem to support the point, for the other side can tell of many cases which "prove" the opposite. In some cases the economist speaks for the rank and file of the workers against impetuous leaders. At other times the leaders have the greater vision and the followers need more caution. On the other side of the controversy, the interests and views of stockholders may be in opposition to those of the company officials. And in the middle stands the long-suffering consumer.

16. Time-Elasticity of Demand Following a Forced Wage Increase.—A major difficulty in reaching conclusions about the best course of action is that of deciding how long a time period to consider. When a wage increase is forced upon an employer by a strike or the threat of a strike, he may not immediately reduce his working force. He may have orders which he must fill or materials which it is economical to use up. His demand therefore may appear to be very inelastic. But after a few weeks or months he may be able to substitute machinery or other forms of labor for the type of labor whose wages have been raised under compulsion. Where unions are strongly organized and unemployment is small, the substitution of other forms of labor may be very difficult and time-consuming. In some cases the employer may find no satisfactory solution and may decide to close down entirely. In other words, the longer the period of time, the more elastic the labor demand is likely to be in the face of wage increases.

The passage of time may also influence the demand for labor by the reaction of consumers to the higher prices which employers often seek following wage increases. If purchases decline, production and employment must decline, too, and high-cost firms may be forced out of business.

Under some circumstances the elasticity of demand may decrease, not increase, with the passage of time for adjustments to the wage increase. This could occur when the immediate reaction of an employer is to reduce the number of workers hired either because he wants to vent his anger against them for forcing the wage boost, because he lacks the working capital for a larger payroll, or because he reduces his output in order to get a more profitable relationship between *MFC* (marginal factor cost) and *MFR* (marginal factor revenue). After he has cooled off, the employer's desire for profit may outweigh his impulse toward revenge, or he may see the light and seek to learn how to cooperate with labor and earn its goodwill. He may accumulate or borrow the needed working capital, or may succeed in raising the selling price of the goods made with the labor concerned. This latter outcome is likely whenever the wage increase is industry-wide (like that of the soft coal miners). The possibility that such a price increase may occur tends to diminish the initial contraction in labor demand.

There is also the possibility that the wage increase may improve the quality of the labor performed so that the employer gets more for his money than he expected, conceivably even more than he received before. This is a likely possibility when the wage increase is in the lower brackets and permits the workers to buy more food, to ride to work instead of walking, to have a happier home life, etc. It may result from a changed attitude toward the employer if he can convince the workers of his desire to work with them instead of against them, but such an outcome is improbable if the wage boost is the outgrowth of a protracted strike. Another possibility is that the prospect of loss may shock the employer into more vigorous efforts to improve efficiency and reduce costs.

Whenever any such improvement in labor efficiency does occur, the result cannot be interpreted strictly in terms of the demand-elasticity for labor, since the "labor" ultimately becomes superior to that obtained at the outset. If labor demand schedules could be constructed in terms of work performed per unit of pay received instead of wage per unit of labor time, then the picture would be one of a decrease in supply in the schedule sense, causing the initial rise in wage, followed by an increase in schedule supply as work output increased and the wage rate per unit of work dropped toward the original figure.

17. Workers May Accept Improved Conditions of Employment in Lieu of Wage Increases.—Increased wages are not the only demand made by organized workers. Improved working conditions

are often requested at the same time. These include all the things suggested above among the nonwage inducements that employers might offer (cf. Section 8). In addition there is the question of hours of work, a topic most frequently raised by the workers. They may want a shorter working week, or a different distribution of the hours of work throughout the day or the week.

Another worker demand is the change of rules regarding the minimum output per worker. This may involve direct reductions in output quotas or may seek the same end by requiring that more men be employed for the same task. The full-crew provisions of railway labor contracts are well known. Employers denounce them as "featherbedding." When unions are strong they may force employers to agree to hire "stand-by" workers who perform no useful service whatever, as in the case of the organized musicians.

18. When Should Employers Try to Cut Wages? And Workers Accept?—If we turn now to the downside of the wage-fluctuation problem, we note that the prime mover is the employer, not the worker. He may cut wages in either of two ways: either reduce the pay of those already employed or offer lower wages to new recruits. Even in periods of business decline the number of voluntary quits may exceed the necessary layoffs and therefore require some recruiting of new employees. To offer these workers lower wages than the rates paid to existing employees does not arouse the same opposition that occurs when, on the upswing, new workers are given more pay than those of greater seniority. Union rules offer the only serious obstacle. Even though it may be feasible to make wage cuts in this manner, there is also a danger. The new employees may feel that they need not work as hard as the better-paid workers of longer employment. This possibility of inferior work is also the major objection to reducing wages of existing employees. They may find ways of slowing up which express their resentment without making their shirking obvious enough to be grounds for dismissal.

If an employer adopts the expedient of cutting the wages of those already employed, it may be because he wants to increase profit, or because he needs to reduce costs to keep out of the red. Occasionally an employer may use wage cuts as a relatively painless method of reducing his labor force by provoking some to quit. However, the ones who leave are apt to be more energetic and resourceful than those who remain, so there is loss as well as gain.

Whether wage cuts are a method of last resort after all other possible economies have been exhausted or are the first thing the manage-

ment tries is largely dependent upon the organizational strength of the workers. *If there is a strong union, wage cuts are not likely to be proposed* until the management can see no other way out. Union leaders will probably demand proof of alleged business losses and must be shown that there are no other feasible methods of cutting expenses.

If the company argues that wage cuts will result in more jobs by reducing costs and thus prices, representatives of the workers ought to study the probable demand-elasticity for the products of their labor. It may be that a price cut would not increase sales appreciably. There is also the possibility that price cuts would not follow wage cuts or would not be as large as warranted by the reduced payroll.[4] That is why some unions seek conditional contracts which link wage rates with rates of profit. Workers on their part would seem obligated to give a pledge not to reduce working efficiency if wage cuts prove necessary.

If sales of a firm decline one-fourth, the necessary labor will decline about one-fourth. An employer might discharge 25 out of 100 workers. Or he might cut 10 out of the 40 hours each employee worked per week. If the union is dominated by older men who would benefit from the application of the seniority rule, discharges may be favored. If the younger men run the show, or if the older men are willing to sacrifice for the benefit of those more recently hired, the shortened work week may be advocated. The latter expedient is probably best for the community in the long run, particularly in times of depression, although there are many exceptions to this generalization.[5] The employer retains his work force against the day of rising demand once more. The junior employees are saved from having to go on relief. The senior workers probably contribute more through sharing the work than they would have had to pay in taxes or would have given to charity organizations for the support of the unemployed. But what they lose in dollars they can gain through satisfactions of a nonmaterial type, a greater feeling of unity with other workers, the bond of shared sacrifices, the gratitude of those helped, a better community for their children, etc.

Employers' ideas about probable future demand for their product will influence their wage and employment policies. The pessimistic employer who thinks demand will decline still further in the future

[4] The percentage reductions in wages and prices will obviously be unequal, for wages are only a fractional part of total cost.

[5] In 1933 the government tried to promote this policy first on a voluntary basis and later through NIRA "codes" given the force of law.

will cut his labor force more quickly than the optimistic employer who thinks the recession will be temporary. If demand for the product is thought to be fairly elastic, wage-rate cuts are more likely to be tried than a discharge policy. The employer will reason that lower wage rates mean lower unit costs, which will permit lower prices and that these, in turn, will increase sales so that he will need most or all of his working force. No amount of optimism, however, will overcome a shortage of working capital, which discourages the accumulation of inventory or prevents operating at a loss when sales decline.

19. Elasticity of Labor Supply Following a Forced Wage Decrease.—The reaction of labor supply to a wage cut is a function of many things including, first, the time which elapses between the wage reduction and the workers' decisions to change their status. If they continue to hold their jobs while looking for others, the decision to stay makes for zero elasticity at that time, but if some are able to find more attractive jobs after looking around, their decision to quit will make the supply curve for the longer period somewhat elastic.

The second determinant of labor supply-elasticity is the trend in business conditions and employment. If things are on the upswing, immediate elasticities will be somewhat greater and ultimate elasticities very much greater than in a period of recession. But in good times employers are not likely to cut wages anyway. Hence our chief attention should be devoted to the downswing. Here the major questions are whether the employees will quit immediately as individuals or by a concerted strike and further, whether they will change their minds and want to come back after the passage of time has depleted their financial resources and changed optimism into pessimism. When the latter happens, the elasticity of supply becomes less than it was at the outset. However it is not likely to become zero, for some workers will undoubtedly find jobs elsewhere and others will tend to shun a firm with a record of wage-cutting.

A third major factor determining supply-elasticity of labor when wages fall is the degree of union organization and financial strength. A very strong union may be able to prevent strikebreakers from crossing its picket lines and thus raise supply-elasticity to infinity, i.e., reduce labor supply to zero. Unemployment insurance and relief systems are a fourth determinant and serve to increase supply-elasticity when wages are cut. In fifth place might be mentioned forces influencing the mobility of labor such as efficient employment exchanges, vocational training programs, moving costs, etc. All of these determinants of supply-elasticity of labor may differ among

places, occupational groups, and time periods. Hence it is dangerous to make sweeping statements about *the* elasticity of supply of labor in the short run or long run.

There are, of course, a multitude of supply curves for labor, even for each type of labor. When, for instance, the managers of a certain firm use a labor supply curve in deciding upon changes in wage or hiring policy, they use one kind; when they are planning a new venture to take several years for completion, they use another. One might say there are at least as many supply curves for labor as there are individual people planning to employ labor in different places, times, and conditions. There are also the supply schedules as seen by the laborers themselves or their agents. Economists might add a large number of supply curves based on observed records of the past and on their own concepts of over-all supply in particular vocations, regions, types of labor, or phases of the business cycle. Finally, the discussion of this section has concentrated on only one directional type of labor supply curve, that in which the wage is falling. However, most of the comments are broad enough so that when considered in conjunction with Sections 4 to 7 above the reader should be able to work out his own supply functions for cases where the wage is rising.

20. Summary of Forces Determining Specific Wage Rates.— Although much of this chapter has dealt with special situations that arise from time to time in the labor market, certain generalizations may be derived even from an institutional approach. For instance, specific money wage rates may be said to depend upon three groups of forces which may be subdivided as follows:

I. Employers' appraisal prices, which in turn depend upon:
   - A. The customary wage rates in the area, the employer's own precedents, or wages paid by other employers.
   - B. The expected demand prices for his product in various quantities
   - C. The expected technical efficiency of labor (the production function)
   - D. The time urgency of profit prospects contingent upon getting labor

II. Workers' asking prices, which are a function of:
   - A. The customary wage rates in the area, the wage rates previously received, or the rates newly established for other labor groups.
   - B. The cost of living

C. Expected efficiency as workers
D. The strength of labor organization and the guesses of its leaders about the gains they can make or hold
E. The time urgency of money income contingent upon getting a job

III. Relative bargaining strength on each side, which depends upon:
A. The number of competitors (i.e., of substitute opportunities available to the other side)
B. The willingness of competitors to engage in wage competition in buying or selling labor
C. Legal maximum or minimum wages
D. Unconfessed time urgencies (not expressed in bids or offers)
E. Expected support or opposition from other groups (e.g., the public, or class allies)

Changes in wage rates result from changes in the variables under each of these three headings plus the time allowed for these changes to take effect.

Chapter 18

EQUILIBRIUM WAGE RATES: SOME THEORIES APPRAISED

1. The Theory of Equilibrium Wage Rates.—In Chapter 17 specific wage rates were shown to result from various demand and supply forces under modern capitalistic institutions. We now turn to the opposite approach and ask how these wage rates may affect labor demand and supply. In particular, we want to know when specific wage rates may be in "equilibrium." This is the problem of normal or stable wage rates. Since our economy is dynamic, wage equilibrium can exist only in the abstract. Wage theories often summarize the general conditions of this equilibrium under static, *ceteris paribus* conditions. Some of these deductive theories are very helpful in understanding the economic world in which we live. Others have led to erroneous conclusions because their premises were not realistic or inclusive enough. Both types of wage theories will be presented and evaluated in this chapter.

The basic principle of price equilibrium is readily stated. The specific price in the market must not of itself stimulate any change in either demand or supply. When a wage rate is the price under consideration, that wage must not cause employers to change their demand for labor nor workers to change their supply of labor. Temporary equilibria are established whenever a wage contract is signed or a verbal agreement is made. If this wage causes either party to change in the future the quantity of labor sought or offered, it is not an equilibrium wage for the long run. During the time required to make the change in labor demand or supply, other economic forces in the environment must be assumed constant. For example: a wage rate of $1.86 per hour for coal miners may be accepted by the owners as the result of bargaining on a certain date. They agree to pay this amount, but they make plans for the introduction of labor-saving machinery. When it is obtained, the mine owners reduce their demand for labor. This proves that the agreed wage was above the longer term equilibrium wage. But it would not be proved if, after the contract date, the demand for coal slackened and caused workers

to be laid off, if the price of coal-mining machinery fell to make its installation more attractive, or if any other change in the economic environment occurred to affect the demand for labor.

The questions of this chapter therefore may be grouped under the demand and supply headings so familiar in economic analysis:

I. How do specific wage rates influence the demand for labor?
   A. Demand equilibrium for one type of labor hired by an individual firm?
   B. Collective demand equilibrium for one type of labor?

II. How do specific wage rates influence the supply of labor?
   A. The supply offered by single individuals?
   B. The collective supply of one type of labor?

III. Why do different types of labor usually receive different wages?
   A. Demand differences?
   B. Supply differences?

IV. How true are certain wage equilibrium theories?
   A. Demand approach: the marginal productivity theory?
   B. Supply approach: the cost of subsistence theory?
   C. The bargaining theory?

V. How does the presence of labor-unions affect equilibrium wage rates?

VI. Why do wage differences persist among members of the same labor group?

2. Demand Equilibrium for One Type of Labor Hired by One Firm.—In earlier chapters we have seen that the individual firm is in equilibrium when its marginal cost equals its marginal revenue. The output which gives that combination is the most profitable one for that firm under certain assumed conditions of demand for the product and cost of production. The major part of our attention centered upon demand conditions and the *shape* of the cost curves. A change in demand was shown to alter the output which made $MC = MR$. Very little was said about a change in the *position* of the cost curves. To this subject we now turn our attention.

An increase in wage rates, for instance, would raise the variable cost per unit of output. This would raise also the MC and ATC curves and reduce the output rate which maximizes profit. Reduced output would lead to reduced inputs of all variable factors, including labor. In other words, an increase in the wage rate tends to decrease the equilibrium labor quantity for each firm, i.e., its demand for labor.

Another way of presenting the same kind of argument is to change back to the approach of Chapter 8, Section 7. This developed the

concept of diminishing returns of grain output per unit of fertilizer input. If we substitute labor for fertilizer as the variable input, the result will be similar. After a certain point is reached in most enterprises, more labor means less output per laborer. Both the average product and the marginal product diminish.

The diminishing average product may be called "Average Factor Product" (*AFP*), meaning the average product per unit of the variable factor used, which in this case is labor. The diminishing marginal product may be called "Marginal Factor Product" (*MFP*). Since we want to compare wages with the product of the labor for which the wages are paid, both should be expressed in dollar terms. Therefore the product must be considered sold and its value computed. This revenue from sale may then be called "Average Factor Revenue" (*AFR*) and "Marginal Factor Revenue" (*MFR*).

AFR and *MFR* curves (often shown as straight lines) usually slope downward for one or both of two reasons. The first was explained above as a manifestation of diminishing returns. In wage theory it is often called the "diminishing marginal productivity of labor." But we must note that there may also be a stage of increasing returns, and a turning point of constant returns. A review of Figure 26 in Chapter 8 will help to recall the fact that in processing and manufacturing there are often "flat" cost curves. This is the same as saying that the productivity curves have a flat top. In many enterprises the addition of workers of a given type does not change the output per worker, whether computed on an average or marginal basis. The *AFR* and *MFR* curves are horizontal straight lines, at least for the range of production that might be influenced by changes in wage rates (see Figure 60).

The second possible reason for a downward slope of *AFR* and *MFR* curves is the downward slope of the demand curve for the output of the firm. This was described in Chapter 4 as occurring rather frequently outside of agriculture and certain standardized raw materials. If the Average Product Revenue (*APR*, formerly *AR*) and Marginal Product Revenue (*MPR*, formerly *MR*) decline, the *AFR* and *MFR* curves must decline likewise even though $AFP = MFP$. That is, $AFR = AFP \times APR$ and $MFR = MFP \times MPR$.[1]

If, now, we assume further that the individual firm buys its labor at a fixed wage rate regardless of the number of workers hired, a

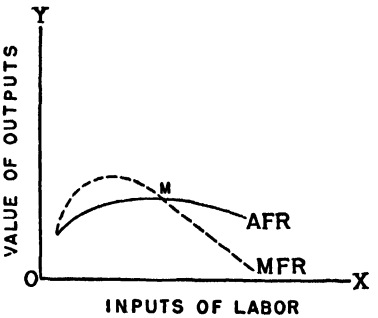
[1] The *APR* and *MPR* terms in the above formulae must themselves be averages if *AFP* and *MFP* represent more than one unit of the product, as is likely to be the case if the input units of labor are large.

maximum profit input point may be found from the intersection of *MFC* and *MFR*. When the firm is very large or the labor supply is very small, efforts to obtain more labor may force upward the wage

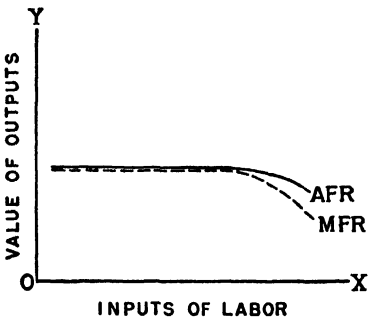
FIGURE 60

FACTOR INPUT PRODUCTIVITY AS VALUE OF OUTPUTS

A. Diminishing Value Returns in Agriculture and Some Manufacturing Enterprises



B. Constant Value Returns in Certain Manufacturing Enterprises



rate that must be offered. In this case *AFC* (the wage rate curve) is not horizontal, but rises. The *MFC* curve must rise more rapidly, according to the same principle that causes *MFR* to decline more rapidly than *AFR*. Where *AFC=MFC*, the ideal amount of labor to hire is shown by the point where *MFC* = *MFR*. Three possible

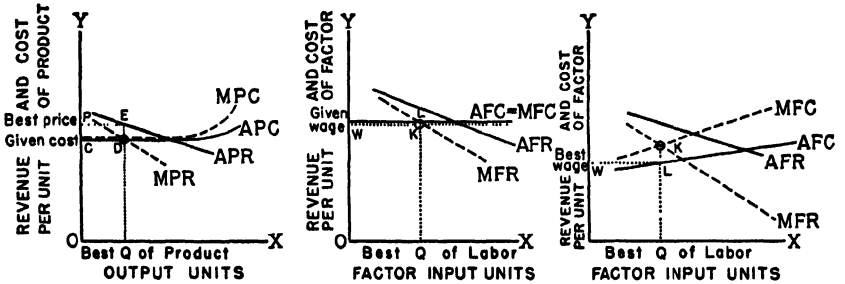
FIGURE 61

EQUILIBRIUM OF THE FIRM UNDER THREE ASSUMPTIONS
(Constant Marginal Product of Labor (*MFP*) in Each Case)

A. Price Varies with Quantity Sold (Wages fixed)

B. Fixed Wage (Price varies) *

C. Wage Varies with Quantity Hired (Price varies) *



* *AFR* and *MFR* curves would also slope downward if there were *diminishing* marginal productivity of labor and either constant or varying price of product.

solutions to this problem of the equilibrium of the firm in regard to labor are given in Figure 61 above.

3. The Usefulness of Labor Demand Equilibrium Formulae Is Limited.—The foregoing analysis is too limited in its assumptions to serve as a good guide to the actions of individual firms in the labor market. It appears to say that if organized workers were to force wages up, the marginal curves would intersect farther to the left and less labor would be hired. While true as a tendency, this influence is usually offset by other adjustments employers can make. Other costs may be reduced. Profits may be pared. Selling prices may be increased. Labor itself may become more efficient.

The marginal identity formula seems to show that wage reductions would increase the quantity of labor hired by the individual firm. This conclusion also is a low percentage generalization. There are many exceptions. Most firms hire only that amount of labor needed to produce the volume of output they expect to sell. Wage reductions would increase profit, but not sales, unless there were active price competition. Firms hesitate to engage in price competition for many reasons (cf. Chapter 12), such as possible adverse effects of retaliation by competitors. And when a whole group of competing firms cuts prices, the price-elasticity of demand for the product is much less favorable to the individual firm than if it alone makes the price reduction. Wage declines are less frequent and less permanent than wage increases. When they do occur, they often injure labor morale and reduce its efficiency, although voluntary reductions to keep a firm from closing entirely may meet with more favorable response.

Labor demand equilibrium formulae may also be used in another way. Assuming that the supply schedule of labor is given, the $MFC = MFR$ point indicates the *best quantity* of labor for the firm to hire. But here, again, the optimum labor input indicated by the formula is only a tendency. For many firms, the quantity of labor hired is determined by the production rate and that rate is a function of the orders on hand, not the firm's selling prices and wage rates. It is true that orders are a function of selling price, but other forces are also influential, such as selling efforts. In other cases the quantity of labor hired is determined by the equipment and the technology of the production process. A machine which is designed for one-man operation cannot be made to produce more by adding another operator to stand alongside the first. Technological factors of this type change very slowly.

In conclusion we must note that the equilibrium formulae for wage rates, labor quantities, or both are useful if their assumptions are kept in mind. They indicate the relation between certain variables. The omission of other variables is legitimate. But it is not legitimate to reason as though these other variables do not exist. On the other hand, one should not exalt the action of these other forces into another universal rule as limited as the $MFC = MFR$ formula. Sound reasoning and accurate prediction require that monistic explanations be discarded. The forecaster should consider the effects of many variables, including the background against which the specified variables operate.

4. Collective Demand Equilibrium for One Type of Labor.—The collective demand curve for a given type of labor undoubtedly slopes downward to the right. This negative slope would arise from cases of diminishing marginal productivity in the individual firms. But even where marginal productivity is constant, there are stronger and weaker demanders of labor. Maximum demand prices differ, and when they are grouped into a collective schedule the result is total demand varying inversely with the price. A change in the wage rate therefore tends to change the number of firms using that type of labor.

The equilibrium wage relative to collective demand is one which does not stimulate any change in the quantity of labor demanded by the employing group as a whole. Specifically, this means no change in the total demand which might result either from entry or exit of firms and no change from the expansion or contraction of existing firms.

5. General Demand Equilibrium Considers Roundabout Effects.—It is now necessary to introduce the roundabout effects of the wage rate upon the demand for the product of labor. There are two main possibilities. First is the effect of a wage rate upon the substitution of that type of labor for another factor, or vice versa. Factor substitution usually takes considerable time and therefore occurs only when there is a marked change in the unit cost or the efficiency of one of the factors. When rising wages or new inventions lead to the substitution of machinery for labor, a form of cross-elasticity is manifest. The increased demand for machinery may reduce the price of that machinery and this, in turn, may create a stimulus to further substitution in the same direction. On the other hand, if the price of the machinery rises, there will be a decreased amount of substitution for labor. Eventually, if no other disturbing changes occur,

an equilibrium will develop between these two classes of factors of production.

The second roundabout effect is still more complex. In its simplest form it is the way in which increased wages, for instance, may increase consumer buying power and thus lead to increased sales. Production will then be likely to expand and more labor will be needed for the job. There are many possible complications, most of which can only be mentioned here. The workers in a shoe factory may buy some of their shoes from the store outlets of that factory, but employees of a steam-shovel factory do not buy steam-shovels. To raise the wages of the latter could have only a very roundabout effect upon sales. Increased wages may be saved, or used to pay off debts. Imports may be purchased instead of domestic goods. Further comment upon this circularity must be deferred until the payment of other factors of production has been analyzed. Even then, the supply effects must be included, too, if we are to complete the picture of general equilibrium for the economy as a whole.

6. Specific Wage Rates Influence the Supply of Labor.—The supply of a given type of labor depends upon forces which are both objective and subjective. The objective determinants include the number of people in a given place and their abilities. These forces are basic, but it seems best to defer their analysis until after we have examined the subjective elements. The amount of labor which a given group of people is willing to offer depends upon their rational and emotional reactions to the situation. It will be influenced by the wage rate, the conditions of work, attitude toward the employer, personal savings, economic aspirations, etc. These determinants can be seen best by considering the possible reactions of individual laborers in different situations.

7. Labor Supply Equilibrium for the Individual Worker.—We may begin by assuming that the wage offered by the employer is the only determinant of the number of hours worked and that all other forces are held constant. We must assume further that the individual worker is free to expand or contract his workday or workweek at will. Under these conditions the labor supply equilibrium point for each individual will be that at which the increment of satisfaction from one additional unit of labor income is just equal to the increment of satisfaction from the amount of leisure time that must be given up to gain that income.

For instance, Joe Doakes may be unwilling to give up a ninth hour every working day for $1 per hour, his wage rate for each of the first

8 hours. But at the same time he may be willing to work the ninth hour if paid $1.50 for this 1 hour of overtime. He values his sixteenth hour of leisure per day ($24 - 8 = 16$) more than $1, but less than $1.50. Somewhere between these two figures is an equilibrium wage rate per *ninth* hour of work at which Joe would say it was a toss-up whether to work or to play. The money wage from a tenth hour of work would be less attractive than that from the ninth, while the desirability of a fifteenth hour of leisure would be greater than that of a sixteenth. Joe might still be willing to work this tenth hour for $1.50, or he might demand $2. His reaction depends upon such things as his physical energy, his desire to spend time with his family, his union affiliations, his acceptance of certain customs of wage payment, etc. A high percentage generalization may be made, however, to the effect that for hours of work beyond the customary amount, the supply curve for the labor of an individual worker slopes upward to the right.

We must now examine reductions in working hours with similar assumptions that workers are perfectly free to choose the rate of pay according to their reactions to different work loads. But we must also be realistic enough to assume that the workers in question have established certain customary levels of living (consumption of material goods) which they want to maintain.[2] The typical worker is not willing to work for a lower hourly rate when the workday is reduced. On the contrary, he usually wants a higher rate. It is more common for him to want to avoid any reduction in "take-home pay" than to desire more leisure at the sacrifice of money income. This is a valid generalization under the assumptions and is amply borne out by the history of labor wage rates in United States manufacturing firms after the second world war. In times of business recession, however, the picture is somewhat different. The worker is confronted with a third alternative, that of losing his job. In such cases he does his best to retain the going rate even though the working day is shortened. It is very uncommon for a worker to offer to work for less per hour if daily hours are cut from 8 to 7. Hours of work rarely approach zero, but if they did, a rising hourly supply price would almost certainly appear.

8. There Are Many "Typical" Supply Curves for Individual Workers' Labor.—The supply curve for a given type of labor by an individual laborer may now be drawn. The most significant portion

[2] These correspond to the customary levels of leisure implicitly assumed for wage changes upward.

of this curve is that in the vicinity of the customary workday and wage rate. The shape of the curve depends upon the direction in which wage rates are assumed to move. Figure 62 shows three possibilities. Curve *SA* shows a rise from existing rates in terms of *marginal* increments, like the *MPC* curve for the individual firm. Curve *SA'* shows the same thing mathematically translated into *average* rates per hour (8 hours at $1 plus 1 hour at $1.50 brings in $9.50, or an average of $1.056 for the entire 9 hours).[3] When business conditions are not good, overtime is unlikely. If it is demanded by a particular employer, workers may waive their customary right to time and a half (see *SA''*) if not under a strict union contract. Curve *SS'* shows what is most likely to happen to the individual labor supply if an employer attempts to reduce wages in good times. Employees refuse to work for him and strike or seek jobs elsewhere. When there is some unemployment and a recession threatens, workers may be willing to work the same number of hours at reduced pay, shown by *SB'*. If they have full freedom to choose, however, they may want to work longer hours, as indicated by *SB''*.[4] Finally, we may note that in good times when the declining variable is *work hours offered* by the employer, not wage rates, the individual supply curve of labor in terms of wage rates is likely to rise to the left, like *SC*. In bad times, reducing hours may bring no demand for increased pay, as shown by *SC'*. In short, there are many typical supply curves for a given type of labor of individuals. They differ with the direction of the wage change, the prevailing volume of unemployment, the power of unions, and a host of other things. And individual exceptions from general rules are, of course, legion.

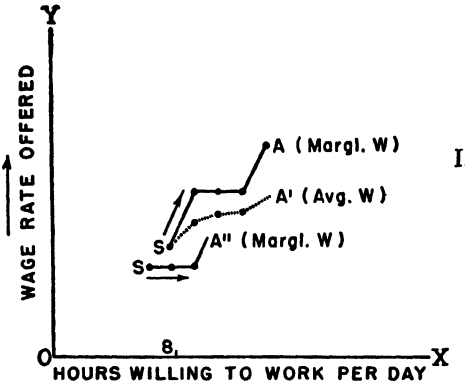
9. Labor Supply Equilibrium for Labor Groups.—A wage rate may be said to be in supply equilibrium at a certain place for any given labor group when it does not stimulate net entry or net exit of labor from that group. If changes in labor supply do occur, it indicates either that the wage rate is above or below the equilibrium wage or that some nonwage force is in operation, such as military conscription, a plague, government efforts to encourage entry into a certain vocation by offers of free training or early retirement, etc. Supply equilibrium for the group need not include labor supply equilibrium for all members in that group. An increase in the

[3] A curve computed in this way is quite different from the sort of curve to be presented below where, to attract *more laborers*, the hourly rate itself is raised.

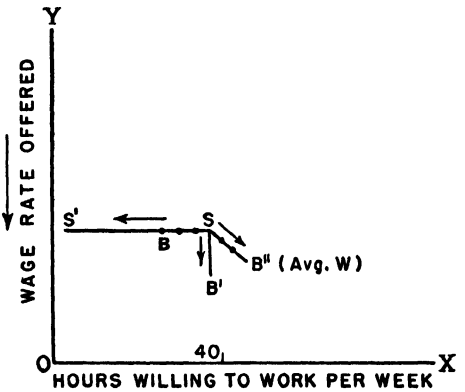
[4] This idea of a possible negative slope in a labor supply curve will be introduced again under the following discussion of collective labor supply.

FIGURE 62

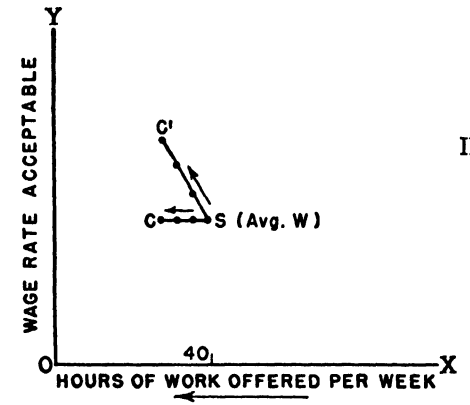
LABOR SUPPLY CURVES FOR INDIVIDUAL WORKERS



I. Wage Increases and *OVERTIME*



II. *WAGE* Decreases May Bring Various Supply Reactions



III. *HOUR* Decreases Bring Various Wage Acceptability Reactions

average number of hours members are willing to work might be offset by exit from the group.

The aggregate labor supply for a group is a function of the number of people in that group together with the number of hours per week, month, or longer time period which each member of that group is willing to work. The supply may be considered either as a number of hours at a given wage rate, or a series of supply possibilities for a given series of wage rates. Labor supply schedules are almost always of the successive type. They involve the expected or realized *changes* in supply following changes in wage rates. One must therefore remember to watch the starting point of these wage changes, their direction, their timing, and their amplitude.

10. The Collective Labor Supply Varies Directly with the Wage.—The collective supply of labor of a given group rises and falls as the wage rate rises and falls. There are several reasons for this. In the first place, the individuals who comprise the group differ in their willingness to work for any given wage. They will drop out of the group at different wage rates as those rates fall. Changes in the individual labor supply may also reinforce this direct relationship between wage rates and labor hours offered.

In the second place a rise in wages stimulates those in other groups to enter the given group. This occurs in four ways. First, there is the transfer of effort from one occupation to another, i.e., to the one made relatively more attractive by its rise in pay. If radio technicians, for instance, receive a wage increase and others do not, some electricians may shift from other jobs to this type. This may happen quickly if the occupations are very much alike, but it usually takes considerable time. People are reluctant to desert a given trade for one which may be only temporarily ahead in income prospects. They may prefer to stay where they are and try to get their own wages raised. The threat of transfer may be helpful in wage bargaining. There are also union rules which prospective entrants must overcome. Mature and experienced workmen may be required to take a long apprenticeship. High initiation fees deter entrants. There are many methods that may be used to keep green pastures for those already there.

A second method of entry is to migrate from a low-wage to a high-wage region. Such migration occurs chiefly between parts of a given country. Immigration restrictions and differences in language and culture greatly limit movement between countries. This internal or external migration may add to the supply of a given type of labor

immediately, as when Jewish doctors and scientists left Europe for America during the 1930's. Or the immigrants may need training and union cards like the country people who went to seaports to work in shipyards during the second world war.

The third method is the training of youth. Students in their teens are influenced in their choice of vocations by relative income possibilities. Many other factors are also involved, but in a pecuniary culture with considerable freedom of vocational choice, wage or salary possibilities are very important. Entry by this method, however, is necessarily slow.

Finally, very high wages for jobs requiring relatively little skill may attract women, children, and elderly people out of the home. This occurred very noticeably during the recent war. War fever and national propaganda also contributed to this movement, but large weekly pay envelopes were not unimportant. In general, the more highly paid occupations for a given level of ability attracted the most entrants of this type. Some of the elderly people were skilled workmen who had retired but returned temporarily to former vocations.

It should be obvious from the above discussion that the definition of an equilibrium wage rate for the collective supply of a given type of labor should state the time period involved. Even that does not help much. More useful is the concept of elasticity of supply relative to wage-rate changes. These changes may start from a hypothetical equilibrium wage or just from any prior wage. The cross-elasticity of supply is revealed by the transfer of workers from one occupational group to another.

The supply curve for labor of most types will probably have positive slope in the short run if that time period is defined in terms of individual willingness to work. This positive slope usually holds also for the moderately long run where changed attitudes influence entry or exit. One interesting exception to this generalization will be discussed in the next section. The very long-run supply of labor changes with the birth and death rates in different regions and economic groups. Economic aspects of this problem must be deferred until the next chapter.

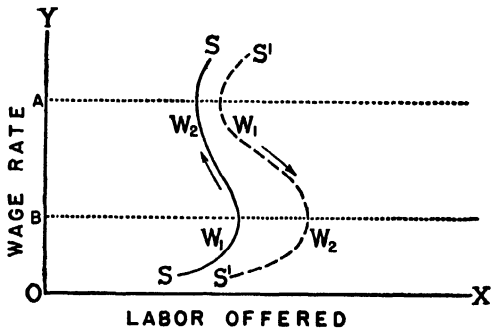
11. A Section with Negative Slope Appears in Some Collective Labor Supply Curves.—For some large groups of labor, particularly the unskilled and semiskilled, higher wage rates often cause a *decrease* in the amount of labor supplied. From an individual viewpoint this refers not to overtime pay, but to wage-rate increases on an hourly

or piece-rate basis. Many workers seem to prefer a shorter workweek more than higher pay. Or they split the difference, taking part of the gain in goods and part in leisure. From a collective viewpoint the increased pay of the head of the family may permit his wife to quit working or his boys to go back to school. Their exit from the labor market reduces the total labor supply in a labor group such as the "unskilled" which is large enough to include the father, too. In terms of narrow vocational groups it is a special type of cross-elasticity of supply. A wage increase in group A shifts workers not from B to A, but from B to C (i.e., to home or school).

Sometimes this phenomenon has been diagrammed as in Figure 63. The dotted line *S'S'* indicates that wage *decreases* may lead to a still more pronounced negative slope. The connection, however, is roundabout. A general decline in wage rates is usually found only during business recessions. At the same time unemployment grows.

FIGURE 63

A Negative Section Likely in Supply Curve for Unskilled Labor



The workweek is likely to shorten, not to lengthen. Pay envelopes become slimmer. The wage earner may even be reduced to receiving unemployment insurance or relief. In such cases, other members of the family often seek work to bring in supplementary income. They increase the aggregate labor supply in the general unskilled or semi-skilled group. Entry of this type is more pronounced than exit in the case of wage increases. Labor supply curves of this type have only a short negative section. At very high or very low wages, their slope is positive.

12. Different Types of Labor Often Receive Different Wage Rates.—We turn now to the question of why one type of labor nor-

mally receives more pay than another. We can ignore the question of the precise equilibrium wage for each group and speak in terms of norms or average wages for a group over a period of time. For centuries most judges have received more than most court clerks and most surgeons more than most nurses. Since wages are prices, the explanation of pay differentials must be found in demand and supply differences. Popularly stated, the labor demand is higher or the supply is lower. More accurately, the demand is higher at a given wage if the supply is the same in both groups at that wage; or the supply is lower at a given wage if the demand in both groups is the same. The next few sections will elaborate and illustrate this theme.

13. Reasons for Differences in the Demand for Different Types of Labor.—The demand for a given type of labor such as welders, machinists, salesmen, pilots, or bookkeepers depends chiefly, as has been said before, upon the ideas of businessmen about the profit that can be made by hiring them. Insofar as calculation of profit opportunities is rational, at least the following variables are considered: (1) the expected market demand for the product, expressed in terms of the quantity at a certain price, or a series of price-quantity pairs; (2) the expected technical method of production, expressed in terms of the factors to be used, some relatively fixed and others usable in varying proportions depending upon their probable prices and efficiencies; (3) the institutional restrictions upon demand such as monopsonistic agreements or practices among employers, maximum or minimum wage legislation, laws governing conditions of work, customs governing overtime pay, collective bargaining contracts, etc. Although hard to describe as rational, prejudice also enters the demand picture, as when employers dislike hiring women for certain jobs, refuse to take men over 40, or reject applicants who are colored or Oriental.

Back of the first two classes of demand determinants given above lie basic cultural conditions of the society in which the workers live. The demand for automobile mechanics in Detroit, for instance, grows out of the demand for automobiles from millions of people who have become accustomed to the possession and use of that luxury. We have developed knowledge of methods of mass production which lead to a demand for much unskilled and semiskilled labor instead of the more highly skilled labor that would be required were a smaller number of cars produced on a custom-made, hand-tooling basis. If in the future we demand airplanes in large numbers instead of auto-

mobiles, there will be some change in the types of labor demanded. The same will be true if technicians devise more efficient methods of producing automobiles than those now employed.

There is also another cultural determinant of labor demand that should not be overlooked, the opportunities for trade. The laborers producing a given article need not be the consumers buying it. Trade permits specialization which creates a demand for labor far different from that which would exist were exchange more limited. If trade were nonexistent there would be no commercial demand for labor at all! Each man would produce for himself alone. When trade exists, labor will be demanded to produce the articles for which a region is best fitted economically. Whatever affects the opportunities for trade, such as transportation costs, tariffs, foreign exchange costs, etc., therefore influences the quantities of particular types of labor that will be demanded in each given region.

The third class of causes of differences in the demand for different types of labor includes institutional developments, which often change quite rapidly. These are usually restrictive, such as anti-child labor laws, minimum wage laws, prohibitions upon the manufacture or sale of alcoholic beverages, and interunion disputes which may curtail an employer's ability to hire members of a rival union. Some institutional changes may also promote the demand for certain types of labor. Outstanding in this group are war and public works. The specific types of labor demanded under such spending programs will also be a function of technology in the art of destruction or construction, not to mention the technical aspects of making the weapons or the public improvements desired. The shift from oxidizing to atomic fission explosives is a recent change of major importance. Among institutional determinants of demand might also be listed such governmental policies as stimulate or restrain the general level of business activity in peacetime. Regardless of their cause, the different phases of the business cycle obviously influence the specific as well as the general demands for labor. Fluctuations in the demand for construction labor are a well-known case in point.

14. Reasons for Differences in the Supply of Labor in Different Groups.—The collective supply schedules of labor groups differ because of differences in the number of people able to do that kind of work and in their willingness to work for the wages offered. The number of people in a certain place at a given time is a result of past events. One would have to go far back into history to explain, for instance, the number of people now living in Los Angeles and still

further back to explain the present population of Paris. Births, deaths, emigration, and immigration have all played a part. At any given time the labor resources of an area resemble the industrial equipment and the technology because each is a cultural heritage. In retrospect they differ only from mineral resources in that the history of their origin is more recent. Looking toward the future, they have the additional difference of being expansible.

The population of one region differs from that of another not only in numbers, but also in skills. This, too, is a cultural heritage. Although there are some changes from generation to generation, the vocation chosen by most children closely resembles that of their parents. Most sons of professional people go to college and enter the same or other professions. The present generation of coal miners is composed largely of sons of coal miners. Farm children become farmers, and so on. In a dynamic and growing country like the United States there are many exceptions to this rule, but not nearly enough to invalidate it. Even the educational opportunities for a given generation reflect the desire of the older generation that their children should be like their fathers.

There is also another viewpoint from which the skills available at any given time may be dependent upon the past distribution of similar skills, but the connection is more difficult to establish. It is the distribution of innate abilities. These abilities differ from person to person and have some tendency to be inherited. The skills which people develop are partly the result of the abilities which they possess. Therefore the skill distribution of one generation reflects in part the ability distribution of that generation, which, in turn, influences the next. This is not meant to deny the greater importance of educational opportunities as compared to heredity in determining any given distribution of skills. The argument is merely intended to indicate that heredity must not be discarded entirely.

Willingness to work for certain wage rates is also a cultural heritage. Low-income families have low aspirations. Their children do not appraise their worth to employers as high as do rich children of no greater ability or educational achievement. The wage customs of vocational groups are very strong in some countries. This is particularly true for *relative* status. Wages for a given type of labor may rise and fall considerably with the price level, but they will not change very much in relation to other wage rates.

The foregoing comments regarding differences in the supply of labor supplement, but do not contradict, the arguments of Section 8 about the elasticity of labor supply. They merely furnish the start-

ing point relationships. When changes in demand alter wage rates, all sorts of supply changes may occur. Willingness to work may change; workers may migrate; new skills may be learned; and even the birth rates and death rates may be altered. All these comments, and many more, are needed to paint a complete picture of the labor supply differences which help to explain wage-rate differences among labor groups.

15. The Cost of Production Theory of Differential Wage Rates.—Cost of production differences are urged, by some, as the best explanation of differences not only in commodity prices but also in wage rates. The argument is partly true, but it accounts for only a portion of the supply-side approach and ignores the demand side entirely. The analogy with commodity prices breaks down at another point. The parents who pay most of the cost of rearing and training children have no expectation of getting all their money back. They are not in the business for profit. Nor are the childless taxpayers who contribute to the support of public schools.

The most that can be said for this theory is that it accounts for *some* of the differences in the supply of workers. Training for some vocations is obviously less costly and therefore open to more people than for other vocations. Or suppose the voters decide that more doctors are badly needed. Laws might be passed guaranteeing free tuition and living expenses for all who wanted to train for that vocation. If the cost of other educational opportunities remained unchanged, one would expect an increase in the number of doctors relative to those in other professions. The resultant decline in the typical doctor's earnings might then be ascribed to a change in the cost of production of doctors.

Some classical economists argued that the wages of manual laborers could not rise for very long above the cost of subsistence of those workers. If they did rise temporarily, the death rate would fall and the total number of workers would rise until competition forced wages down again to a bare "cost of production" level. The theory states an obvious tendency. But it overlooks the changing psychological and diet habit bases of what is called minimum subsistence. The cost of production of an American laborer is much higher now than 100 years ago and very much higher than that of an African kaffir or a Chinese coolie.

16. The Marginal Productivity Theory of Wage Rates: a Critical Appraisal.—Under certain conditions the differences in wage rates between different groups of workers may be explained by the

so-called marginal productivity theory of wages. To be applicable, this theory requires that the demand for labor be expressed as a schedule or curve measuring the collective diminishing marginal value productivity of labor as seen by the group of employers involved.[5] The supply must be homogeneous and unorganized. The labor group must be defined as the total number which actually find employment.[6] Under these circumstances the wage rate will be determined at the point of intersection of the demand and supply curves.

This point then *reveals* the marginal productivity of the workers employed, but for various reasons the reverse interpretation has become widely accepted. The marginal productivity of the group is said to *determine* the wage. One reason for this often misleading interpretation is the fact that if the marginal value product of the workers rises because of higher prices for the product from which the demand for labor is derived, then the marginal productivity curve rises and lifts the equilibrium wage rate correspondingly. Another reason is that if the size of the group seeking employment is increased, the supply curve shifts to the right and the intersection with the demand curve comes at a lower point, i.e., the marginal productivity corresponding to the equilibrium wage rate is lower. (The converse would also be true in each of the last two illustrations.)

The marginal productivity theory of wages has very limited applicability because it is based upon assumptions which rarely apply to current wage situations. However, it may be used to explain in part why large groups of labor get lower wages than small groups, such as the unskilled as compared with the skilled. The theory here renders a service by showing that the size of the group is more important in wage determination than the skill of the workers. As the number of workers in any occupation rises, the pay will tend to fall, even though there is no decrease in the skill of each individual worker.

There are some cases in which a larger group has a greater demand for its services than a smaller group, with the result that the favorable demand differential for the larger group offsets the adverse force of its size. Automobile mechanics are more numerous and get more pay than blacksmiths; radio announcers get more than hog callers.

A second possible use of the marginal productivity theory of wages will be demonstrated in the next chapter, which endeavors to

[5] This has been more precisely defined above as marginal factor revenue. See Section 2.

[6] Another alternative when comparing groups would be the stipulation that an equal percentage of each group finds employment.

explain differences in real wages among various regions. In dealing with very large groups of workers, the requirement that supply must be homogeneous becomes far-fetched, but very useful conclusions may be reached.

17. The Influence of Skill on Wages.—The correct explanation of the influence of skill on wages may be seen by distinguishing between two different situations. In the first it is apparent that a journeyman carpenter is paid more than an apprentice, or a speedy and accurate stenographer more than a novice. Here the comparison is between *different degrees of the same type of skill.* Since pay is in proportion to work performed, the more highly skilled worker gets more pay per hour than the less highly skilled. But in the second situation, the carpenter must be compared with the stenographer, and who is to say the one is more skilled than the other? It is impossible to erect a satisfactory explanation of relative wage rates between labor groups by comparing the *degree of skill in different vocations.*

A comparison may be made between the time required to acquire the skills of each trade, but the cost of production argument is not enough. One must explain also the relative demand for each type of labor and the size of each labor group. Long years spent in training do not guarantee that the skill will be demanded once acquired (cf. wheelwrights) or that the skilled group will be so small as to command high pay for individual members (cf. public school teachers, or shipyard welders after a war). Not all people have the ability, the opportunity, or the financial resources required to get certain scarce skills, and in this sense skill influences wages by limiting the size of the group. But it is a roundabout argument that really supports the rival approach through marginal productivity.

18. The Bargaining Theory of Wages: a Summary and Appraisal.—Another explanation of relative wage rates among groups says that they are in direct proportion to the bargaining strength of the worker and in inverse proportion to the bargaining strength of the employer. This theory has several advantages. It brings both demand and supply forces into the picture. It says to each of the contending parties: "You can improve your position by increasing your bargaining strength." They know from experience that it is true. The theory avoids the fatalism of either the marginal productivity or the cost of production theories. It is also more realistic.

On the other hand, the bargaining theory may be misleading and is clearly only part of the explanation needed. It tends to make

people think that bargaining power is something that either side can increase at will. It ignores the fact that bargaining weakness may be due to the presence of forces very difficult or undesirable to change. An employer's bargaining position with a given group of skilled workers is weakened by the presence of high profits, which are otherwise very desirable. A union's bargaining position may be weak because of the large volume of unemployment of its membership, a condition that is usually completely beyond the union's control.

The bargaining theory needs to be supplemented by explanations of the size of the labor group, the technical substitutability of one factor for another, the relation between labor demand and product demand, the interfirm competition of employers of labor, etc. All of these forces must be understood if one is to know why wage rates are what they are and why they change as they do. The bargaining theory is merely the capstone. That is why the other partial explanations were given first and it was saved for the last. With this general statement of the advantages and disadvantages of the bargaining theory we may now proceed to examine specific acts by which the one side or the other may increase its bargaining power.

Workers may improve their bargaining position in many ways. Individual workers may form unions and local unions may affiliate with others in the city, state, or nation. Alliances with unions of other types are often helpful, as through sympathy strikes and boycotts. Unions also accumulate funds which are useful in case of strikes to feed workers and their families and thus reduce the economic pressure to return to work. Union funds may be used for organizing other workers, for educational campaigns to increase group solidarity, and to work for favorable legislation. The latter method is one of the most important because laws govern not only the right to organize and to strike but also the methods that may be used in bargaining. That is why unions oppose legislative restrictions upon the right to negotiate closed-shop contracts and to engage in sympathetic strikes. They oppose the use of the injunction in labor disputes. They favor unemployment insurance for their own direct gain when out of work and also because it discourages unemployed workers from taking positions as strikebreakers.

On the other side of the picture the employers also use many methods to improve their bargaining strength and to weaken that of the workers. They seek laws to outlaw the closed shop, ban sympathetic strikes, get the right to use injunctions, etc. In case of strikes it is very important to get public sympathy. This may involve securing newspaper publicity which paints the workers' demands as unreason-

able. Knowing that the general public condemns acts of violence and the destruction of property, employers occasionally employ men to pose as strikers and to urge or to perform such acts of violence. Another employer tactic is to hire people to promote dissension among the workers and weaken their solidarity.

Against this background, the bargaining theory seems to reduce to the statement that *wage changes* are best explained by acts which change relative bargaining strength. Certainly it does not *fully* explain the differences in *relative wage rates.* Locomotive engineers are as well organized as any group, but they do not get as much as airplane pilots. Organized coal miners do not get as much as organized musicians. There are also occasions where no bargaining is involved at all. Sometimes employers voluntarily offer more pay to their workers in order to keep them or to attract others. Skillfulness in bargaining at times may cover up weaknesses which the other side does not perceive. In short, enthusiasm for the many merits of the bargaining theory must not be allowed to make it *the* theory of wages. There is as much danger in its monism as in singling out any other partial theory for elevation to a place where it excludes other explanations.

19. Unions May Strengthen the Demand for Their Services by Opposing Substitutes and by Aiding Employers.—One of the most interesting developments of recent years in the labor field has been the increasing efforts of labor-unions to influence the *demand* for the labor of their members. They feel they have done all they can to raise wages by action on the supply side of the labor market. The next step is to reduce the elasticity of demand and to increase its intensity. There is some parallel here with the activities of monopolistic sellers of commodities who seek to raise prices by similar techniques. In the labor field, the action is of two main types: to decrease the demand for substitutes and to help employers pay higher wages.

The antisubstitute campaign has three objectives, to oppose substitute labor, substitute methods, and substitute products. It proceeds by various tactics, chiefly exclusion and cost-boosting. For instance, unionists traditionally oppose nonunionists who may want to take their jobs as strikebreakers, as rival bidders for jobs, or as candidates for union membership. This is the exclusion tactic. It has also been employed for more than a hundred years against the use of machinery or other labor-saving devices. In recent decades a supplementary tactic has been developed in the form of "featherbedding" rules by which the employer is forced to pay the partially displaced worker

for as much time as it would have taken him to do the job in the old manner. For instance, faster trains do not save the railroad anything in terms of pay for the train crew because engineers, conductors, and brakemen generally get as much for 4 to 6 hours' work on the fast trains as they received on the same run formerly taking 8 hours. The gains from speedier schedules are therefore reduced, or the cost of introducing the labor-saving improvement is increased. From either viewpoint union action is calculated to discourage the employer from introducing the labor substitute.

Exclusion and cost-boosting tactics are also applied against rival goods in the production of which labor belonging to other unions (or to no union) is employed. Rivalry in the building trades, for instance, makes carpenters oppose the use of prefabricated houses produced by members of other unions and leads them to get city ordinances passed which either prohibit the erection of prefabricated dwellings or make them so expensive that they offer no economy. If that fails, the organized carpenters, and other building trades affected, may refuse to work for any contractor erecting them or may set up picket lines which other union workers refuse to cross. Another illustration of major importance is the cooperation which unions have given employers in efforts to raise tariff rates, secure interstate trade restrictions, and even city laws increasing the cost of goods "imported" from outside.

The cooperation against substitute products extends also to efforts to help employers pass on the cost of higher wages to consumers. This may take the form of pressure on Congress, the OPA, and the wage-regulating agency in time of war restrictions. In peacetime it may involve seeking industry-wide organization of both labor and employers with simultaneous or joint contract negotiation. In this way no one employer is put in a bad competitive position when he is forced to pay higher wages and all have the same incentive to raise prices at the same time.

One should also note a second group of cases in which the unions have influenced the demand for their labor by improving its efficiency. This involves conferences with employers about production methods, improving morale, job-training, etc. The Amalgamated Clothing Workers have a good record in this respect. Obviously employers can pay more if unions help workers to produce more.

A final comment on the monopolistic practices of organized labor might be that there seems to be very little that might be called oligopolistic activity. Unions rarely engage in competitive selling activities by making wage cuts corresponding to the price cuts of merchants

and manufacturers. Therefore they do not have to consider what would happen to the demand for their labor if rival sellers were stimulated to cut wages, too. Employers on occasion have tried to play off individual workers against one another in such a way as to raise this problem, but the position of unions has generally been too strong.

20. Wage Objectives of Union Labor May Vary.—When a business firm operates monopolistically in setting the price for its product, a statement of the maximum profit price in formal terms is not difficult, but a similar statement for union wage policy is much more difficult. The business firm may logically be assumed to seek maximum net income in the total sale of its product for a given period. Unions are more often inclined to seek maximum gross return per unit of labor sold. They usually want the highest possible wage rate per hour or per week. Most unions give less attention to annual wage rates and only rarely do they plan how to maximize gross income for their entire membership considered as a group. In formal terms, they do not give much attention to the elasticity of the long-run demand curve for their labor. They are more apt to be concerned with the short-run bargaining range and to let the future take care of itself. And a bargained wage is notoriously unpredictable.

With increasing maturity, unions depart somewhat from the foregoing practices and a few show true long-range vision as indicated in the preceding section. But there is no uniformity. The most that an economist can say is that if union leaders seek short-run maximum wage rates they will probably strive for a higher figure than if they have long-run perspectives. The same is true if union policy concentrates on high hourly rates for the individual worker instead of maximum annual earnings of the group as a whole.

There is also the problem of those unions which are run (like some business firms) chiefly for the power and income of their officers. In such cases the wage policy is no more predictable than commodity price when profit maximization ceases to be the dominant motive. Racketeering obeys no formal rules. Recent demands for special retirement plans and medical aid further complicate the picture and make intergroup wage differentials greater than they appear from wage comparisons alone.

21. Wage Differences among Members of a Given Vocational Group.—In addition to the problem of wage differences among vocational groups which has been studied in the preceding five sections, there is also the task of explaining wage differences among members of a given group. These develop chiefly when workers are not or-

ganized, but may exist for the same type of work performed even by organized labor in different cities or firms. This is due to several peculiarities of the labor supply. First, inertia or immobility keeps workers at a given job even though the pay is higher next door. Workers like to work among old friends and acquaintances. They may even like their boss! Second, transfer to a new job may involve costs or losses which offset the gain from increased pay. Chief among these are the expense of moving and the loss of seniority. The latter may carry certain measurable retirement benefits together with the important, though intangible, assets of promotional opportunities and security against unemployment in bad times. Third, workers often lack knowledge of greener pastures elsewhere. Or if they hear about higher paying jobs, they may be suspicious about working conditions. Fourth, the other jobs may have a slightly different title or set of duties from that which a worker thinks he wants. This is one reason why even in the same firm there may exist wage discrepancies which cannot be justified upon grounds of differences in work performed or skills needed.

Intragroup wage differences may also develop because of slight differences in output per man day. Unions oppose such pay differentials, but employers often use output bonuses for their incentive value. This sometimes leads to reclassification by subdivision of a vocational group more detailed than the traditional one of apprentice, journeyman, and master craftsman. Nepotism and other forms of favoritism are all too common. Employers are also guilty of prejudice, such as that against women, married women, Negroes, Jews, Germans, Japanese, or men with gray hair. Where prejudice is community-wide, like that against races or women employees, the group discriminated against often becomes resigned to its fate and accepts as a self-valuation the relatively low figure named by the employers.

22. Summary of Equilibrium Wage Theories.—Following the introductory chapter which gave the institutional approach to wage determination in particular instances, this chapter has presented a more formal statement of the principles which explain wage rates in general. The problem was stated as one of explaining wages as prices and therefore the special circumstances of demand and supply were classified as follows:

I. The demand for a given type of labor is a function of:
   A. The expected intensity and elasticity of demand for *products* requiring that labor

B. The technical methods of production (which influence marginal factor revenue schedules and interfactor substitution)
C. Existing institutional obstacles or aids

II. The supply of labor of a given type is a function of:
A. The total population in a region (past birth and death rates plus migration) and their age distribution
B. Ability to perform that type of work (includes cost of training)
C. Willingness to do that work (many subjective factors involved)
D. Whether labor is organized or unorganized
1. The type of union activity
2. The objectives of union leaders

Both demand and supply curves shift through time as adjustments are made to changing wage rates. Long-run curves generally have greater elasticity than short-run curves, but the latter are the ones which govern particular wage rates. Once a wage is determined for the members of a group, the marginal productivity of a typical worker may be defined, not vice versa. The marginal product is only a point on a schedule and cannot be determined in the absence of a supply schedule. Efforts to explain relative vocational wage rates by reference to skill fail to consider other determinants of the supply schedule and ignore the intensity of demand. An exception in favor of the skill approach may be made for comparative wage rates of workers performing the same type of labor.

When demand and supply curves are discontinuous in the neighborhood of their intersection, bargaining will set wage rates and the outcome cannot be predicted. The bargaining range, however, is set by general demand and supply conditions which are only partly subject to control by the rival groups. The position of the wage within that range is a function of the relative strength of the bargaining parties. Union organization aids labor in bargaining, particularly by limiting the size of the labor group. Sometimes unions try to raise the demand curve by curbing the use of substitute products, labor, or methods. Worker efficiency may be increased. Employers also organize to improve bargaining strength. Both sides seek public approval and government support. For these and other reasons, wages of organized labor are indeterminate in the sense that they cannot be described by any single principle, neither marginal productivity, cost of production, nor relative bargaining strength. In more advanced studies, several type situations are defined and used in more detailed analysis than is possible in the present volume.

Chapter 19

COMPARATIVE REAL WAGES

1. Statement of the Problem.—Another problem of wage theory is the explanation of differences in *real* wages between different regions and different periods of time. We want to know whether the level of living is higher in one country than another, and why. Or we wonder how much a group of workers has benefited by its wage increases in the last ten years. This cannot be found by comparing money wage rates alone. For instance, machinists who now earn $3,500 per year may or may not be better off than when they earned only $2,000 per year. It depends chiefly upon whether the cost of living has risen less than 75 per cent or more than 75 per cent. Many other things also must be considered in making comparisons in this field. This chapter will consider the more important ones in answering the following questions:

1. What are the basic variables involved?
2. What determines the *general level* of per capita prodictivity?
   (a) The Law of Proportionality in factor ratios.
3. How explain the *changes* in real wages of a given labor group?
   (a) By changes in per capita productivity through the positive and negative effects of:
      (1) Capital accumulation?
      (2) Inventions?
   (b) By changes in bargaining strength and skill?
   (c) By changes in the labor supply?
4. What do statistics show regarding the effects of changes in the foregoing variables?
5. How explain the *differences* among average real incomes per capita in different countries *at the same time?*
   (a) What are the nonlabor factors of production which differ in quantity and quality from region to region?
   (b) How are levels of living affected also by:
      (1) War and war preparation?
      (2) The rate of capital accumulation?
      (3) The distribution of wealth?

2. Definition of Terms and Statement of the Basic Data Needed. —*Real* wages may be defined as the amount of goods and services that money wage earnings can buy when spent at the prevailing level of prices for a given consumption budget. To compare real wages in different places or at different times, we need to know (1) the average money earnings per family in each group, (2) a representative consumption budget appropriate to both groups and stated in terms of physical quantities of goods and services, and (3) the retail prices of the items in that budget for each place or time period.[1] Each of these data problems must now be examined.

The average amount of money earnings per family seems a more logical starting point than the wage rate per worker. It includes the full- or part-time income of all members of the family considered as a consuming unit. Annual earnings are probably better than monthly or weekly earnings. The length of the typical workweek may be less now than it used to be. Or the coal miners in one region may have only seasonal work while in other places most of them are employed all the year round. Statistical difficulties, however, often force us to use data which are not as good as we should like. Hourly wage rates for individual workers are usually easier to get than average annual earnings per families. Fortunately, for most purposes their use does not seriously distort the conclusions.

Standard consumption budgets are still more difficult to obtain. Statisticians have not been so active in this field as in that of money wages and family earnings. It is very hard to determine precisely the typical consumption budget of a family of long ago. We must rely chiefly upon diaries and autobiographies which are far from perfect sources of statistics. Since 1920 scientific budget studies at different income levels have become more frequent. It is almost impossible to get a single consumption budget which is fairly representative of two groups when they represent different culture patterns. The patterns of life which are accepted as normal or desirable differ greatly between Orient and Occident, urban and rural communities, warm and cold countries, or even two time periods a century apart. The cost of fuel for heating should be a large part of a Maine budget, but not of a budget for Florida. Most Americans are not interested in the rice and fish diets of the Japanese, nor would automobiles be very attractive to Eskimos.

The third set of data is more easily obtained than the first two.

---

[1] If the two regions are in different countries, we must also know the exchange rate that should be used in comparing money prices and wages.

Retail prices in different regions can be surveyed fairly quickly as needed. Many countries publish cost-of-living indices which have sufficient continuity for use in comparing two time periods. Newspapers, books, and court records give major commodity prices far back into history. Many of the myriad items in a consumption budget are not very important and their price differences may be ignored.

Because of the data difficulties outlined in the preceding paragraphs, economists use various devices to simplify the problem. Sometimes they compute how many hours a certain type of worker must work to buy a single article, such as an all-wool suit or a pair of shoes. The highest level of living (real wage) is said to exist in the country, or the time period, where the fewest hours of work are required for such a purchase. At other times economists group all the people in a country together as consumer-producers regardless of their actual roles in the economy. The total national income divided by the total population will then give a per capita income figure whose spatial or temporal differences may be measured. Subtractions can also be made for the effects of war or unusual capital accumulation before reaching final conclusions as to relative real incomes. If there are marked differences in the distribution of income, the data will need further correction before they reveal the status of the working class or a certain labor group.

The ultimate form of simplification is to describe the independent variables one by one and tell how each affects the results. This is the procedure adopted in the remainder of the chapter. The first main topic is the explanation of changes in real wages which occur from one time period to another. The second is the difference in levels of living in different countries (or regions) at the same time.

3. Per Capita Productivity Changes with the Quantities of the Factors.—Except for war periods, the chief reason for changes in levels of living is changes in levels of output. Per capita consumption is closely related to per capita production. The money income received by workers, capitalists, landlords, etc., is spent to buy the goods produced at the time the income is paid out. There are some time lags and cyclical variations in investment, but it is true enough for our purposes to say that current consumption roughly equals current output.

Two other approximations are also useful. Changes in the national income occur much more rapidly than changes in the population and therefore may be said to reflect changes in per capita productiv-

ity.[2] Finally, since workers comprise by far the largest segment of the population, changes in per capita productivity may be presumed to indicate changes in real wages. Even though some people live without working, most of them perform some kind of labor for all or part of their incomes.

These assumptions narrow the scope of our task. We must now explain changes in per capita productivity. It is logical to make the first approach to our answer by using the principle of diminishing average returns. This principle was first used in Chapter 8 to help analyze the cost curves of the individual firm. Here it is applied to the economy as a whole. After a certain point is reached in the expansion of a region's population, additional workers will reduce the per capita output. This point of diminishing returns has been reached and passed in most if not all of the countries of the earth. At present the average output per worker is *inversely proportional* to the number of workers. The more the people, other things being equal, the less the output per person. At the same time, the average output per worker is also *directly proportional* to the quantity of other factors with which labor works to produce the total product. The more the capital with which each worker works, the greater his output will be.

4. The Law of Proportionality and Some Applications.—The direct and inverse relationships described above may be summarized for all factors under the so-called "Law of Proportionality." This law may be stated formally as follows: *When one factor, A, increases in proportion to other factors, B, C, etc., being combined with it, the average productivity of A falls and the average productivities of B, C, etc., rise.* The law could also be rephrased to apply to marginal products and the point of diminishing marginal returns. Both statements of the law require that there be enough of the variable factor present to push production past the relevant point of diminishing returns.

If production were to take place with fixed proportions of the factors, then each would be essential to its complement and no change in their quantity ratios could occur. Thus in certain chemical processes like the production of calcium chloride by treating calcium carbonate with hydrochloric acid neither chemical alone will make the desired product and an excess of either will also be useless. A less perfect illustration might be found in the field of labor, where to

[2] The changes in value terms must, of course, be adjusted for changes in the price level.

build a modern house the services of a group of specialists are needed, each one complementing the other. The carpenter cannot build the house without the aid of the plumber, electrician, painter, mason, etc. There must be at least one of each, or one person must act in a second or third capacity. There are also cases where at least two of a given kind must be present to help one another and the point of diminishing returns does not set in until the third member of the group appears.

Applied in the abstract without consideration for human values, the Law of Proportionality leads to some interesting conclusions. Densely populated portions of India or China would benefit if their capital equipment were doubled rapidly. Applying the same Law of Proportionality, they would also benefit if their population were decimated by plague. Like the Black Death of the Middle Ages, a great plague would bring anguish to many hearts, but its purely economic effects in agricultural areas should be beneficial. The average product of the survivors should be raised by the proportional decrease in the number of workers. The poorer grades of crop land could revert to pasture. The same benefits from a plague would not be so certain in an industrialized country. There it might be necessary to make expensive changes in machines and methods if the labor force were suddenly reduced one-tenth. Rapid emigration might have an effect quite similar to a plague and would cause less sorrow.

If population were to increase rapidly through immigration, per capita outputs would fall. Real wages and living levels would decline. The same would be true for a slower increase through a falling death rate or a rising birth rate. These methods take a long time. In the meanwhile there might be offsetting improvements in technology or increases in the stock of capital goods. And, of course, a reduced level of consumption is possible only if at the outset people are living above the subsistence level. If workers already are starving, one would not expect immigration to occur, nor the death rate to decline.

Factor classes should be subdivided for the best application of the Law of Proportionality. Certain types of workers or equipment may be very scarce at the same time that the class as a whole is too abundant. There are certain countries like India whose population is so fecund as to add continually to its unskilled laborers when what is most needed is people with engineering training, management ability, and manual skills. The population should not be considered as a unit and only its aggregate increase considered, for obviously the country would gain, not lose, by the immigration of persons

possessing the knowledge and skills which are most scarce. And a similar comparison might be made of different types of capital equipment. An increase in farm tools might be much less stimulating to a region's per capita productivity than an increase in electrical generating capacity or transportation equipment.

The foregoing arguments may be summarized by saying that it is to the interest of any given labor group to encourage complementary factor groups to expand and particularly those factors which are most scarce in the light of prevailing or potential technology. In other words, every self-seeking group should stimulate the expansion of those other factors whose marginal productivity is highest. On the other hand, if some factors are partial substitutes for others, their increase should logically be opposed by the rival group even though in the economy as a whole the substitute has a higher marginal productivity than some other factor. That is why certain unions have opposed the introduction of labor-saving machinery and at the same time labor in general has benefited by its introduction. The Law of Proportionality may be summarized by the statement that that factor gains whose (service) quantity diminishes relative to other factors, and that the same factor loses when the reverse is true. How much it gains or loses depends upon the degree of relative change and the productiveness of the factors whose quantities change.

5. Capital Accumulation Increases per Capita Productivity.— When we compare average real wages in the United States today with those of a hundred years ago, we find that they have increased several fold. The Law of Proportionality offers a clue to the explanation. Population must have increased at a slower rate than the other factors. Capital goods have been accumulated very rapidly, and they have been improved by numerous inventions. In the United States the saver and the inventor have won the race against the stork by a wide margin. In certain other countries the contest between these positive and negative forces has been more even. The following paragraphs will explain some of the reasons.

Capital accumulation in the physical sense occurs when there is an addition to society's stock of machines, buildings, railroad trackage, office equipment, etc.[8] Some of the currently available land, labor, and materials are used to make these production goods more rapidly than they wear out or are discarded as obsolete. In addition to pro-

[8] The monetary background of savings, investment, and capital formation will be explained in a later chapter. At this point we are concerned with societal saving in terms of goods, not individual saving in terms of money.

duction of equipment in excess of replacement needs, there is continual improvement in quality. The output per worker is increased either by giving him more machines with which to work or better ones. The term *capital accumulation,* therefore, should be broad enough to include all changes in capital goods which make man's efforts more productive, whether quantitative in the accumulation sense or qualitative in the performance sense.

The accumulation of capital goods has economic, psychological, and political foundations. It tends to be most rapid in countries which are already rich in capital equipment, thus keeping them ahead of those vainly trying to catch up. "To him that hath shall be given . . ." Large individual and corporate incomes tend to promote savings and thus indirectly encourage investment and capital formation. From another viewpoint, large incomes may be said to furnish the necessary margin above survival needs which permits time and energy to be devoted to invention. The more this is done, the more rapid the rate of technological change. This in turn creates many opportunities for the profitable employment of savings by introducing the new to supplant the old. Another economic basis for our rapid growth in wealth and income has been the abundance of our natural resources at the time when our country was settled. Conditions were extremely favorable for the exercise of individual initiative in production. The free enterprise system was also favorable to the accumulation of capital and the invention of labor-saving devices. These various economic forces are and have been very interdependent.

The attitudes of individuals under capitalism might also be listed among the psychological forces, but there are others more clearly in this field. Individual capital accumulation is fostered by a desire to increase one's area of power. The owner of capital goods can give orders to those who must work with those goods in order to make a living. He can use his equipment in his struggle with other capitalists. Other individuals want to live without working and invest their savings in stocks and bonds. The sellers of these securities get control over additional funds and thus increase their power to buy and use goods and services. On the other hand, the fear of future expropriation of property or income from property may discourage saving. It may also discourage investing in the capital goods whose increase raises real wages. The fear of war might likewise cause people to put their savings into easily concealed and durable things like gold, silver, and jewels, instead of into machines and buildings. But these deterrents are often exaggerated, as will be shown in a later chapter.

The political forces influencing capital formation are both direct and indirect. Governments may make the major decisions regarding capital accumulation and force the people to cooperate. The Russian "Five Year Plans" offer a good illustration, also Hitler Germany's "Guns, not butter." In this country a local school building program is ordered by the voters of a district. Or a TVA may be proposed and financed by the representatives of those voters in Congress. Progressive income and inheritance taxes influence capital formation in a different fashion. Many other acts of governments have indirect and complex effects in this field, such as subsidies, agricultural price parities, tariffs, price ceilings, etc. Further comments on capital formation must be deferred to the chapters on interest.

6. The Effects of Inventions upon Real Wages May Be Good or Bad, Depending upon Circumstances.—As suggested above, any invention which improves the efficiency of capital goods or the way in which they are used will increase total product and therefore the per capita real income of the group. This generalized approach, however, says nothing about the way in which inventions may change the *distribution* of the enlarged total product. New methods, materials, or products are introduced by capitalist businessmen presumably for their own benefit and not with a desire to benefit labor. Just what is their effect upon the functional shares of a country's total income? This question can best be answered by considering first the still narrower problem of the individual firm.

In a capitalistic economy inventions are usually sought and introduced because of the prospect of financial reward. They are either cost-reducing inventions expected to widen profit margins or new products designed to attract a profitable volume of sales. Costs may be reduced by mere rearrangement of a given group of factors or by substituting some for others. Inventions which have the latter result are usually classified as capital-saving or labor-saving, while others are neutral, like many new products. Land-saving inventions may be grouped separately if desired, such as building techniques which permit erecting taller structures than before.

Nearly all inventions benefit laborers as a group by improving their efficiency, by increasing the quantity of other factors with which labor works, or indirectly by providing desirable consumer goods. Cost-reducing inventions may lead businessmen voluntarily to reduce the selling prices of their goods in order to get a larger total profit through expanded sales than would be possible with smaller output and larger profit margins per unit. This will also

tend to force downward the prices of substitute products. Consumers, including laborers, will benefit.[4]

But what about labor-saving inventions? Do they benefit or harm labor as a whole? It is obvious that in most cases their introduction means the displacement of certain individual laborers who are forced to seek work elsewhere at presumptively poorer terms. Where the demand for the product is very elastic and price competition is active, reduced prices may increase sales so much that the displaced workers may be reabsorbed in the same firm with little loss in time or job-rating, but this is the exception rather than the rule. Labor in general may benefit as consumers if, according to the argument of the preceding paragraph, the labor-saving invention leads to reduced prices of consumption goods or better quality. The degree to which prices are cut determines how much of the entrepreneur's initial gain is passed on to society as a whole. A precise balance can never be struck between the amount of gain received by worker-consumers and the loss suffered by the particular displaced employees. The gain and the loss occur to different people and are incommensurable, but society as a whole clearly gains.

Labor-saving devices are both substitutes for and complements of labor. In effect they constitute an increase in the supply of labor and thus reduce its marginal physical productivity. At the same time they increase the supply of capital (by improving its quality) and thus, according to the law of proportionality, raise the marginal productivity of labor. Which of these forces is the greater determines the net effect of inventions upon the average real income of laborers. The history of rising real wages in the past fifty years indicates that the complementary effect has outweighed the substitutionary effect. And there is no conclusive evidence that technological unemployment is cumulative.

7. Other Effects of Inventions.—The foregoing analysis does not treat a closely related problem, the proportionate shares of the total product going to laborers as a group, to capitalists as a group, etc. (Cf. Section 8.) The question of what governs how much individual laborers receive seems more important in terms of national policy. In fact, individual workers rarely concern themselves with anything beyond the specific and immediate effect of a certain change, such as an invention, upon themselves. They leave to economists the task or the pleasure of figuring out the larger issues. Nevertheless,

[4] If the new products are not consumer goods, and that is often the case, the reduced prices may not benefit consumers in the short run, but only the intervening producers or distributors. Again the degree of price competition is important.

they sometimes use his arguments about shares of national income when contending for higher wages or legislative support. Businessmen who are concerned about profits may invoke the economist and phrases like *economic progress* when they want to get workers to stop impeding the introduction of labor-saving devices. At other times these same businessmen may themselves oppose the introduction of machinery that would make their own obsolete.

There are two other interesting problems connected with inventions and real wages. The first deals with the way in which capital improves itself without having to subtract anything from current income in the form of savings. This is made possible by the joint effect of inventions and depreciation accounting. As machines, buildings, and other capital goods used by business firms wear out in use, accountants include as an expense a fractional part of their total cost in each accounting period. The depreciation "expense" is not paid out for the capital good because that good was paid for at the time of purchase. No income is paid out at all for this "expense," but it is accumulated as an increase in cash or in some other asset. By the end of the useful life of the capital good, enough wealth of some sort will have been accumulated to equal the cost of that good. It will have paid for itself and can be replaced without the need to call upon the owner for new investment.[5] If in the meantime better types of machines have been invented and offered for the same price, these may be purchased instead of making identical replacements. As the stock of capital goods in a region increases, the volume of depreciation allowances likewise mounts. New inventions in increasing number may be absorbed without forcing businessmen to enter the market for capital funds, or even to save out of their own net income.[6]

8. The Effects of Unions upon Real Wages.—Union leaders often claim that the improvement in labor's economic status in the

[5] Depreciation reserves are merely an accounting entry and do not represent cash or any property in particular. But they indicate the presence of liquid or tangible wealth equal to the original cost of the capital good at the end of its useful life, if calculations have been correct and business has gone as planned.

[6] The influence of the business cycle has been ignored in the foregoing analysis. It cannot be elaborated fully here, but two brief comments may be made. First, depreciation techniques reduce business demand for funds which other individuals save. This failure to borrow and spend saved funds may create a recession unless offset in some other way. Second, optimism and pessimism regarding the future trend of business are potent causes of the amount of capital goods purchased or constructed. On the upturn, when things look good, capital goods are demanded in large quantities. Labor is in demand. Wage payments rise. On the downturn, the opposite is true. Workers must not ignore this cyclical relationship when they examine the causes of changes in real wages.

last fifty years or more is to be attributed largely to the organization of workers into unions which increased their bargaining power. They reason from particular cases where they have been able to force employers to raise money wages to the general conclusion that workers on the average are benefited whenever any group among them improves its status. Or they argue that real wages have improved most rapidly during the recent decades of union expansion and therefore, *post hoc ergo propter hoc,* the former was caused by the latter. Despite some obvious loopholes in the logic employed, the conclusions deserve serious consideration. Maybe they contain truth for reasons other than those most commonly alleged.

In the first place, organized labor may improve its own position in particular cases if by superior bargaining strength it becomes able to win larger money wages than before. This could be done at the expense of either (1) the owners of the business, (2) those who buy the products of that firm, or (3) those who sell to the firm (other workers, land owners, lenders of capital, sellers of materials). Union labor's usual assumption is that the ones who lose when organized workers gain are the "employers" or "capitalists" who own the business and whose profits share is reduced. This may be so at the outset, but the owners usually succeed ultimately in shifting part or all of their immediate loss to other groups. They may raise the price of the articles sold if the wage cost increase applies to competitors as well as to themselves. When this is done, laborers as a whole will suffer, but not as much as the organized workers gain. The buying public includes non-labor consumers, and some of the employer's added burden is very likely to remain on his shoulders or to be shifted to non-labor factors. If the wage boost prompts employers to bargain more effectively with other groups of workers, the net gain for labor as a whole will be reduced.

According to principles previously explained, employers will probably try to substitute other factors for the laborers whose wage has been bargained upward. Substitution takes time, but eventually it brings a reduced schedule demand which may offset part of the wage rate increase. Its effect upon the total revenue of the union members depends upon what may be called the long-run elasticity of demand for that labor. In dealing with real wages, additional variables enter the picture and make generalization increasingly difficult. But the unionist is undoubtedly warranted in arguing that the average union member today receives a larger money wage and a larger real wage than he would get if he were one of an unorganized group. Some unions restrict membership and reap the monopoly gains of

obstructed entry. Others contribute chiefly superior bargaining strength and skill.

The net effect of unions upon the real income of workers as a whole, including both organized and unorganized labor, involves the analyst in such a long list of possible variables working in different directions at different times that it seems best to appeal to statistics instead of to logic. Figures revealing labor's share in national income indicate that the percentage of the national income going to labor in the United States did not change much from 1899 to 1918 and from 1920 to 1939. The effects of wartime inflation and deflation brought an increase from an average of about 60 per cent in the first interval to 67 per cent in the second, the transition occurring in 1919 and 1920. If labor's share has not experienced a secular increase, then what becomes of arguments about the benefits of union organization? Few deny that the members of organized groups have benefited financially, and their gains are likely to have exceeded such increases in the cost of living as may have been the indirect effect of their gains. But the statistics seem to indicate that the *relative* position of *labor as a whole* has not improved.[7]

Other statistics show that the real income of the people as a whole, including workers, of course, has increased several fold during the past seventy years. Can this *absolute* improvement in the status of labor as a whole be credited to labor organizations? Or is it entirely the result of invention and capital accumulation as described above? The record of organized labor contains chapters in which the introduction of machines has been fought bitterly, either by outright opposition or by various forms of sabotage, including slowdown and featherbedding, which have prevented the expected labor-saving economies from being realized.

On the credit side may be put instances of unions improving the efficiency of their workers in various ways. Some unions have fine records of cooperation with management in the introduction of time-saving devices and in suggesting improvements (inventions) which reduce cost. The same unions and others have instituted training programs to improve the skill of their membership. Most of the gains in this line, however, have probably been indirect, i.e., those resulting from the improvement in labor morale and effort resulting from increased wages and a greater feeling of job security.

9. Indirect Effects of Union Activity.—Worker efficiency has been increased in indirect ways by union activity in various fields.

[7] Simon Kuznets, *National Income and Its Composition, 1919–1938,* New York: National Bureau of Economic Research, p. 22.

Collective bargaining contracts have brought improved working conditions, some participation in management, and a general sense of importance from being a member of a group possessing obvious power. Unions have led the way in movements for increased free public school opportunities, for the reduction of child labor, and for the introduction of workmen's accident compensation. Social security laws have worked in both directions. By improving health and the feeling of economic security they have helped efficiency. On the other hand, unemployment insurance may have decreased the morale of a minority. Some employers have been resentful against social security taxation and have tried to take it out of the workers indirectly.

One long-run offset may also be mentioned, the progressive reduction in the length of the work week. When the work week was very long, reductions actually raised the total weekly output in most cases. But now that a forty-hour week is fairly common, further reductions will probably reduce the total product per worker. At the same time they increase the workers' leisure. Even with reduced consumption of material goods, total satisfactions may rise. Under such circumstances one can hardly say that the worker has reduced his level of living by bargaining for a shorter work week.

Indirect gains for society may result from the way in which labor successes in bargaining force employers to search more aggressively for labor-saving devices. The increased power of unions in recent years undoubtedly has stimulated technological change in industry. It has given employers the same sort of stimulus that is traditionally provided by price competition. In other cases where business firms have used monopolistic techniques to appropriate for the small group of stockholders the gains from autonomous inventions, union demands for higher wages have been a good way to transfer these gains to a larger group of consumers.

10. An Increase in Labor Supply Usually Reduces Real Wages. —The preceding six sections have dealt chiefly with factors causing increases in real wages. Now we must consider forces which cause decreases. The most obvious one is the destruction of capital goods, which is the reverse of the process of accumulation. Earthquakes, fires, insects, and other natural calamities take continuous toll. Wars cause more severe, but fortunately sporadic, destruction. Less obvious, but important, is the increase in the labor supply of a given region which results from (1) an excess of births over deaths, (2) net immigration, (3) people seeking employment who did not

work before (women, children, aged), and (4) laborers becoming willing to work longer hours.

The last two of these variables have an effect upon the level of consumption which is different from the first two. When a person works longer hours than customarily, he is likely to receive additional pay. This increases his personal income and hence presumably his level of consumption. The decline in marginal productivity per time unit of labor which generally occurs from this type of addition to the labor supply constitutes an *increase* in marginal productivity *per laborer*. Similarly when the members of a worker's family who were not previously employed add their efforts to the total number of labor hours being worked, marginal productivity tends to fall, both per hour and per laborer. But it rises *per family unit,* and hence the level of consumption increases. Another interesting paradox in this connection is the fact that these aspects of labor supply usually diminish as time goes by and their downward secular trend is interrupted only by war, depression, or other crisis. In such times they expand as offsets to the decline in civilian consumption that would otherwise occur. The plane of living may fall in wars and depressions, but it does not fall as far as it would if these additions to the labor supply did not occur.

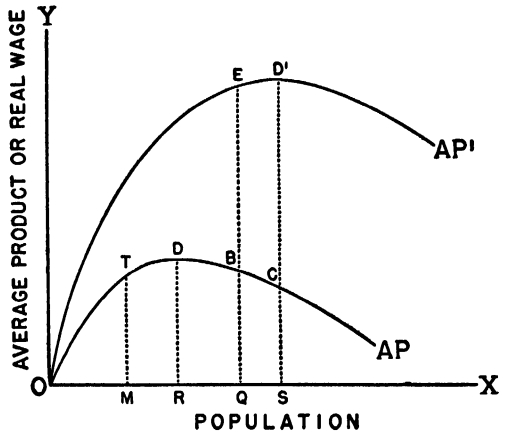
The first two types of addition to the labor supply may be grouped under the general heading of a net increase in population or of the working force of a region. Sometimes immigration is the more important of the two, as in a large part of American history. At other times natural increase is the significant factor. Whether a net increase in population will raise or lower the marginal product of a region depends in the first instance upon the capital structure of the area. If it is primitive as in India, China, and much of Central Europe, then the principle of diminishing marginal returns will operate. If, however, the region is well equipped with capital, labor *may* be applied under conditions of increasing returns. The same may be true of the settlement of virgin territory where land is the abundant factor. It is also conceivable that a region with a relatively stationary population might experience such an increase in the non-labor factors that it shifted from a condition of diminishing returns to one of increasing returns, as illustrated in Figure 64. (Compare also Figure 27, page 103.) If the working population is at first OQ with an average product curve AP, and then the latter curve rises to AP', the point of diminishing average returns may shift from OR to OS, so that there could still be an expansion of population beyond OQ by immigration or natural increase with benefit rather than harm

to the average real income of the people. A situation of this type might arise in any region where a large new factory is built or an irrigation project completed.

The form of capital and the structure of capitalistic organization tend to become adjusted to the size of the working force available, so that a war which is very destructive of human life might also push

FIGURE 64

EFFECTS OF POPULATION CHANGES UPON THE AVERAGE LEVEL OF CONSUMPTION



a country back from a decreasing returns position like QB to one of increasing returns like MT, although the probability is slight (see Section 9 below). Even in the primitive society of the Middle Ages, the Black Death, which took one third of the population of western Europe, may have had this sort of effect. Ordinarily, however, a regional economy operates so far past the point of diminishing returns for labor that a plague, famine, extensive emigration, or rapid industrialization would merely relieve population pressure momentarily and not nearly enough to get back past the point of optimum population. Most people who favor quantitative restrictions upon immigration into their country are probably arguing from a sound economic foundation, once the materialistic objective of maintaining or increasing consumption levels is accepted as the major goal. Exceptions would be found when the immigrants were more advanced industrially than the natives, as in the Americas in the seventeenth century and Palestine in the twentieth.

11. The Dynamics of Real Wages—Statistical Investigations. —Attempts have been made to investigate the statistics which are

available to determine the relative influence of changes in the supplies of both capital and labor in western countries during recent decades. The most comprehensive study made in this country has been that of Paul H. Douglas, who came to the conclusion that the marginal productivity of labor in manufacturing from 1890 to 1922 declined 0.25 per cent for every 1.0 per cent increase in the labor supply. Similarly the marginal productivity of capital declined 0.75 per cent for every 1.0 increase in fixed capital.[8]

More significant, perhaps, is the cross-elasticity between the increase in capital and the increase in the marginal productivity of labor. This is determined as 3/4 of the total increase in product attributable to capital increase, or approximately 1/5 (18.9%). In other words a 1.00 per cent increase in fixed capital accumulated (in manufacturing) will *raise* the marginal productivity of *labor* approximately 0.20 per cent. On the other hand, a 1.00 per cent increase in labor will *reduce its own* marginal productivity about 0.25 per cent. Therefore, if inventions are ignored and the only two variables are capital accumulation and population expansion, the rate of capital accumulation must be more rapid than the rate of population increase if real wages are to rise, or even to remain constant.

According to the record, this condition was amply satisfied from 1899 to 1922. Capital in manufacturing increased much more rapidly than labor (1922 indexes: 431 and 161 respectively). As a result the *marginal* productivity of labor rose 49 per cent during the same interval. This figure exceeds considerably the 21 per cent increase in the real wage index for manufacturing obtained by dividing a money wage index by a cost of living index, but both results affirm the fact of an increase.[9] Perhaps part of the divergence may be explained by such factors as a time lag between increases in marginal productivity and the payment of higher money wages (the real wage index for 1923 was 128.6); the inclusion of nonmanufactured goods and services in the cost of living index; the use of retail prices in this index; an improved bargaining position upon the part of manufacturers, etc. An average value product index per employee during the same interval showed an increase of 36 per cent.

12. Differences Between per Capita Real Incomes in Different Countries at the Same Time.—The foregoing discussion of the dy-

[8] For the economy as a whole the figures were probably closer to 0.33 per cent and 0.67 per cent respectively. See Paul H. Douglas, *The Theory of Wages*, New York: The Macmillan Co., 1934, p. 493, and by implication, pp. 490–491. Also see Ch. 5.

[9] *Ibid.*, pp. 121, 125, 146, 512, and Chs. 5, 8, 20 *passim.*

namics of real wages provides tools of analysis which can expedite the answer to our next question about the reasons for differences in levels of consumption between countries at any given time. The basic reason is differences in the average productivity of the people. This is (1) inversely related to the size of the population and directly related to (2) the total hours which those people work, (3) the skill and morale of the workers, and (4) the quantity and quality of other factors which cooperate with labor. Among these other factors the more important ones are (a) the quantity of machinery and capital equipment of all kinds, (b) the amount of developed natural resources, especially fuel, power, metals, and arable land with good growing climate, and (c) the degree of managerial aggressiveness in applying new and more efficient methods of production. There are also various intangibles such as (5) the presence of an efficient system of finance and distribution and (6) a large accessible trading area permitting extensive geographical specialization and the economies of large-scale production.

Applying these principles, we may say that the high plane of living characteristic of the United States is largely the result of our having relatively large amounts of these "other factors" per worker and the fact that both management and workers have a high general average of skill and morale. In New Zealand, which had an equally high per capita real income before the war, the nonlabor resources are not so numerous, but the population is small per unit of agricultural land, workers are capable, and management is energetic. A special advantage of that country may perhaps lie in the morale-boosting effect of extensive labor organization and bargaining together with an ample social security program.

At the other extreme are the "backward countries" which lack capital accumulation, managerial enterprise, skilled workmen, and developed fuel or power resources. Technological knowledge is so cosmopolitan and so easily borrowed that the failure of a backward country to make use of it must be blamed not upon its absence, but upon the deficiencies just named. The accident of a late start down the road of industrialization is the main reason why such countries as Brazil and Russia lag behind.[10] The handicap of others such as India and China lies partly in the historical accident that brought them the benefits of modern medicine before the benefits of modern agricultural and industrial technology. They learned how to reduce

[10] Another effect of a late start is that the possibility of borrowing ideas and importing equipment and engineers permits the *rate* of industrial growth to be greater at the present in the adolescent countries than in the mature ones.

the death rate (especially infant mortality) before they accumulated equipment for increasing per capita productivity. In Western Europe and the United States the industrial revolution preceded the medical revolution and the plane of living rose before population growth could hold it back. Other underpopulated countries in the western hemisphere and the British Dominions have the blessing of abundant land to thank for their early start in the struggle for a higher plane of living. A detailed analysis of economic geography and economic history would be necessary to find the particular causes applicable in each case and to assign them their proper weights.

Real wages may also decline relative to other countries if a given nation decides to devote more of its resources to capital formation and accumulation than others do. This happens when, under the stimulus of a "Five-Year Plan" or other visions of greater power or productivity in the future, a nation's leaders decide to tax the people heavily to get more funds than would be made available by voluntary saving and lending. A less obviously painful method than taxation is the forced saving that occurs when deficit spending by the government raises prices more rapidly than the rise in many individual incomes.

13. Intercountry Differences in Living Standards Affected by War.—If a nation chooses to divert a sizeable proportion of its resources of men, materials, and machines into the accumulation of instruments of destruction, the standard of living may decline even though per capita productivity remains high. This qualification must be added to the generalization thus far made about the influence of factor quantities upon real wages. On the other hand, if rearmament begins during a depression, it may even raise the real wages of the working class as in Nazi Germany from 1932 to 1939.

Of course if war actually occurs, it affects the standard of living in various ways. First, it will probably increase the volume of production set aside for military purposes, thus tending to leave less for civilian consumption. Or it may stimulate people's willingness to work so that their services are available in far larger quantity than before the war. This is particularly true of management, which occupies a crucial position through its ability to discover and apply new methods of production which constitute in effect an increase in the intangible factor, knowledge. Women and retired people join the working forces, and regular employees work overtime. This latter change, however, may be more than offset by the draining of man-

power into the armed forces. No generalization can do more than mention the possible variables and point out the direction of their impact. Each situation will be different in over-all effect, depending upon circumstances, but most common is a decrease in civilian consumption, since both the relative and the absolute amounts taken for war usually increase.

Second, if a war results also in physical destruction at home as by gunfire, bombs, or incendiaries, there will be a decrease in the stock of capital goods such as factories, transportation facilities, office buildings, and homes. This will clearly diminish the average and marginal productivities of labor, so that a decrease in the level of consumption is inevitable. The disorganization of customary markets and sources of materials will also reduce output. The loss of life which accompanies such destruction would tend to raise the marginal product again were it not for the extra wounded to care for, the shattering of morale, and the drain on physical energy caused by insufficient shelter, clothing and food. However, even though the marginal product were to rise to its former level, the total product would be diminished. If war demands remain constant, the remainder for civilian use would have to decline.

A third way in which war affects the average level of consumption is through its effect upon international economic relations. War usually cuts off more trade connection than it establishes and therefore diminishes the quantity of factors which cooperate with the people of the warring country through division of labor and exchange. Isolation forces that country to produce substitute products at greater cost than for the exports which secured the former imports. A partial offset may be found through war-created opportunities to get imports without exporting. Foreign allies may lend or give (cf. lend-lease) more freely than in peacetime, so that domestic consumption may be maintained even though the production of civilian goods per capita declines. On the export side there may also be some saving through default on debt charges payable to enemy countries, or even to allies. Other invisibles that may affect the picture one way or another include especially tourist travel, shipping services, and the international movement of bank balances or short-term capital. In connection with trade a final comment should be made about the effect of possible changes in foreign exchange rates. They usually move adversely to the warring country and therefore tend to reduce its volume of imports of civilian goods even more than would have occurred under war demands, shipping shortages, and blockades.

14. "Economic Parasites" Reduce the Level of Living of Those Who Work.—Countries differ in the percentage of their population which does not work. The idle have to be supported by the efforts of those who labor. *Except as they contribute indirectly* to the satisfactions of the working population, the idle may be called "economic parasites." This is clearly true of the unemployables, the sick, the aged, the crippled, and the mentally incompetent. Young children are parasites in one sense, but they give satisfaction to most parents whose labor supports them. From society's viewpoint children are people being prepared for the work of the future. If society decides that its youth should be trained better than formerly, the age for compulsory school attendance may be raised or the better students may be subsidized for higher education. Americans believe that the welfare of their country is improved by a long training period for its youth and are willing to have children remain economic parasites until they are sixteen, eighteen, or older. The high per capita income of the country helps to make this possible. If the output per family were much lower, as in many foreign countries, children would have to go to work at an early age in order to survive.

Our great wealth also makes possible an early retirement age. Many believe that it is good for the economy to force people to stop working at 65, 60, or earlier. The usual argument is compounded of a belief that there are not enough jobs for everybody and a sense of justice which says that the older people have worked enough and ought to be supported for their remaining years by the younger people whom they worked to rear.

In poor countries a higher percentage of the women have jobs outside the home than in rich countries. Here again there is a mixture of motives. The poor feel it necessary. Those with higher incomes believe that women contribute most to society by staying at home and caring for the men and the children. In times of depression, working women are criticized for "taking away the jobs that rightfully belong to men." Most men don't want their wives to work and are willing to support them in exchange for their services in the home. Housewives are not "economic parasites" in the usual sense of the term. They contribute services and satisfactions directly instead of working outside of the home to earn money with which to buy goods and services for the family.

The unemployed are "parasites" through no fault of their own. They are able to work and want to work, but cannot find employment. Some of them live on their savings, others on their relatives. Some get unemployment insurance, part of which they themselves

may have contributed while working. Others get a dole. Regardless of where they get the funds which they spend for food, clothing, and other items, the unemployed do not currently contribute to the production of the goods which they consume. The larger the fraction of the total population which they represent and the more they consume, the less the product that remains for those who do work.

There is also the problem of the armed forces. These men are nonproductive in the usual sense. They are "economic parasites." Yet a majority of the citizens whose level of living is reduced to support the military derive satisfaction from the thought that their country is "well defended" or is able to force its will upon weaker states. Others condemn the diversion of an increasing fraction of the national income into armed forces which are "too large" or more expensive than some other method of "defense." A few citizens get only apprehension, not satisfaction from their nation's military parasites. They fear that the armed forces may grow so large as to dominate the government and force civilians to vote ever larger appropriations. They also fear that the military influence in government may provoke war, not prevent it. The economic destructiveness of atomic war makes such apprehensions quite understandable. It is difficult for an economist to resolve such arguments. But the greater the number of military parasites in proportion to the number of workers, the lower the average level of living in terms of the goods and services which make up conventional budgets. Here again the great wealth of the United States and its small armies in the past have kept us from feeling the military burden very much. Poorer countries have suffered severely when a quarter or more of the national income has been diverted to preparedness expenditures.

15. Large "Unearned Incomes," a Problem of the Unequal Distribution of Wealth.—Capitalistic societies differ in the distribution of wealth among their populations. This is one cause of differences in their average levels of living. Some economies have a fairly large "middle class" like the United States. Others have chiefly the extremes of rich and poor with few between. Most of the very rich derive enough income from their wealth so that they do not have to work. They are "economic parasites." Many rich men also work. They combine earned and unearned incomes. In the middle classes, wages and salaries are supplemented to a much smaller degree by income from investments. There are others, such as widows, orphans, and retired people, whose very modest income is derived completely from property. We never think of calling these poor

people "the idle rich," but only those who are fully capable of work and play around instead because of some fortunate inheritance.

All property income is "unearned" to the extent that its receipt requires no current work on the part of the recipient. Whether its beneficiaries are in high or low income brackets, they are "economic parasites" when they use unearned income to obtain a share of the products of wage and salaried workers. As with similar groups examined above, some social justification is possible on various grounds. Some property owners have been thrifty savers in the past, some are philanthropic in the giving of time and money, and some cultivate the arts or promote research for the benefit of society as a whole. There is also the argument that large unearned incomes in the hands of a few are an inevitable by-product of the institution of private property which is believed to be essential to a free and progressive society. The beneficiaries of the system cultivate an acceptance of such ideas by the general public. In this they resemble other "economic parasites" such as the military who have been quite successful in getting taxpayers to believe that increasing armaments spell increasing national and personal safety. Since there seems to be an element of truth in such arguments, it is best for the workers to be happy in their support of the drones. This gives them satisfactions which lighten their load and make it tolerable, provided the "economic parasites" do not become too numerous or demanding.

The foregoing discussion is not intended to imply that the owners of private wealth should be expropriated or that state aid should be denied to youth, the aged, and the unemployed. Nor does it suggest that maximizing the average level of consumption for the masses is the proper major objective of social policy. It is merely intended to explain further variables influencing the level of consumption of those whose work is responsible for the welfare of the whole group. In some regions the bargaining power of the owning classes is probably stronger and that of the working classes weaker than in other regions. This causes differences in the distribution of income which must be taken into consideration if one is comparing and explaining the relative levels of consumption of farm labor or factory workers in different regions. If the workers at the bottom are kept so poor that they have insufficient nourishment, clothing, shelter, and education, their physical efficiency and morale may also be impaired. As a result an area where title to wealth is very unevenly distributed may have an average productivity lower than one otherwise identical, but with a much more even distribution of wealth. Considerations of this type probably help to explain why Poland or

Spain, for instance, were less prosperous before 1939 than Denmark or Norway.

16. Roundabout Effects of Differences in Income Distribution.—A final comment should be made about the long-run effects of various possible distributions of the national income. A complete discussion of this topic belongs properly in a chapter on economic progress, but briefly the argument may be given that in a capitalistic economy the volume of saving is probably enhanced by considerable inequality. If all of the unearned income in this country were distributed equally among the productive workers, the latter would probably save less in the aggregate than they together with the rich save now. This is so because those with low incomes generally have a much greater preference for additional present goods and services than for future security through saving and investing. Those with large incomes can buy the present goods they want with only part of their incomes and can easily create a surplus with which to satisfy their demand for expected future income or power.

The effects of the rate of saving upon real wages are roundabout and dependent upon other variables. For instance, savings which are invested in capital goods may promote economic progress. Savings which are hoarded promote unemployment of men and machines. The former result raises the level of living, the latter result lowers it. Savings by individuals are not as important as formerly in this country. Business firms accumulate most of their own savings for capital replacement and for expansion. They do not have to rely upon selling stocks or bonds to individuals as much as was necessary in an era when most profits were distributed as dividends. Bank credit and government credit are much more readily available for capital expansion than they used to be. Therefore, the deflationary effect of saving is probably more to be feared than the handicap-to-progress effect of inadequate saving. Further comments on this subject will be found in the chapters on interest.

The implications for social policy from this very brief analysis are somewhat as follows. Governments should do everything possible to promote full employment of resources in order that real wages may be as high as possible. This includes acts to reduce saving, to stimulate the investment of savings in capital goods, and to offset hoarded funds by government spending. No action should be taken in one field without considering its probable adverse results in another. If high taxes on large incomes discourage the investment of savings in new ventures, the loss to society may exceed, or be less

than, the gain. If removing all taxes or profit income stimulates investment, it will also distort the distribution of incomes in favor of the rich and promote greater saving. The net balance may be good for the average citizen, or it may be bad. Unstated results also enter the picture. And there are other objectives which guide men's actions. Maximizing per capita income as a goal in our society must be supplemented with the maintenance of freedom, the securing of economic justice, and the preservation of peace. Those who contemplate legislation to bring the worker a larger share of the national income should put a whole group of the pros and cons of many alternatives on the scales, not just one or two, before deciding which course is the best.

17. Summary of Economic Principles Explaining Differences in Real Wages.—Real wages differ from place to place and time to time according to differences in money wages and the cost of buying a representative budget. Comparisons can best be made in terms of the marginal product on the assumption that the workers of even large groups are roughly homogeneous, or have similar differences. Using this approach the comparative real wages of any labor group are seen to be a time or place function of differences in:

1. Quantity of labor supplied by the labor group (includes number of laborers, hours worked, efficiency of work)
2. Quantity of other factors, especially capital, which usually increases through time but is destroyed by wars
3. Efficiency with which the factors are combined; especially the role of inventions of various kinds

Additional determinants which do not fit into the above approach include:

4. Number of workers in the representative family group
5. The strength of unions in:
   (a) Restricting entry (keeping down the size of their group)
   (b) Winning larger money wages at expense of employers, consumers, or other labor groups
   (c) Getting paid for work not performed (featherbedding)
   (d) Improving the cooperation or energy output of members
6. Amount of national product diverted into:
   (a) Capital formation by private initiative or government plan
   (b) Preparation for war

(c) Support of non-working groups ("economic parasites")
  (1) Unemployables
  (2) Youth, aged
  (3) Women
  (4) Unemployed
  (5) Armed forces
  (6) Idle rich

7. Miscellaneous determinants of the size of the national product
  (a) War destruction or war stimulation
  (b) Size of the free trade area
  (c) Distribution of wealth and income
  (d) Uninvested savings

Chapter 20

WAGE THEORY AND UNEMPLOYMENT

1. The Wage Theory Approach to the Problem of Unemployment.—Unemployment exists when people who want to work cannot find work at terms satisfactory to them. The unemployed include people who have never worked before, those who have quit former jobs, and those who have been discharged or laid off. A complete study of the problem would explain the reasons for quits, discharges, layoffs, and the failure of people to find the jobs they want. It would also consider the effects of unemployment on individuals, on the economy, and on social and political developments.

In this chapter we shall discuss only those aspects of unemployment which are appropriate to the wage theory section of a book on price economics. The *effects* of unemployment upon wage rates have already been discussed in earlier chapters. Only a portion of that material need be repeated here. We now want to study how wage rates, wage offers, or wage demands *cause* unemployment. The specific questions to be answered are the following:

1. How do wage rates or their changes affect unemployment?
   - (a) When product demand declines?
   - (b) When other factors become relatively cheaper than labor?
     - (1) Because workers demand increased pay?
     - (2) Because of decreases in the prices of other factors?
     - (3) Because of changed technology?
     - (4) Because of reduced worker efficiency?
   - (c) When employers do not pay enough to satisfy workers?

2. Unemployment as an Individual Problem.—Statistics of aggregate unemployment represent the total of individual unemployment situations. If a million men are listed as looking for jobs, each one of them has his own problem, but certain typical situations may be generalized. At the risk of oversimplification, we may think of unemployment as resulting when two people cannot agree about a job. The worker asks too much or the employer offers too little. They do not agree, and they do not make a deal. The disagreement may relate to wage rates or to other terms of employment. If an

employer wants a skilled worker and the applicant is without experience or skill, willingness of the unemployed man to work for little or nothing will not get him the job. Similarly, a union card or a college degree may be required before negotiations begin. The trouble may be merely that one or both lack information about the opportunities. The worker does not know that there is an employer offering work at acceptable terms. Or the employer does not know about Bill Smith who is willing to work at the job offered.

The individual problem may even lie in the realm of personalities. The boss does not like Jim Brown and fires him. Or employer Cross has a bad reputation. He is known as a slave driver, a man reluctant to make promotions, an enemy of unions, "unfair to organized labor," or the like. Workers may quit because an employer does things they don't like or fails to do things they want. Employers may dislike Negroes, Jews, women workers, dark-skinned people, cross-eyed individuals, etc. It is not our task to inquire into the reasons for these antipathies on one side or the other. Nor are we concerned with their merits. They are mentioned here only because if they were omitted, the hasty reader might think that the following discussion was intended to cover the entire subject instead of just one part of it.

3. When Product Demand Declines, Labor Demand Declines Also.—If the demand for a manufactured good declines, output is likely to decline, too. When the production rate is reduced, fewer workers are needed. Employees are laid off, and unemployment figures rise. The sequence is obvious, but the implications should be examined with care. Each statement must be accepted as a generalization to which there are some exceptions, but not enough to invalidate the rule.

Take the case of automobiles. A decline in schedule demand means that fewer automobiles are bought per week than before. Fewer Fords will be produced, unless there is a temporary willingness to produce for inventory. Fewer employees will be needed in Ford plants. The wage rate has no direct connection with this type of unemployment. But there are some indirect connections of importance. If Ford were to reduce his price while other automobile producers kept theirs unchanged, Ford sales would probably rise, perhaps back to their former level. This would benefit Ford workers, but it might not benefit Ford. The price cut might reduce his total net profit below that which existed with the reduced volume of sales and the unchanged price. However, if Ford workers wanted so much to keep working that they offered to work for less, Ford might ex-

pect to save as much on wages as he lost on the price cut. This might be sufficient inducement for him to run the risk of price-cutting in a highly competitive industry. Such is one possible indirect connection between wage rates and employment.[1]

The foregoing illustration leads to the generalization that rigid wage rates are *one* of the things which promote unemployment in times of declining demand. But it does *not* show that in most cases of declining demand the number of workers could be held constant by wage cuts. The argument is rather that *some* unemployment might be averted in this way.

How much good wage cuts would do depends upon several other things. If the employer will not cut his price, there will be no change. If he does make a price cut, the resultant increase in sales and employment will depend upon the amount of the cut and the elasticity of demand for his product. Demand-elasticity in turn depends upon what prices are asked by competitors, the success of their advertising, quality changes in rival products, and many other possible variables. All of these things have been discussed in earlier chapters (cf. Chapter 4).

4. Wage Cuts Will Not Cure a Depression, May Make It Worse.—During a recession we usually hunt around for something to do which will stop the decline. Workers say prices should be reduced so that people can buy more and employers hire more. Employers are more apt to say that wages should be reduced so that prices can be reduced, more goods sold, and more workers hired. Most producers during recessions are unwilling to risk price cuts to test the elasticity of demand for their product.[2] It is doubtful if wage reductions would increase very much their willingness to cut prices. Fear of equal or greater price cuts by rivals remains a strong deterrent. Therefore, on this ground alone, economists would be warranted in opposing wage cuts as a general policy, although admitting the presence of exceptional cases.

There is also a roundabout effect of wage cuts to be considered. Wage-rate reductions may decrease the total of wage payments or may increase them, depending upon the elasticity of demand for labor, which in turn depends upon the elasticity of demand for the product

---

[1] Note that the assumption is one of a general decline in demand, not one of Ford falling behind in sales because competitors put out better cars.

[2] The producers referred to are, of course, those who can control their selling prices and increase output if market demand increases with a drop in price. The argument excludes farmers and other sellers in markets where there are very many rivals on the supply side.

whose price is now presumed to be cut.[8] If we start with 1,000 workers earning $50 per week, the total payroll is $50,000. A decline in demand and sales brings a 20 per cent layoff. The total payroll drops to 800 times $50, or $40,000. The question now is whether that $40,000 will be raised or lowered by a wage cut to $40 per week. If there is no price cut (and if sales do not expand for any other reason), the number of employees will remain at 800 and the total payroll at the new rate will be $32,000. This reduction of $8,000 in earnings will reduce spending by approximately that amount. Sales will decline and the demand for workers will be less than before. Laborers in general will have lost, not gained, by the wage cut.

Using the less likely assumption that price cuts do follow the wage cut, we note that sales probably would have to increase more than 25 per cent to raise the number of workers from 800 to a figure above 1,000. A labor force of 1,100 would mean a total payroll of $44,000, up $4,000 from the previous total. It is this $4,000 increase which is important, *if it occurs,* not the fact that $44,000 is still $6,000 below the original total payroll of $50,000. But sales may not increase that much. If they expand employment only up to 900 persons, the payroll of $36,000 would permit fewer purchases than the $40,000 received at the old wage of $50 for 800 employees.

The next question to be considered is the effect of wage cuts upon the volume of spending out of a given total wage income. Consider, for instance, the two alternatives of $40,000 received by 800 workers at $50 each and $40,000 received by 1,000 workers at $40 each after wage rates have been cut. It is probable that the latter group would spend more than the former. The savings of the average family after a $10 weekly wage cut are likely to be small if not negative. Approximately all the $40,000 would be spent. But the 800 fortunate workers still getting $50 per week would continue to save, perhaps more than before, so that their aggregate spending might be only $36,000. This would seem to imply that recovery would be promoted more by a wage cut than by wage rigidity. The argument, however, neglects the probability that the wage cut would *not* restore employment to 1,000 workers. The more probable result during a business recession would be only a very slight proximate increase, such as from 800 to 825 workers. This would leave total payrolls at $33,000, much below the $36,000 spent by the 800 workers receiving $50 per

[8] This is a reasonable but not a necessary assumption. There would be some cases where employers would be willing to hire more workers or to reduce their layoffs if wage rates were cut, even though sales volume were not expanded by price cuts or otherwise.

week and saving $5. The ultimate roundabout effect would be still worse.

There is still another possibility which might be significant in the short run when unemployment is not large. It is based upon the observation that when workers lose their jobs, they do not curtail their spending very much very soon. Ultimately they may have to reduce their level of living considerably, but at first they resist this change. They draw upon savings, borrow from friends and relatives, get loans on their life insurance, automobiles, or furniture, etc. Within a few weeks they may begin to draw unemployment insurance. After that runs out, they may get a relief dole, or charity aid. The 200 workers laid off in our illustration may spend as much as $40 per week during the first few weeks of unemployment. Total spending does not drop from $50,000 to $40,000, but only to $48,000. After several months the drop might be to $30 per week for the unemployed, or a total of $46,000. This is clearly larger than the amount that would be spent if the entire working force of 1,000 were employed at the lower weekly wage of $40. If total consumer spending is an important force in determining the level of business activity, it would seem better, at least at first, to keep wage rates rigid.

From several lines of reasoning we reach the conclusion that *as a general rule* wage cuts are more likely to accelerate a decline in total spending than to increase it. They should not be recommended as an appropriate cure for a business recession or depression. Wage cuts are not likely to increase very much the proximate demand for labor. Their roundabout effects on aggregate consumer spending would probably be deflationary. Investment to increase capacity or to modernize will not be stimulated during a recession by anything so simple as a few reductions in wage rates. However, certain exceptional cases can be found. Occasionally wage cuts keep a firm from shutting down entirely. There are also certain social advantages of private employment induced by wage cuts instead of relief or government work projects. But reducing the length of the workweek is a much quicker and surer way of securing this objective than the highly uncertain method of reducing wage rates.

5. Wage Rates Promote Unemployment When Other Factors Become Relatively Cheaper Than Labor.—Firms also dismiss workers when it becomes profitable to substitute other factors for labor. Changes of four types create that kind of situation. First, workers may raise their wage-rate demands. Second, interest rates or rents may fall. Third, the daily output of the average worker may

decline. Fourth, the efficiency of other factors may rise, as when "labor-saving" devices are invented.

A fifth possibility may be mentioned briefly before passing to a more extensive study of the other four. It is a shift in consumer demand from one type of product to another. This will almost always bring some change in the kind of labor demanded, as when automobiles were purchased instead of carriages. The shift may also change the total quantity of labor demanded if the second article requires less labor per dollar value of finished product than the former. Beer versus milk is perhaps a good illustration.

6. Unemployment May Be Caused by Wage Demands Being Too High.—When workers raise their wage demands, employers may be unwilling or unable to pay the higher wage. As a result the workers and employers cannot agree on a mutually acceptable wage rate. Workers leave their jobs or refuse jobs offered to them. They may be counted among the unemployed. Eventually they may get jobs if employers agree to meet the wage demands or if the workers reduce those demands.

This analysis implies no blame upon either party. Each may have what it considers good and sufficient reasons for its refusal to accept the terms of the other. The workers may feel that their pay has not risen as fast as that of comparable workers. Perhaps wage rates have not kept pace with the rise in the cost of living. Or the employer is making huge profits and should share with his workers. On the other side the employer may feel he cannot pay higher wages without losing his profits or even going into the red. Perhaps he is now making large profits for the first time in many years. Or he may not see how he can shift part of the labor cost to his consumers through higher prices.

To pass judgment regarding partisan arguments of this type requires a criterion of economic justice, a philosophy of social welfare, or some ethical standard. If the economist wants to suggest what should be done in a given situation, he should first be clear in his own mind about his scale of values. He should recognize whence it came and what other people share it with him. When he speaks or writes his views, he should declare his objectives. This done, he can be as strong in his praise or condemnation as the facts warrant. Economic science requires only that economists be scientific in their approach to economic problems. It does not demand that they remain isolated in an ivory tower far from the issues of the day.

7. Unemployment May Be Caused by Decreases in the Relative Prices of the Other Factors of Production.—Another cause of unemployment may be found in the changing relative prices of the factors of production. Insofar as they are substitutes for one another in production, a fall in the price of one will tend to cause a decrease in the demand for others, and vice versa. Thus a fall in long-term interest rates may lead to some *substitution* of capital equipment for labor of certain types. The same tendency will prevail if there is a reduction in the market price of capital equipment, or a fall in rents and land values. Another illustration may be found when an employer fires one employee to hire another who has similar ability, but who is willing to work for less pay. In the days before workers became extensively organized, this sort of displacement was a common occurrence. One of the objections to immigration from Europe was the fact that the immigrants were willing to work at lower wages than those who had been here a generation or longer. The same is true today, but on a lesser scale, when migration occurs from farm to city or from the South to the North or West. The unemployment which results from causes of this type is probably not large in the aggregate at any one time, but it is a continuing irritant.

8. A Decrease in Labor Efficiency Reduces the Demand for Labor Like an Increase in Pay.—If the daily output per worker declines, the employer is hurt as badly as if he had to pay that much more in wages. He will strive to substitute other factors for the labor which has become more costly per unit of product. The long-run effect will be much the same. In the short run there may be some differences. There may be a gradual decline in efficiency over a period of good business which is not crucial for the employer until prices drop or sales fall off. It is then hard for management to make low productivity an issue. On the other hand, individual laborers may be fired if their output performance is not up to a set standard or not equal to the average for the group. And if the issue arises in a controversy over a labor contract with a union, the outcome may be voluntary quits or forced layoffs just as if the dispute were over wage demands. "Featherbedding" rules are well-known points of disagreement between employers and employees.

9. Unemployment May Be Caused by Changes in Technology.—Inventions may cause unemployment by reducing the amount of labor needed to produce a certain good. This occurs when a new process, machine, or product improves the physical efficiency of the

labor employed with it so that an employer can produce more goods with the same amount of labor or the same quantity of goods with less labor. When he can sell profitably all that he can produce, the employer may not diminish his demand for labor. But when the demand curve for his product has appreciable inelasticity, the employer who introduces a labor-saving invention is very likely to discharge some of his workers. He can maximize his profit at less than capacity output, which usually means that he employs fewer workers than he hired before the change. Another possibility is that he will discharge skilled workers and hire unskilled workers instead, perhaps an equal number but less costly.

The loss of employment suffered by the technological victims of particular innovations is individually serious, but its effect on the economy as a whole is usually beneficial, not harmful. Offsetting increases in demand may occur in the market sense if the enterprising firm is led by competition to reduce its prices. If these products have a fairly elastic demand, they will be purchased in appreciably larger quantities and additional labor will be needed for the extra production (or not as many men need be laid off at the time of technological change). The decreased price of the first product may enable some people to purchase a larger quantity or variety of other goods. If the technological change requires machinery, people will be needed in its production. If larger profits are made by the innovating firm, its owners will have larger incomes which they may spend for consumption or may invest in creating new production facilities. To the extent that demand does increase in one or more of these ways, employment will rise. In the short run, occupational and geographical immobility may keep the displaced persons from shifting from areas of contracting employment to firms or regions where employment is expanding. In the long run labor-saving inventions do not increase the total volume of unemployment. At least they have not done so in the last 100 years.

The foregoing argument should not obscure certain important areas of unemployment which are so large as not to be dismissed lightly by reference to "the group as a whole." Technological change has been particularly rapid in agriculture, and displaced persons have found it difficult to move to cities or to find work once they got there. There are also numerous instances of areas where the major industry was forced to close down because of technological improvements which aided competitors elsewhere. The substitution of other fuels for coal, for instance, displaced many coal miners when their mines lost market outlets. Stranded skills are notorious. The general

trend in mechanical improvements is to reduce the time required to train a worker for his task. Many with long apprenticeship now find their years of training no longer of any importance in the labor market. It is very difficult for such persons to make the mental adjustment necessary to shift downward in the economic and social scale. Therefore, they tend to remain unemployed longer than those who lose their jobs for other reasons than technological change. Finally, the rate of technological change at times may exceed the rate of reabsorption of displaced labor, as in depressions.

Sometimes changes in the demand for labor are the cause and not the effect of the introduction of labor-saving devices. In the first paragraph of this section the implication was that employers sought and introduced inventions because of their desire to increase profit. They may also be motivated to seek cost-reducing innovations by a desire to avoid loss. This latter situation may occur when workers take advantage of their organized strength or the scarcity of labor to force their employer to raise wages. As indicated above in Section 16 of Chapter 17, this may reduce profits so much as to prod the employer into more active cost competition than before. He will hunt energetically for ways to reduce his wage bill without diminishing his output. If these can be found and introduced, the workers of the firm may experience a setback which they did not contemplate at the time they demanded higher pay. However, for the working force of the economy as a whole the ultimate outcome will not be the same as that for the firm, since there are offsetting forces similar to those described above.

10. Unemployment Caused by Appraisal Price Being or Becoming Too Low.—The parallel but opposite situation to that of Section 6 occurs when employers reduce their wage bids or merely offer wage rates that potential new employees consider too low. The effects of wage cuts under various circumstances were treated in Sections 18 and 19 of Chapter 17. At this point it is merely necessary to repeat that wage cuts will not produce immediate unemployment if they are accepted by the workers. But if the laborers' asking wages are not reduced to match the wage cut, unemployment will result.

The wage bid of an employer tends to reflect his appraisal of the worth of the labor to him and will fall when that labor becomes less valuable. There are at least three major causes: (1) labor-saving changes in technology, (2) reductions in relative prices of other factors, and (3) decreases in the demand for the good made with the

labor in question (or in the demand for the labor service where it is purchased by the ultimate consumer). Each of these causes has been treated in a separate section above. There is also a minor cause which may be described as changes in employer appraisals of labor's bargaining power. These result from observed variations in labor's organizational strength, the volume of unemployment, and the extent of governmental support.

11. Cyclical Unemployment Is the Worst Type.—The major problem in the theory of unemployment, numerically speaking, is why there are *many* people out of work at certain times and what can be done about it. The problems of temporary displacement by business failures, changes in desires, introduction of labor-saving processes and equipment, bargaining tiffs with employers, or slowness in learning about alternative job opportunities and deciding to accept them are all simple compared with the complexity of cyclical fluctuations in the volume of production and employment. Although extended answers cannot be given at this point, it is fitting to review the arguments found in the chapters on wages which indicate in what directions solutions of cyclical unemployment may be sought. These include (1) reducing laborers' asking prices, which may be too high, (2) reducing the cost of complementary factors such as interest rates on capital, (3) readjusting the relative prices of substitute goods, (4) increasing total demand for goods by expanding the volume of government spending, (5) furnishing consumers with subsidies so that those who want to spend more can do so, (6) subsidizing producers by guaranteeing them against losses of certain types, (7) improving producer and consumer confidence about the future in order to reduce liquidity preference, and (8) trying to prevent the use of any one or more of these devices from doing more harm than good through its direct and circular effects upon other causes of the volume of unemployment.

Section 4 analyzed the wage-cut proposal and reached an unfavorable conclusion. Since this chapter deals with business cycles only in connection with wage theory, the other seven proposals cannot now be treated at similar length. We should add only the comment that some people urge raising wages as a method of preventing a recession or starting an upturn. Organized labor is particularly pleased with this argument. But stable consumer buying or even buying increased by wage boosts is not enough to keep total demand stable or increasing. Therefore, it is inadequate to the task. Business demand for inventory accumulation, for modernization of equipment, and for ex-

pansion of capacity may still fluctuate, and usually does. Consumers also vary their own demands by borrowing more at one time than another, though their incomes remain constant. A buying boom for homes, automobiles, refrigerators, and other durable consumption goods is more stimulating to the economy than average or subnormal borrowing and buying in this field. The problem of business cycles is so complex that an analysis of their causes and cures would require at least a whole volume. Only certain negative arguments regarding wage-rate proposals can be mentioned here.

12. Summary.—One of the causes of unemployment may be found in wage rates. At some times wage rigidity is a contributing cause, as when the demand for a particular commodity falls. If the decline in demand is general, as in a business recession, wage rigidity probably does more good than harm. In most cases reduced wage rates would increase unemployment through their roundabout effect upon consumer purchasing power. To seek increased wages at such times probably would also cause unemployment. Strikes would be longer and less successful. Some firms would close entirely. Others would become more pessimistic about future profit prospects. Total business spending would decline.

Technological changes, demand shifts, and business failures are sources of unemployment at all stages of the cycle. Wage rigidities may be a contributing cause when these changes occur. But that is not sufficient reason for favoring wage cuts to avert this type of unemployment. It seems better to help workers find other jobs at prevailing wage rates and to care for their minimum needs by unemployment insurance and the like during the transition period.

Workers may voluntarily reduce their efficiency by setting limits upon daily output. They think to improve their health thereby, or to spread the work to make it last longer. But at the same time they may promote unemployment. The same is true if they demand higher pay. Gains may exceed losses, or vice versa. The economist can only say that workers should understand the risk before acting. If he is given some goal of social welfare as a judgment criterion, he might attempt an evaluation of particular proposals. But generalizations are difficult unless the whole problem is broken down into smaller groups of typical cases.

Chapter 21

RENT

1. The Problem to Be Considered.—The question of definition is much more important in the discussion of rent than it was for wages. Some prominent writers use the term *rent* in one way, some in another. The man-in-the-street is apt to have a still different idea. We shall see later that the terms *interest* and *profits* present similar semantic problems. Of the four terms in common use in distribution theory, only the term *wages* can be used widely without much fear of being misunderstood.

In this book on price economics, rent will be treated primarily as a price paid (and received) for the use of any durable good over a specified period of time. Adjectives will be used when other meanings are intended. Rent is a use-price, not an exchange price. In this respect it resembles the concepts of wages and interest which are used in this volume. The thing (factor source) which is rented, however, differs from the things which are hired or borrowed. For instance, the category of durable goods includes most tangible objects other than human beings and currency. Durable goods are often sold outright. Human beings are sold only under conditions of slavery. Currency is not bought and sold except in the foreign exchange markets.

Using this concept of rents as prices, we must ask much the same series of questions that we have asked about the other prices previously studied. Some special ones are also included in the following list of topics for the chapter:

1. How are specific rents determined? (Primarily an "institutional" approach.)
   (a) What forces determine the bid prices of demanders? (The concept of "income surplus.")
   (b) What forces determine the supply prices of suppliers?
2. Why do rents differ? (Primarily a "theoretical" approach.)
   (a) The average level of rents at one place or time as compared with another? (The Ricardian concepts of the relative intensity of use and the "product surplus.")

(b) The rent of one durable good as compared with another? (The Ricardian concept of "differential surplus" based on differences in productivity.)

3. What are the effects of rents upon:
   (a) The supply of durable goods? (The concepts of reproducibility and durability.)
   (b) The prices of durable goods? (The concept of "capitalizing" rents.)

4. Do durable goods have "normal rents" comparable to "normal prices" of commodities?
   (a) In the long run? (No change in physical supply.)
   (b) In the short run? (No change in demand attitudes.)

5. How do royalties differ from ordinary rents?

2. Alternative Definitions of "Rent."—The definition of rent as a price paid for the services of a durable good is the concept most widely accepted. That is one of its merits. But economists have often used the word in other ways. Rent has been described, for instance, as the income derived from the use of any durable good, or from land alone, or from any natural resource. In a more general way it has been called the income from any factor fixed in supply, or from any which is scarce for natural or artificial reasons. Some writers emphasize the supply-side approach by identifying rents with surpluses over the costs of production of the thing rented (the factor source). Sometimes the surplus is an excess over the reservation price of the supplier. Chapter 22 will treat this problem of surpluses as a general one applying to all factor incomes. A few writers stress differences in the desirability or productivity of rented objects. The more desirable or more productive ones earn a rent. The least desirable and least productive ones do not.

When rent is defined as a price, it is sometimes called "contractual rent." If the owner of a durable good uses it himself instead of renting it to another, he is said to receive an "imputed rent" equal to the contractual rent he might have obtained. By a similar line of reasoning, the lessee (renter) is also said to receive an "imputed rent" equal to the contractual rent which he pays. But if the "rent" is thought of as something produced by the rented object, the rent imputed to that object may be greater or less than the contractual rent. Because of the confusion that arises from this alternative approach to "imputed rent," the term "income surplus" is offered as a substitute. The idea will be developed in Section 5 below.

The rental agreement between the owner and the user may be of various types. Most formally it is a written contract known as a lease which stipulates the rights and obligations of the two parties during a specified period of time. Ordinarily the lessee is expected to maintain the property in good condition so that at the end of the rental period it may be returned to the owner in substantially the same condition as received, normal wear and tear excepted. The contract may variously specify that certain upkeep expenses shall be borne by the lessor, for instance, or that depletion royalties should be paid by the lessee. Informal arrangements are also common as in the renting of dwellings where month-to-month payments without leases frequently continue indefinitely until terminated by one party or the other. Occasionally lessees sublet the property they have leased to some one else and thus become lessors even though not owners, as when a tenant housewife rents out rooms.

Sometimes the rental agreement is on a contingent basis, with the lessee agreeing to pay the lessor a fraction of the future gross product or gross income. Such agreements are common in farming (share-cropping) and in leases of retail stores. A special form of contingent rents known as royalties will be discussed in a later section. Royalties are usually on a percentage or per unit basis, as ten cents per ton of coal mined, or 15 per cent of the publisher's net on every book sold. The total royalty for any given *time* period is not known in advance. It depends upon the volume of output or sales. Sometimes there is a minimum amount per month or year with a royalty-type payment above that.

3. The Peculiar Nature of Rentable Goods.—The differences between rents and other factor prices stem chiefly from differences in the circumstances of supply. In the first place the units offered for rent are usually much less homogeneous than hours of labor, heterogeneous though the latter may be. Parcels of land differ from one another in size, shape, slope, improvement, location, etc. Buildings usually are rented with the land on which they stand. This introduces still greater heterogeneity. Rental units also differ widely in size, and there is no standard smaller unit into which they may be subdivided. It is therefore rarely possible for potential lessees to decide whether to rent more or to rent less of a given type of property. Usually they must choose between this aggregate or that, neither of which is quite what was wanted. Occasional exceptions may be found in the availability of office, storage, or factory space. Sometimes, but rarely, there are relatively homogeneous acres of farm land for lease on something other than an all-or-none basis.

In the second place, rentable goods differ from other goods in their greater durability. They may be used many times before wearing out or otherwise losing their usefulness. Single-use goods like food or fuel cannot be rented. Repeated-use goods like pianos or land may be either rented or purchased outright. Many durable goods also are purchasable only in large, expensive units so that those with limited capital are forced to rent them if they are to use them at all.

Thirdly, these expensive articles usually take a long time to produce. Some rented goods are not reproducible at all, such as land. Slowness to wear out and slowness to produce increase the fluctuations of rents, as a later section will point out. However, one must realize that the contrasts in durability and reproducibility mentioned here are intended to distinguish rented goods from nonrented goods, not from other factor sources. Laborers are also "durable" and require a long time to reproduce. Capital funds are "durable" in a different sense. They can be produced quickly by banks, although saving large amounts is usually a long process. A more complete comparison must await a later chapter.

The fourth peculiarity of rentable goods is related to their ownership rather than to the goods themselves. Many property owners do not possess more than one rental unit. If they do own more than one, the units are likely to be heterogeneous, like most houses or stores. This contributes to the bargaining which is so common in rental markets. Lessors cannot readily establish a fixed rental price for one or more units if they have only one of a kind to rent. Customary rents are less common than customary interest rates or wages.

4. The Demand for Rentable Goods; Problems of Measuring Incremental Revenue.—The lessee often wants the use of only one rentable unit. His problem is therefore not how many to rent, but how much to pay. He must also decide which of several different ones that may be available will give him the most for his money. This is true whether he is renting a home or leasing an office or a factory site. The ultimate consumer's choice will be determined by weighing the relative subjective merits of the different possibilities against the rents which owners demand for them. This is a utility or indifference problem resembling that arising in the purchase of consumer goods and needs no further treatment here. Our chief concern is with the businessman who must decide how much he can afford to pay for the services of a particular good and then must use his bargaining power to get the good for a smaller rent if possible. His calculations of maximum rent are often guesses rather than com-

putations. He may decide that he is able to pay as much as his competitors are currently paying and therefore offers the going rent. Or he may find that he can get the property only by giving a long lease which requires such a distant look into the future that he cannot possibly make any precise calculation of revenue productivity.

Insofar as a producer does try to compute his probable gain from the lease of a rentable good, his task is that of estimating how much of the total revenue of his enterprise can be ascribed to the use of the rented factor of production. This may be done by the marginal factor revenue approach of the chapter on wages in the rare cases where the rented units are homogeneous. Usually the object to be rented is the only one of its kind used by the firm and therefore the incremental revenue to be attributed to it cannot be computed by applying the marginal revenue formula: $TFR_n - TFR_{n-1}$. For instance, a small merchant rents a store and thus becomes able to conduct business, or he does not and has no total revenue whatever.

Large firms may rent several pieces of real estate. Even though these differ from one another, a manager may be able to calculate how much would be added to his total income if one more building were leased and incorporated into the operations of the enterprise. Another circumstance which may permit the calculation of incremental gain occurs when a going firm of whatever size contemplates renting quarters different from those currently used. The manager may estimate how much the new property will add to (or subtract from) the total present revenue and then compare this amount with the difference in probable rent of the new and the old locations. He has no need to compute what the *total* worth of the rented property will be to his firm.

Nearly all of these decisions involve not only deciding whether the lessor's asking price can profitably be paid, but also which of two rental possibilities offers the best chance of gain. Here a second comparison is introduced: the excess of the estimated income increment from the use of property A over the stipulated rent of A as compared with the similar excess for B, C, etc. The property with the largest margin should obviously be chosen unless there is an opportunity to use bargaining to diminish the asking rent of one of the lessors. The excess in favor of property A when the owner gets the rent he asks may become smaller than that for property B if bargaining forces down the rent for B while the lessor of A remains adamant.

5. Demand Appraisals by the Income Surplus Approach.—If a given rentable good is indispensable to the operation of an enterprise

and, when rented, no other like it will be needed, its worth to the firm may be calculated as a remainder or surplus by assuming that all other costs are known and fixed so that they can be subtracted from total revenue. With this approach an imputed normal profit must be included as a cost, thus reversing the usual view which makes rent a cost and profits the residual amount. Even though the estimates of future revenue, costs, and normal profits are difficult to make, there seems to be no other practicable way to calculate how much rent an enterprise can risk paying for a given piece of property under these circumstances. For instance, before they decide in favor of a certain site, certain chain stores make careful surveys of the number of families in the neighborhood or the number of persons who pass by a given location during a representative week. These surveys afford a basis for estimating gross revenues, while past experience tells the probable costs of business with the expected volume of sales.

A numerical illustration may help to clarify the argument about the use of the income surplus approach to rent *ex ante* (in prospect).

Ex ante: Assume that the manager's expectations are:

Total revenue		$10,000
Total operating and overhead cost excluding rent	$7,500	
Normal profit	1,000	
Total explicit and imputed costs		8,500
Remainder or *income surplus* payable for prospective *rent*		$ 1,500

If a business operation is viewed in *retrospect* instead of in prospect, total revenue and costs will be known. The rent that could have been paid for the property can easily be calculated if the firm makes normal profits. However, if profits exceed normal, then the economists will want to know why. Consider, for example, the following *ex post* illustration for the same firm whose hypothetical *ex ante* calculations led to a payment of $1,500 annual rent.

Ex post: At the end of the year the firm's books show:

Total revenue		$12,000
Total operating and overhead cost excluding rent	$8,300	
Normal profit (imputed)	1,200	
Total explicit and imputed costs		9,500
Remainder or income surplus attributable to rented factor		$ 2,500
Rent actually paid		1,500
Excess, unaccounted for, which may be attributed either to the rented factor, to another factor, or just to luck		$ 1,000

Was this $1,000 excess profit above normal due to the fact that the property was rented for the period at less than it was really worth? This question cannot be answered merely by a process of subtraction based on items in the profit and loss statement. The investigator must examine the circumstances of the case to discover other possible causes, such as (1) whether the property was obtained at a lower rent than competitors paid for similar property, or (2) there happened to be an unanticipated increase in demand for the products of the type of property under lease, or (3) the manager introduced a new cost-reducing process during the year, or (4) a bond issue was refunded at a lowered rate of interest, etc. Further comments on this problem of determining the components of profits in excess of normal must be deferred to the chapters on profits, but at this point it is necessary to insist that to impute any definite amount of rent income to either an owned or a leased factor one must answer questions such as these.

6. Other Ways of Determining Demand Bids for Rentable Goods.—Not all firms make estimates as detailed as the foregoing calculations of income surplus. Many small firms merely look to see if the lessor's asking rental is out of line with that paid by competitors for similar buildings and locations. If the rent is higher, they then guess whether it is worth the difference. If the rent is lower, the enterpriser hunts for a reason. In times of rising business volume and mounting profits, lessors may ask old tenants to pay more. The lessees then have to figure whether they can do better by moving elsewhere. Or they may bargain with the landlord to try to get him to take less than the rent he asks. The cost of moving is usually so great that lessors generally win the argument. They lose if the tenant is able to find a long lease elsewhere at a better rent, or if he is so weak in his field that the higher rent forces him out of business.

The nature of rental markets is such that a demand schedule has little significance in explaining how rents are determined. Under the assumption of supply heterogeneity, each rentable good must be considered separately. Therefore, demand must be for the rent of one good only. A series of potential demand bids by different firms for rental of this good might be ranged from high to low in cumulative fashion. The result would be a formal demand schedule. But a lessor rarely knows about all these potential bids. Therefore, he cannot choose the highest. There is no "intersection of demand and supply curves" in the usual sense. The lessor is likely to accept the

first offer which equals or approximates his reservation rental price. Or he will engage any prospective tenant in a bargaining match.

7. The Determination of Lessor's Asking Rentals.—The owners of rentable property usually do not go to as much effort to determine its probable marginal productivity or income surplus as do potential lessees. Many of them have no thought of using their property themselves and therefore have no alternative save to ask the customary rent or to name a high figure and then yield as little as possible in subsequent bargaining. In reaching a decision about what rent to ask for leases of more than a year's duration, a careful lessor may survey economic trends which influence the probable demand for his property and its productiveness in business. These include movements of population, changes in transportation routes, trends of the business cycle, rise or fall of war demands, etc.

Sometimes the lessor is influenced by his ideas of a fair return on the total cost of the property to him. This may involve merely the application of a certain percentage, such as 1 per cent per month gross on the investment. Or it may include a more carefully calculated depreciation allowance for improvements plus enough to cover upkeep, insurance, taxes, and interest on cost. Occasionally a lessor is influenced to raise his asking rent by the fact that he has recently made improvements which were costly to him and were intended to make the property more attractive to potential lessees. The larger the investment in land, buildings, machinery, or other rentable goods the more carefully are rental calculations likely to be made. Millions of small owners do nothing more than imitate, guess, and higgle.

Renting resembles selling in another respect: the owner frequently incurs "selling expense" in trying to attract lessees. Since most rented property is immobile and cannot be brought to a central display place like a store, it is often quite necessary for lessors to use special contact methods like advertising in newspapers or listing with real estate brokers. The expense of advertising or commissions to brokers should be counted as a cost of doing business to be recovered, if possible, through gross rents. Another cost comparable to selling expense is that of product improvement through initial design or more often through remodeling, repairing, and redecorating.

One of the biggest "costs" to property owners is vacancy. That is why landlords will often make improvements or rental concessions to achieve the certainty of keeping a tenant because they fear having to make greater ones if a vacancy develops and they have to hunt a new lessee. In the rental business "time is of the essence" just as

in the case of the worker hunting for a job. Every day lost without a rent-paying tenant deprives the owner of income with which to meet his unavoidable and continuing fixed costs. The higher the probable vacancy rate, the higher must be the asking price per unit for the periods of occupancy. The landlord often must choose between asking a high rent and risking a high vacancy rate or asking a low rent and having almost continuous occupancy. His problem is something like that of the merchant trying to decide whether to seek his maximum profit through a few sales at a high profit margin per unit or many sales at a small unit profit. Workers resisting wage cuts or negotiating for wage increases face a similar problem. Prevailing rental policies depend very much upon the general state of business as revealed by the average vacancy rate in the community. When it is low, lessors can afford to be "tough," but when it is high, a more conciliatory spirit is profitable. The labor float and vacancy rates are clearly analogous.

8. Vacancies Compared with Unemployment.—Vacancies of rental property resemble the unemployment of labor in that in each case a factor of production remains out of use because owner and potential user fail to agree on the terms of use. There are very few specific cases in which property could not be rented if the rent were low enough or a single worker employed if he were willing to work for next to nothing. But both people and property have maintenance expenses which they seek to cover and which tend to set a minimum below which wages and rents will not fall. Labor's minimum is probably more rigid than that for property, since eating cannot long be postponed. The minimum rent for property in the short run would have to cover taxes and any interest due on mortgage funds. Normal maintenance might be slighted, and insurance reduced or cancelled. The long-run expenses of depreciation and imputed interest may be treated as sunk costs and temporarily ignored. However, ideas of customary, fair, or normal returns usually put a brake upon reductions in both wages and rents long before the "logical" minimum is reached.

Individual instances of vacancy are caused also by the imperfection of the market, the failure of potential lessors and lessees to know about one another's bids and offers. But even as in the case of unemployment, the chief reason for extensive vacancies is the general decline in demand for factors of production which occurs during a business recession. In such times reducing rents is even less likely by itself to bring full utilization of rental property than reducing

wages is to bring full employment. Business property is empty because of the unprofitableness of business enterprise, and rent reductions would bring such a small cut in total cost as to do little to improve profit prospects. Residence property is in less demand chiefly because workers are unemployed and have doubled up. Rents would have to be reduced a long way to give very many unemployed workers the opportunity to have living quarters comparable to those they rented when employed. In short, the demand for residences declines more rapidly than owner attitudes become readjusted downward, and much more rapidly than the physical supply wears out in use.

9. Rents Differ with the Relative Intensity of Use: The Product Surplus above the Intensive Margin.—We turn now from the question of how specific rents are determined to that of changes in the level of rents. This is not a question of relative price levels at different phases of the business cycle. It is related chiefly to changes in the intensity of use of land. In many respects the question has the same answer as that regarding the differences in rents between regions. In one area or at one time the land is more intensively used than in another area or at another time.

The solution of this problem in classical rent theory used two principles. The first was the theory of diminishing marginal physical returns with land fixed and labor as the independent variable. The second was the competitive interchangeability of units of labor. These two principles were combined in Sections 3 and 4 of Chapter 19. The average real wage of workers was shown to equal approximately the marginal product of any one worker. It decreases as the number of workers increases in a given region. Competition prevents any worker from getting more than others like him so long as there is perfect mobility from job to job.[1]

Ricardo used these premises to argue that rent would emerge as soon as population increase forced production past the point of diminishing average returns. See Figure 27 of Chapter 8, or Figure 65 below, which shows a simpler case in which diminishing returns occur immediately. The landlord need not pay each of his two workers more than the marginal product of 7, but his total product is $9 + 7$. Total product, 16, minus total wages, 14, yields a product surplus of 2. Since the workers cannot claim this product, the landlord does. It is his rent.

As the population rises further, rents also rise. When three

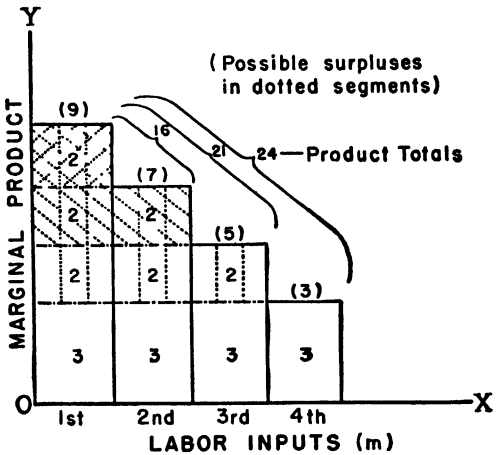
[1] The inaccuracies in these assumptions are not so great as to invalidate their use in the present analysis.

workers must earn a living from this land, their marginal product will decline to 5, and total wages will be 3×5, or 15. Total product is $9 + 7 + 5$, or 21. Rent is now 6. Similarly, with four workers, rent would be 12, and so on. This explains the rise in rents in regions with rising population. The declining marginal product is often said to represent the "intensive margin." (Cf. Fig. 66.) The argument also may be used to explain the difference in representative rents in regions which have different population densities, but which are otherwise similar. The rent of the fixed factor varies directly with the intensity of its use.

FIGURE 65

ESTIMATING RENT BY THE RICARDIAN PRODUCT SURPLUS APPROACH

(Diagrammatic Presentation)



TABULAR PRESENTATION

Inputs of Labor m	Total Product TP	Average Product AP	Marginal Product (Wage Rate) MP	Wage Total $m(MP)$	Product Surplus or Rent R $(TP - mMP)$ or $m(AP - MP)$
1	9	9	9	9	0
2	16	8	7	14	2
3	21	7	5	15	6
4	24	6	3	12	12

The tabular presentation which accompanies Figure 65 shows that rents may be calculated in either of two ways by the Ricardian approach. The first formula is the one used in the numerical illustrations given above:

Rent_{I} = Total Product minus m times Marginal Product or,
$R_{\text{I}} = TP - m(MP)$ (where m = units of labor)

The second formula is based upon the fact that the average product exceeds the marginal product. It is sometimes more useful than the first.

Rent_{II} = m times (Average Product minus Marginal Product) or,
$R_{\text{II}} = m(AP - MP)$ (where m = units of labor)

The application of either formula will give the same result. For instance, when m is 4,

$\text{Rent}_{\text{I}} = TP - m(MP) = 24 - 4(3) = 24 - 12 = 12$, or
$\text{Rent}_{\text{II}} = 4\ (6 - 3) = 4(3) = 12$

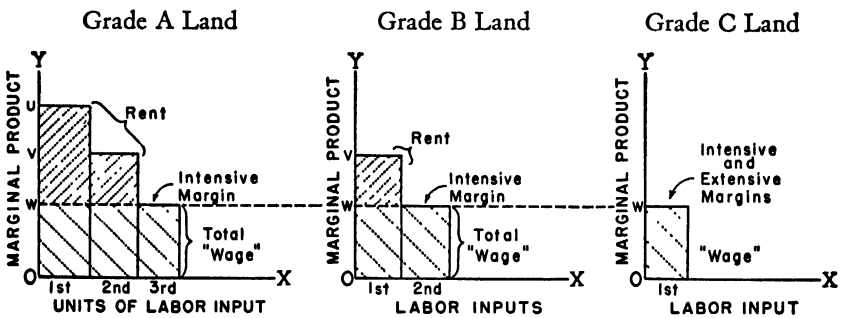
A few qualifications may now be added. The input dose of labor should really be labor-plus-capital. Output almost always requires that the worker use something more than his bare hands. But the result is unchanged. Rent increases while the return to the labor-plus-capital dose decreases. For the moment the question of how labor and capital split their return is irrelevant. Normal profits may also be introduced if desired. A second qualification is a little more difficult. If technology improves so that the total product curve rises, the marginal product will rise, but the rent or product surplus will not. In fact it is more likely to decline. The actual outcome depends upon the shift in the position of the point of diminishing average returns, as shown in Figure 64 of Chapter 19. Rent will also be affected by any change which may occur in the shape of the output curve beyond that point. Finally, the Ricardian approach to rent theory is based on physical product. It may be translated into monetary terms, if necessary, by giving each unit of product a uniform selling price. But this does not make it the same as the income surplus approach of Section 5. The latter is best used for explaining the demand estimates of individual firms. The former is best adapted to explain differences in typical rents between two different time periods or regions. Another application follows in the next section.

10. Rents Differ with the Quality or Productivity of the Rented Good.—The Ricardians also used this product surplus approach to explain the differences between the rents of various pieces of land.

Agricultural land obviously differs in its fertility and distance from the market. Urban land attracts more customers in certain desired locations than in others. But if labor and capital are mobile and their units generally interchangeable, then competition must make it impossible for one of these variable units to receive more than another in the same category. Hence they would distribute themselves over different pieces of land so that wherever employed their return would be the same. That is, the product yielded at the intensive margin would be the same on all land regardless of how productive the intramarginal input doses might be on the better lots. If there was any land so poor that the first input dose yielded no more than the additional product at the intensive margin on all other land, this poor land would produce no surplus. Such land was said to be at the "extensive margin" and was called "no rent" land. The following diagram reveals a hypothetical situation on three grades of land, A, B, and C, of which the last is "no rent" land. The intensive margins are equal on all three and the intensive and extensive margins coincide on Grade C land.

FIGURE 66

RENTS EQUAL TO DIFFERENTIAL PRODUCT SURPLUSES



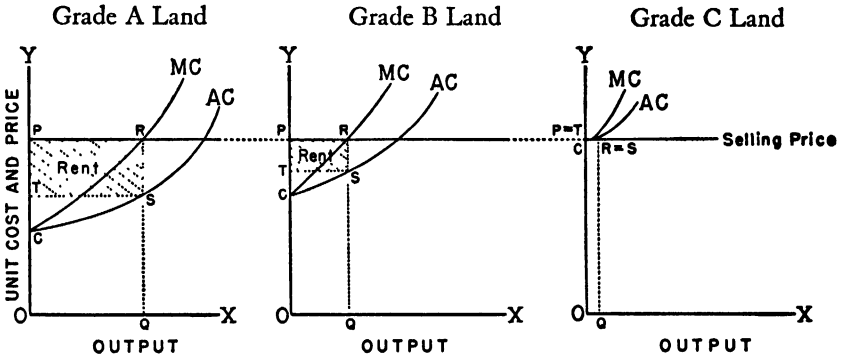
The Ricardian "differential surplus" theory of rent may now be summarized as follows: the rent of any given piece of land will tend to equal the total product of that land minus m times the marginal product when m is the number of units of the composite variable factor and when that factor is applied to such an extent that its marginal product is just equal to the cost of obtaining one unit of it on the competitive market. In symbols the formula is $R = TP - m(MP)$. This resembles the income surplus approach in that it explains the amount of rent as the surplus which a tenant ordinarily could afford to pay and which a land owner has it in his power to

demand. Stated with appropriate qualifications and shorn of reference to the existence of "no rent" land and the alleged historical sequence of utilizing the better lands first, the Ricardian theory is a pretty good explanation of the absolute and relative amounts of rent. But it lacks the market realism of the income surplus explanation offered in the preceding sections. It is best seen as an *ex post* analysis based either on the average output of various grades of land over a considerable period or on the output of a "representative user" of average ability in any given production period.

11. Differential Costs May Be Used Instead of Differential Product Surpluses.—In an effort to bring the Ricardian approach into line with the current pattern of price and distribution theory, it has sometimes been presented in terms of increasing marginal money cost instead of diminishing marginal physical product. For instance, a smoothed curve of marginal money cost may be substituted for the successive product rectangles of the previous figure and the

FIGURE 67

RENTS EQUAL TO DIFFERENTIAL REVENUE SURPLUSES



rent of Grade A and Grade B land may then be shown with or without a contrast with "no rent" Grade C land.

In using this approach one cannot apply a counterpart of the formula $R = TP - mMP$, but must turn to the first formula suggested above: $R = m(AP - MP)$. Revised to conform to the change from physical product to money cost, the rent formula now becomes $R = n(MC - AC)$. The marginal cost remains constant but the average cost rises as one moves from Grade A land to the inferior Grade B and the poorest Grade C, where it becomes equal to marginal cost. In the new formula n replaces m to indicate the

change from inputs as the independent variable to outputs. These outputs are the result of input changes but must be used in any diagram which measures on *OY* the cost per unit of output. The "differential" concept is found in the comparison between average cost and marginal cost when the latter is measured at either the "intensive margin" on any land or the "extensive-intensive margin" on Grade C land. For instance, the rent on Grade A land in Figure 67 is diagrammatically *OQ* times *RS*, which is the difference between *RQ* and *SQ*, or the rectangle *PRST*.

This approach using differential revenue surpluses seems more realistic than that using differential product surpluses. Revenues and costs are both measured in money units. It is easier to include in the composite variable such cost-raising factors as transportation to market, capital improvements, and normal profit. The value approach also is conceptually better adapted to the explanation of urban land rents.

The same sort of diagram can be used to demonstrate quite simply the income surplus approach of which it is a variant. It can be applied to any type of marketing situation and any type of rentable good. It requires no comparisons between "grades" of land, nor stipulations about identity between marginal costs and marginal revenue. Let *OP* represent the selling price for the commodity produced by a firm using property A, and *OQ* the output sold or expected to be sold. Then the total revenue becomes *OPRQ*. The total (nonrent) cost of producing the output sold to yield this revenue may be represented by the area under the curve *CR*, or *OCRQ*. The excess of total revenue over total cost is the income surplus area *CPR* which is, of course, equal in area to *PRST*. A similar approach and diagrammatic technique may be used if the average revenue curve *PR* slopes downward to the right.

12. Normal Rents: Long Run, a Supply-Side Concept.—Since rents are imputed or contractual prices paid for the temporary services of durable goods, it is fitting to inquire whether they resemble the exchange prices of commodities in tending toward a "normal" amount. Normal commodity price has been defined as that price which tended to keep the rate of supply constant in relation to a given demand and therefore to remain itself unchanged. The constant rate of supply was explained as being due to the fact that the normal price was equal to average total cost of production plus a normal profit and therefore encouraged neither entry nor exit, neither expansion nor contraction of production. This definition

would hold in the moderately long run for expensive, durable goods just as well as for commodities which sell at a low price and are consumed in one using. Therefore, something like it should be applicable also to the rents or use-prices of those durable goods.

Thus one may speak logically of normal rents as existing in the long run for buildings and other man-made goods which can be reproduced as they wear out in use. Normal rents of this type would be those which were just sufficient to cover the time fraction of the total cost of reproducing it and making it available to the lessee. These generally would be the sum of (1) depreciation based on the fractional part of the useful life represented by the rental period, (2) upkeep expense, taxes, insurance, etc., borne by the lessor, and (3) normal profit to the lessor, which may be further subdivided into (a) interest, (b) wages of management, and (c) a possible compensation for risk of investment. Since it is *future* production that will determine whether there is to be a constant rate of supply of new buildings, the cost of reproduction should really be used in all of these interval cost estimates, not the prior cost of production of those buildings now being leased.

In actual practice, however, long-run normal rents are rarely calculated with any such precision. Usually they are figured as a gross percentage return on the investment which is approximately equal to what has been customary in the past or what can be earned currently on other types of new investment of similar expense, risk, and effort. Furthermore, most owners of rentable goods buy them secondhand from other investors; that is, they do not make them themselves nor buy them from the original producers. Therefore, a concept of short-run normal rent may possibly prove more useful in the analysis of most rental situations than that of normality in the moderately long run.[2]

First, however, the definition of normal price should be broadened to include the demand side as well as the supply side. A normal commodity price or a normal rent then becomes one which stimulates no change in demand or supply which would alter that price or rent.

13. Normal Rents: Short Run, as Amounts and as Percentages. —Short-run normal rents may be said to exist for any type of rented property when there is no trend upward or downward in the exchange price of that property in current buying and selling. The

[2] "Short run" here refers, as in Chapter 12, to the supply side and to goods already produced; in the "moderately long run" goods can be reproduced. On the demand side, the "short run" refers to fixed incomes, while the "long run" permits incomes to be changed by the prices paid.

number of persons actively seeking to buy is approximately equal to the number actively seeking to sell, and there is only the usual turnover of property that occurs when a few owners desire cash and other persons want to invest funds they have accumulated. If the prevailing stable exchange price at any time happens to be equal to the reproduction cost of the rented property, short-run normal rent will equal long-run normal rent, but ordinarily the short-run normal rent will be less than or greater than this amount and will fluctuate much more rapidly.

The best way of generalizing particular instances of short-run normal rent is to express the amounts as percentage of the acquisition cost of the rented good, such as 12 per cent gross on the current exchange prices of such goods. These exchange prices will then tend to remain stable when the going rents yield a percentage return which is generally acceptable for that class of rental property at that time. Any change in the rental demand for property will cause rents to depart from normal. The same effect will also occur if people's ideas change about the percentage return which current buyers of rental property ought to get.

14. Rents Which Deviate from Short-Run Normal Cause Changes in the Demand for Rented Goods: Capitalization.—Fluctuations in the purchase demand for rental property may be described as varying inversely with changes in the short-run normal percentage and directly with changes in expected rents. For instance, let us assume that investors currently accept 12 per cent as about right for gross investment returns on purchases of small apartment buildings in a given area and that a certain building returns $2,400 per year rent. If that rent may reasonably be expected to continue indefinitely into the future, the probable market value of the property may be estimated by the process of capitalizing the rent, i.e., by dividing $2,400 by 0.12 and getting $20,000. If the rent expectation should then drop to only $2,000, the capitalized income would decline to $16,667. But if the prevailing normal rent concept drops from 12 per cent to 10 per cent at the same time, the market value will not be $16,667, but $20,000, the same as before.

All such calculations are approximations only. The margin of error in predicting an exchange price from a known rent and normal percentage is much greater for rented property than in the case of bonds bearing a fixed interest rate. The market for realty is so imperfect that urgent sellers often sell at prices which are "below the

market" in that they represent abnormal yields. On the other hand buyers are frequently so ill-informed, impatient, or gullible that they pay relatively high prices which show subnormal yields on their investments.

It is important to note that land as well as buildings may have a short-run normal rent even though only the latter, and other reproducible goods, may have a long-run normal rent. When a normal rent is expressed as a percentage, it is the same fractional return at any given time for either short-run or long-run normal rents. In the former case the percentage is based on current acquisition cost, but in the latter on reproduction cost. For instance, during a depression a building whose reproduction cost is $10,000 may drop in market value to $5,000. Yet if people think that 15 per cent is a normal rate of return at that time, the property may remain stable in value when the rent is only $750 per year despite the fact that $750 is only 7½ per cent on reproduction cost. Similarly land, which is generally considered nonreproducible and therefore can have no normal *price* whatever, may at any time earn short-run normal rent on acquisition cost. The concept of short-run normal *rent* is really much broader than that of long-run normal rent. The former applies to all goods, both old and new, but the latter must be confined to new goods alone. To put it in another way, rents may be normal without the current exchange price of the rented good being normal.

15. Other Causes of Changes in the Prices of Durable Goods.— Fluctuations in the normal rent percentage are only one out of several important causes of change in the prices of durable goods. Others include changes in the relative desirability of rental property as compared with other forms of investment, and changes in the supply of rental property. For instance, rental property may become more attractive than other investments if rent incomes seem more likely to rise in the future (or to resist declines better) than dividend or interest income. Changes may occur to reduce the going rate of return on other investments, thus making real estate relatively more attractive than before. Speculative buyers, sellers, and producers may at times be numerous enough so that their actions influence the market price of durable goods. A good example is that of real estate in boom periods, such as that of coastal Florida in the 1920's.

On the supply side there may be changes in the rate at which old units are worn out or destroyed, as when fires, earthquakes, or wars accelerate the normal rate of deterioration. Important variations

may also occur in the rate of production of new structures stimulated by changes in the rentals of existing property of the same type or by changes in the cost of producing such property.

In any case one must remember that some demand and supply variables in the realty market are independent of either rents or normal rent percentages. Their changes are not derived from variations in normal rent percentages by the capitalization process. Rather, they tend to change those percentages. If property values rise when rents remain constant, this change both indicates and determines a fall in the normal rent percentage. The same is true if rents fall while exchange prices remain unchanged.

16. Fluctuations in Rents Tend to Be Wider Than Fluctuations in the Prices of Nonrented Commodities.—When changes occur in schedule demand, rents tend to fluctuate more than other prices in both amplitude and duration. This is chiefly because rented objects characteristically have a longer time-elasticity of supply than objects purchased outright. In the first place rented goods are usually more expensive. Both businessmen and ultimate consumers usually buy the relatively cheap things they want, but rent the more costly ones like stores or homes. A second attribute of rented goods is also related to their expensiveness, the relatively long time required to reproduce them. Therefore, they have a long time-elasticity of supply expansion and their exchange prices or rents can rise farther and stay above long-run normal price longer than those of less expensive articles.

In the third place, the lessors of rentable goods more often follow monopolistic types of marketing policy than do sellers of goods. For instance, certain types of machinery like those used in shoe manufacture and record tabulating are withheld from sale by their producers and are obtainable for use only by lease, while at the same time patent controls prevent the entry of potential competitors. Other owners have a natural monopoly like those who own land containing valuable minerals or land which is located in particularly desirable spots. Entry of potential competitors is slowed by the difficulties of discovery or by the impossibility of finding another lot with the same location. Substitutes of an inferior type may be found or made, and they tend to check the upward rise of rents, but they cannot return rents to former levels since they are either imperfect substitutes or they cost more to produce. This is particularly obvious in the case of land when rising rents cause the use of less desirable land elsewhere, but the use of this land does not force the rent of better land

back to zero. Monopoly-type rents cannot be forced downward without breaking the monopoly, i.e., without finding substitute durable goods which offer the user as much or more for his rent payments than those offered by the monopolist.

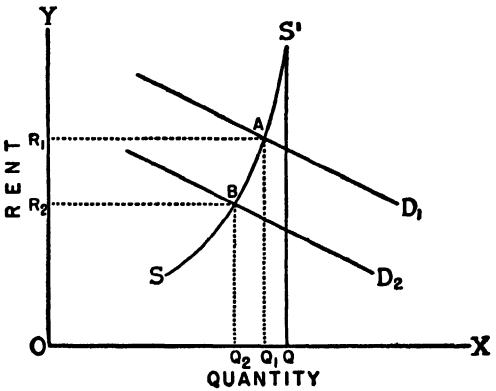
On the downside of rent fluctuations, the characteristically long time-elasticity of supply is due to the extreme durability of most rented objects. For reproducible goods such as buildings, a fall in rents below long-run normal rent will tend to stop construction, but it will not lead to a quick shrinkage in supply. Most buildings have a very long useful life, and this life often can be prolonged by repairs much more cheaply than new buildings can be erected. Hence, in bad times when rents have fallen below the level at which it is profitable to construct new buildings, there is a tendency to decrease rather than to increase the rate of obsolescence. Land in its spatial and location aspects never wears out in use, and its physical supply cannot diminish when contractual or imputed rents decline. Hence, its time-elasticity of supply on the downside is virtually infinite. Stated in another way, the long-run supply curve for land has zero elasticity downward.

17. Substitution and Voluntary Vacancies Reduce the Amplitude of Rent Fluctuations.—Certain qualifications to these broad generalizations must now be made. Usually it is particular types of land or improvements whose fluctuating rents are the subject of analysis rather than the rents of durable goods in general. In such cases supply-elasticity may be achieved by changing the use of certain land and buildings rather than by altering the total supply. For instance, during the war the congestion of certain cities increased the demand for residences while the demand for store leases declined. As a result many store buildings were converted into residences to swell the total when home-building programs proved inadequate. In many agricultural areas farmers have long made a practice of shifting from one crop to another as prices and revenue yields change. This is a method by which an owner-user maximizes his income surplus from alternate uses much as he might maximize his contractual rent by renting to a tenant who intended to plant corn and was therefore able to offer more rent than one who had recently planted wheat and intended to continue that crop. This possibility of substitution into or displacement out of varying employments of durable goods tends to limit the fluctuations in their rents. The more narrowly specialized the use of the good, the less substitution will be possible and the more its rent fluctuations may depart from the general level.

A second qualification of importance is found in the fact that owners of property may temporarily choose to let it lie vacant rather than to cut rents during a period of declining demand. The short-run supply curve for rental property is not perfectly inelastic downward, but slopes to the left as shown in Figure 68 because of the different reservation rent-prices of different owners. The physical supply does not shrink as rent bids fall, but rather the quantity offered for rent. In this respect rental property does not differ significantly from labor or from stocks of commodities in general. Vacancy rates are paral-

FIGURE 68

ELASTICITY OF SHORT-RUN SUPPLY CURVE OF RENTABLE PROPERTY



leled by unemployment and by accumulated inventories of unsold goods.

The degree of elasticity of the short-run supply curves of rentable property varies with the type of property and its ownership. It is probably greater for business property and multiple residence units than for single-family dwellings. Except when protected by long leases, the owners of apartment houses, office buildings, and the like may hesitate to reduce rents to get new tenants during depression periods for fear that the old tenants will be induced to demand similar reductions. Owners of this type usually do not depend upon any particular rentals for their essential income. For them to lose one tenant means that income will be reduced by only a small fraction and they may even have substantial nonrent sources of income to help out in bad times. Rents in this category are therefore likely to be more stable and vacancy rates may go higher than for individual residences. Owners of the latter usually do not own more than a few, and the

loss of a single tenant may seriously diminish a relatively small income. There are also the difficulty and expense of keeping up the grounds around vacant residences, and a provision in many fire insurance policies makes them void after ten days of vacancy. Owners of small homes are more apt to be individualistic in their renting than are apartment or hotel owners and rarely belong to any organization which might urge them to refrain from cutting one another's throats by reducing rents. Under these circumstances a relatively inelastic short-run supply curve is to be expected. Neither group of property owners has anything to compare with the collective bargaining organizations of workers, since prospective tenants have in general no greater bargaining strength than lessors.

18. Royalties Are a Special Form of Rents.—To complete our study of rents it is necessary to consider "wasting assets" as well as land, buildings, and machinery. These include particularly various types of mineral deposits such as oil wells and coal mines where the one who uses the property extracts and sells some of the mineral wealth which makes the land desirable. A parallel situation occurs where land is leased for cutting its timber. There are numerous borderline cases including soil depletion by farm crops. The nature of the operations of the lessor is such that he is not expected to return the property in the same condition he received it. Hence he is made to pay for the damage done by including in his rent what is usually called a depletion allowance. The word *royalty* may be used for this concept, but it is more commonly used for the composite payment which includes both a use-price and a depletion allowance.

Royalties also differ from ordinary rents in that usually they are figured at so much per physical unit extracted or produced by the lessee instead of being stated as a certain amount per time unit of use. Thus an oil lease might be for one eighth of the oil recovered instead of for a given number of dollars per year. Royalties distantly resemble the depreciation allowances of building rents in that the lessor may be thought of as trying to get back through them enough to replace the leased good at the end of its useful life. Mineral deposits, however, are often of uncertain extent, and their current value need bear no close relation either to acquisition cost or to cost of replacement. Hence, a time depreciation method of figuring mineral royalties would be neither logical nor feasible.

The rent or use-price element of royalties paid for the right to exploit mineral deposits will vary with the desirability of using one particular area rather than another. This is usually an inverse func-

tion of the costs of extraction, processing, and transportation. Thus a relatively rich ore body near the surface will command a higher lease than a leaner one of similar extent or one located at lower depths. A mine near rail or water transportation will bring more than a less accessible one. Sometimes it is impossible to distinguish between rent and depletion elements in royalty payments, but both are always present.

The term royalty is also applied to payments made by licensees for the right to use privately owned patents. These royalties are also usually figured as a certain dollar or percentage amount per physical unit of use rather than per unit of time like ordinary rents. Patents have a fixed length of life and therefore may yield something like a depreciation or depletion element in their use, but these amounts would be quantitatively difficult if not impossible to calculate. Even if cost of production or acquisition is known, the cost of replacement is not. Replacement would have to be with something different, but of equal value. Current market value, however, is a function of the productivity of the patent in use, i.e., of something approximately equal to its current total royalty including both rent and "depreciation" elements. Physical replacement cost should relate to the depreciation element alone.

Royalties paid to authors by publishers are composite payments. Copyrights have a limited length of life like patents and mineral deposits, so that depreciation or depletion elements form a logical if indefinite amount of royalty payments. Copyrights also have different degrees of desirability or productivity in that certain publications catch the public fancy more than others. The amount of the author's total royalties rises or falls with the volume of sales, and this fluctuation reveals the rent element. Royalty percentages are usually about the same for all manuscripts of a given class, since relative worth to be revealed by future sales cannot often be determined in advance of publication. Patent royalties, on the other hand, are usually different for each patent. Sometimes they reflect an estimated cost savings to be achieved with each unit produced with their use.

19. Summary of Economic Principles in the Analysis of Rent as a Price.

I. Specific rents are determined in most cases by the lessor whose offer price is accepted by the lessee, or by bargaining on the basis of this rent-price.
   A. Lessor supply prices are chiefly a function of custom modified by guesses as to future trends.

B. Lessee demand prices for business uses are chiefly a function of estimated income surpluses. These involve estimates of future income and future costs other than rent.

II. The general level of rents at any given time or place is a function of the following. It is:
   A. Directly related to:
      1. The intensity of demand for the services of rentable goods (based on desire, income, and expected profit from use)
      2. The supply of other factors (labor, capital) combined with the fixed, rented factors (usually land, but also capital goods in short run). (The more of these other factors seeking employment, the higher the rents.)
      3. The cost of producing and renting reproducible goods. (Long-run normal rent includes depreciation, interest on investment, taxes, and operating costs.) (The higher the cost, the higher the rent.)
      4. The length of time required to reproduce rentable goods. (The longer it takes, the higher rents can be forced above long-run normal by rising demand or sudden destruction of part of the supply.)
   B. Inversely related to:
      1. The supply of the rentable goods in question. (The larger the supply, the lower the rent.)
      2. The length of time it takes for the rentable goods to wear out in use. (The longer it takes, the lower rents may fall below long-run normal.)
      3. The ease of substituting other factors when the rent of one type rises. (The easier it is for users to substitute, the less the rent can rise. Ease is a function of technology and the abundance of substitutable factors.)

III. Differences in rents at any given time are a function of:
   A. Differences in the quantities of the respective rented goods
   B. Differences in their physical productivities
   C. Differences in the demand for them, based on both physical and revenue productivities

IV. The short-run prices of rented goods are affected (through "capitalization"):
   A. Directly, by the rent expected
   B. Inversely, by the short-run normal rent percentage

V. Royalties differ from ordinary rents
   A. They include a charge for the depreciation of wasting assets
   B. They are calculated per unit of output not per unit of time

Chapter 22

SURPLUS ELEMENTS IN RENTS AND IN OTHER FACTOR INCOMES

1. Statement of the Problem.—In economic literature rent is often defined as a type of surplus and is sometimes called "economic rent." This book on price economics uses the word *rent* in the popular sense of a price paid for the services of land, buildings, machinery, or any other durable good. The student should understand both concepts. The task of the present chapter is threefold: first, to show how rent-prices may at times contain rent-surpluses; second, to show how rent-prices may at times contain three other types of surpluses; and third, to show how these four surpluses may also appear in other factor prices, here called factor incomes. Certain implications of this approach will also be considered.

The specific questions to be answered are the following:

1. What are the four types of surplus which may appear in factor incomes?
2. When may they appear in specific rents? in wages? in interest (as a price paid or income received)? in profits (as net income of a business unit)?
3. How may tax policies be influenced by concepts of surplus elements in factor incomes?
4. What are the relations between normal rent, normal interest, and normal profits?
5. Do rent-prices determine commodity prices, or vice versa?

2. Four Types of Surpluses May Appear in Factor Incomes.—Any factor price may at times exceed the cost of supplying the factor service. This excess over cost may be called a surplus on the grounds that the cost is all that actually needs to be paid to evoke the factor service. The surplus is "unnecessary." Or it may be called "undeserved," a "chance gain," the "reward for successful bargaining," etc. When received by the seller of the factor service, the surplus becomes part of his factor income.

The minimum factor supply price here called "cost" may be calculated in at least four different ways. Therefore, there are at least

four different types of surpluses. The size of these surplus elements in factor incomes depends upon both the minuend and the subtrahend in the formula:

$$\text{Factor price} - \text{Factor cost} = \text{Surplus}$$

The following table shows four possibilities under this formula:

Four Possible Surpluses in Factor Incomes

Ways of Calculating "Factor Cost"	Suggested Descriptive Term for Each Type of Surplus
I. Short-run minimum supply price for the factor (owner's reservation price)	I. Lessor's surplus (or worker's or lender's surplus; resembles seller's surplus in commodity pricing)
II. Total *past* cost of providing the factor source and its service (usually equal to the pro rata share of the total cost of producing or buying the factor source plus the current cost of maintaining it and offering its services)	II. Residual surplus or accounting profit. (This view treats factor owning and offering as a form of business enterprise.)
III. Total *future* cost of reproducing the factor source plus the cost of maintaining and offering it	III. Economic surplus, more often called "economic rent" (like II, but better concept in explaining expansion and contraction of supply)
IV. The cost of getting a similar factor at the present time	IV. Opportunity-cost surplus

These four surpluses have been defined in general terms to indicate that they do not apply alone to rent payments for the use of durable goods. They may be found also in wage payments, interest incomes, and profits. That is why we call even the third type "economic surplus" rather than "economic rent." It seems more logical to speak of "surplus" elements in rent incomes than to explain how there may be "rent" in "rents," "rent" in "wages," etc. The fourfold approach to surpluses is also more complete than that which defines "economic rent" as *the* surplus element in factor incomes. The next few sections of this chapter will discuss each of the four surpluses in turn, first in relation to rent-prices, and then briefly in relation to other factor incomes.

3. The Lessor's Surplus (I) in Rent Incomes.—Whenever a rent payment exceeds the minimum amount that a lessor would have

been willing to take at that time for the good which he rented, that lessor receives a type of "seller's surplus." This explains his willingness to rent. That is, in order to be induced to rent, the potential lessor must be offered as much as or more than his minimum reservation rent-price.

The difficulty with the concept lies in the definition of this short-run minimum supply price. The economist would be likely to use the marginal cost idea. He would call this minimum rent the additional cost to the factor owner of supplying the service for that time period. But the thing that counts in the marketplace is the view of the factor owner. This is influenced strongly by custom, by memories of past cost, or by a sense of justice and personal ideas of a fair return. Therefore in most cases it is likely to be higher than the bare minimum conceivable to the economist.

If the highest specific rent offered by prospective lessees does not equal or exceed the factor owner's reservation rent, he will withhold his property from use, at least temporarily. If he employs the property himself, this does not constitute withholding from society unless he devotes it to a less productive use than would be made by a probable lessee (like using good farm land for pasture).

A positive lessor's surplus will emerge whenever a factor owner is successful in bargaining and obtains a higher rent than the minimum he was willing to take. It may also be found in those less frequent cases where the factor units offered by various lessors are relatively homogeneous and therefore capable of being assembled into a supply schedule. Supply interacts with demand to determine a uniform rent-price for all lessors which exceeds that which some of them were willing to accept because of their lower initial reservation figures. Least common of all in rental markets, though prevalent in many other types of seller's surpluses, is the situation in which a lessor rents more than one homogeneous unit at a given time. In this case, he may be considered to receive a subjective surplus on each of the intramarginal units.

Lessor surpluses should be contrasted with income surpluses, which are in a different category entirely. This may be shown by the following tabulation of the two methods of calculation:

	Minuend	— *Subtrahend*	= *Remainder*
1.	Specific rent income	— Owner's reservation rent	= Lessor's surplus
2.	Total business revenue	— Total nonrent costs of doing business	= Income surplus

4. Residual Surpluses (II) in Rent Incomes.—The term *rent* is not often explicitly identified with the residual surplus or accounting

profit concept defined in Section 2, but this viewpoint needs definition for future use in connection with proposed changes in tax policy. The residual surplus sometimes received by factor owners in excess of the cost of obtaining a factor source and supplying a factor service is not the same as the residual surplus of the business enterprise which rents and uses the factor. Both surpluses result from the accounting procedures of business units, but the former may be distinguished from the latter by saying that it is a factor-owner surplus and not a factor-user surplus.

A specific rent payment may yield a residual surplus to its recipient if it exceeds the pro rata cost to the owner of obtaining the land, building, or other rentable good plus the costs which the owner-lessor may have to bear in leasing it. Although the average lessor does not set up an accounting system to compute these costs and to derive a residual surplus or deficit, he could do so. He could think of himself as being in the factor-owner business and could figure his annual profit or loss from his purchase and rental operations. If his calculations revealed a profit, the terminology of this section would call that profit "a residual surplus in a rent income."

5. Economic Surpluses (III) in Rent Incomes.—The third of the factor income surpluses is often called "economic rent," and sometimes just plain "rent." For reasons previously explained, it is here called an "economic surplus." Like other surpluses, it may appear at times in either rent, wages, interest, or profits as defined herein. Economic surpluses differ from residual surpluses in being derived from reproduction cost instead of production cost.

This emphasis upon future cost instead of past cost makes the concept of economic surplus very useful in explaining the expansion and contraction of supply when rents deviate from long-run normal rent. When the rent of a reproducible good rises to exceed its pro rata cost of reproduction (and maintenance), its production will appear profitable. The supply of the rentable good will tend to expand through new production. The increase in supply will force specific rents down towards normal again. Whether rents will return exactly to their former level or not depends upon the future cost of production of the rented good. If this cost is higher than before, the new normal rent will be higher. If the reproduction cost is below the past production cost, the new normal rent will be lower.[1] In times of rising prices,

[1] There is a close parallel here with the concepts of increasing, decreasing, and constant cost for commodities as described in Section 7 of Chapter 13. The historical trend of costs and prices, however, should be given chief attention in dealing with the rents of long-lived durable goods. In the former discussion the emphasis was upon the effects of increased demand upon the economies of scale in production.

the desirability of comparing specific rents with reproduction cost rather than with past cost becomes obvious.

When specific rents fall, economic deficits may appear. These tend to stop current production of rentable goods, and the available supply gradually diminishes through wear and tear, fire, and other calamities. As soon as the place is reached where supply is shrinking more rapidly than the demand, rents will turn upward toward a new normal set by the then current cost of reproduction.

6. Economic Surpluses in the Rent of Land and Other Nonreproducible Goods.—If a good cannot be increased in supply as its rents mount, it is said to be nonreproducible. Land in its spatial aspect is the most common illustration. The reproduction cost of a nonreproducible good is zero. Therefore, one may argue that not part, but all of its specific rent at any given time is an economic surplus. That is why land rent has long been held to be the truest form of "rent." But one must be careful not to reason in a circle. Land rent is "true rent" only when *rent* means "economic rent" which is here called one form of surplus, the "economic surplus." And there remains the difficulty of defining "land" so as to eliminate all produced elements it may contain in any given situation which are capable of being reproduced. Defining *rent* as a rent-price paid for the use of *any* durable good seems a much simpler solution of the semantic problem.

Some goods, like buildings, cannot be reproduced quickly, but can be reproduced in the long run. During the time in which their supply is fixed, they resemble land. That is why Alfred Marshall and others who used the economic surplus concept of rent called building rent "quasi-rent." It resembles, but is not quite the same as their "true rent." Further comments on this term will be found in a later section dealing with "normal rents."

7. Opportunity-Cost Surpluses (IV) May Arise in Rent Incomes from Long Leases.—The concept of an opportunity-cost surplus is derived from the experience of those lessors who are successful in executing a long lease when rents are high and then having this contract remain in effect during a period when rents fall. When that happens, each specific rent payment brings the fortunate lessor a surplus over what he could obtain if he did not have a long lease and was forced to rent his property during the period of depressed rents.

The opportunity-cost terminology has been borrowed from the comparable buying-side situation where it is so well established that there seems no point in translating it into something like "oppor-

tunity revenue" merely for the sake of logical consistency. The opportunity-cost surplus which is a gain as seen by the lessor is of course a loss from the viewpoint of the lessee. Furthermore, opportunity-cost deficits probably occur as frequently or nearly as frequently in factor incomes as do opportunity-cost surpluses. Neither bears any necessary relationship to residual surpluses, but the probability is that opportunity-cost surpluses accompany accounting profits in factor ownership more often than they accompany accounting losses. Long-term leases are more willingly given by property owners during periods of expansion and prosperity than during a recession or depression.

The four classifications of surpluses are not mere terminological exercises, but are intended to help explain the differences in the behavior of various specific rents. They may be used also in the supply-side approach to the use-prices of other factors, as will be indicated shortly. Other terms might do just as well as these, provided they revealed equally well the significant differences in factor supply-elasticity and the manner in which these limit and determine use-price fluctuations.

8. Extending the Surplus Concepts to Wage Incomes.—As indicated above, surpluses of the four types which emerge clearly in the field of rent also occur at times in other factor incomes. Wages, for instance, may exceed the minimum amount which a worker would have accepted for a week's work if he had to take less or go without. Or the earlier hours of a day may be worked more willingly than the later hours. Therefore, part of some people's wages may be regarded as a worker's surplus. Residual surpluses, on the other hand, are not so logical in the wage field because of the fact that slavery no longer exists. An adult worker does not cost himself anything in most cases except upkeep; the past expenses of his rearing and training were borne by his parents or by the taxpayers of the community. Workers do not think of trying to make a profit on this investment. Nor does future cost of reproduction usually bear any close relationship to wages except as a worker believes it necessary to get enough to raise his family on the level to which they have been accustomed. For these and other reasons a normal wage concept is very difficult to define and therefore economic surpluses in the wage field must also be elusive.

Nevertheless, one may cogently argue that certain gifted or fortunate individuals receive salaries far above any conceivable past cost of rearing, training, and maintaining physical strength or mental

energy. These lucky persons may be said to receive residual surpluses in their salaries, surpluses which are akin to the high rents received by certain fortunate owners of plots of land which happen to be in great demand and limited supply. These factor surpluses tend to remain undiminished for long periods of time because rare natural abilities, like rare mineral deposits or land locations, cannot be reproduced by man.

Acquired abilities also often have a long time-elasticity of supply upward because of the long time needed for the formal training and experience which makes great surgeons, musicians, executives, etc. Finally, opportunity-cost surpluses occur in the wage field whenever long-time contracts or the force of custom, bargaining power, or paternalism keep wages high in a given line or firm after declining demand has forced them down elsewhere.

9. Surplus Elements in Interest Incomes.—Lender's surpluses and opportunity-cost surpluses are common in the field of interest and are so obvious as to need no explanation. The concept of residual surplus seems applicable to professional lending institutions, but not to individual lenders. The former operate as business units in securing and lending funds. They show profits or losses at the end of the year when receipts are balanced against expenditures. The small saver, however, experiences only the "real cost," if any, of foregoing spending or liquidity, and that cost is hard to express in monetary terms. But if we shift to the concept of an economic surplus, we find that increases and decreases in the rate of interest alternately stimulate and depress a few people's willingness to lend. Therefore, there must be economic surpluses and deficits relative to some normal figure for these individuals, impossible though it is to define norms in numerical terms for lenders in general.

10. Surplus Elements in Profits.—The surplus concepts in the field of profits are best examined through the eyes of a person purchasing a share in a business unit. There is no contract between the owner of potential equity capital and the firm in the making of which the former may say, "You must agree to give me at least this much or else I will not invest." The business ownership relation is essentially a take-what-you-are-given proposition. When a stockholder is dissatisfied with the dividends he is getting, there is nothing he can do to make the firm return his money unless he can get a majority of the other stockholders to agree to end the life of the business and liquidate its assets. Even then the return of his investment intact is improbable, and when dividends are poor he usually cannot find some one

who will buy his share for the price he paid. For these reasons, the concept of a business owner's surplus in the field of profits is either inapplicable or else the minimum supply price must be held equal to zero.

As regards the other surplus concepts, that of accounting profit in the business of business ownership is probably limited to investment trusts and holding companies. The concept of normal profit is widely accepted, even though difficult to define as noted in Section 9 of Chapter 10. Therefore one may speak of economic surpluses when profits are above normal and economic deficits when they are below. Finally, since profits are essentially residual incomes, they cannot be involved in long-term contracts, and opportunity-cost surpluses can hardly be said to exist, although a case might be made for putting participating dividends on preferred stock into this category.

11. Some Tax Problems Involve Consideration of Factor Income Surpluses.—If factor income surpluses are unnecessary payments to certain owners of the factors of production, people may well ask why they need to be paid. Or if it is difficult to keep factor owners from receiving these surpluses, why shouldn't the government aim at them in its tax program in order that the needs of the state be met as painlessly as possible? The answer must be that it should, provided that we can be sure about three things: (1) that reduction or elimination of the surpluses will not diminish the supply of vital factor services, (2) that we can measure the amount of these surpluses and devise taxes which will absorb them, and (3) that other criteria are not more important than painlessness in the collection of taxes. With these three qualifications in mind, let us briefly examine the taxability of the various surpluses that have been described above.

12. The Taxation of Land and Improvements.—For many years the followers of Henry George have contended that the total rent of pure land is an economic surplus and may be taxed away from the land owner without diminishing the supply of land. However it is difficult to estimate in advance just what the economic surplus will be in any future tax period. Both contractual and imputed rents fluctuate from year to year with changes in demand. Even the future average is uncertain. Hence the tax-collector hunting for economic surpluses may decide to be conservative and take only part of the probable land rent in order that private ownership of land shall not be made completely unprofitable. If the taxing agency should confuse lessor's surpluses with economic surpluses and tax

too heavily the rent income from buildings, as well as from land, society might not suffer much in the short run. But in the long run the supply of buildings would diminish and rents would rise to cover the additional tax. The burden of the buildings tax would be shifted to lessees, and lessors would again get a normal return upon their investment.

In practice we find that although land and improvements are usually assessed separately for purposes of taxation, the same tax rate is imposed upon both. There is no attempt to make property taxation conform to the economic surplus viewpoint except as land may be relatively overvalued on the assessment rolls as compared with improvements. This failure to adopt a seemingly logical policy of imposing the property tax burden largely upon landowners is probably due first to a general feeling that it would be unjust to penalize one group of real property owners rather than another, and second to the political power of the land-owning groups who, of course, cultivate the foregoing concept of justice in their own defense.

13. Taxes on Interest and Profits.—There are two other forms of property income that may be taxed, that from loans and from business ownership. Interest income and dividend income are usually taxed on the same general basis as wage and salary income. But the corporation as a form of business unit is usually made to pay a special type of income tax which resembles real property taxes in the use of two implicit and rather vague assumptions. The first is that ability to produce income is sufficient ground for taxation. The second is that the tax rate, ordinarily not very high, will not discourage either business ownership or the corporate form.

Frequently there is added to the corporation income tax a corporate excess profits tax which is clearly based upon the concept of economic surplus. Proponents argue in its favor that even high rates will not diminish either the number of new corporations or the efforts of corporate managers. Opponents claim that in any time except a war emergency excess profits taxation will be harmful to society, particularly when tax rates approach 100 per cent. They say that normal profits alone are not enough to induce most holders of liquid funds to take the risks of business ownership, but there must in addition be the chance to make and keep extra amounts. This is the same as saying that normal profits cannot be defined as a certain *average* percentage return on investment over a period of time or else that this average is well above the going rate of interest.

A decision between the pros and cons of this argument is rendered very difficult by the fact that it involves subjective imponderables. Owners of enterprise capital may even frighten themselves by shouting too loudly about the alleged menace of this form of taxation just as people raised their own fears about the size of the public debt during the New Deal deficit spending period. Further comments about the subjective and highly unstable determinants of the supply price of venture capital must be deferred until the chapters on profits.

14. The Personal Income Tax Involves Surplus Problems.—Another major source of tax revenue, the personal income tax, may also be construed as being aimed at an economic surplus, that portion of our incomes in excess of the amount needed to keep us functioning as workers. Like other normal supply prices, that for labor is uncertain and variable. The most generally accepted definition, that of the minimum cost of living, is annoyingly vague. It may apply to a single worker or to a worker's entire family. It is likely to change with fluctuations in the worker's bank balance, his trips to the movies, the pressures of union officials, and his wife's ability or willingness to economize. Even though hard to pin down, such a minimum is assumed to exist by legislators who propose that each taxpayer be allowed a minimum subsistence exemption for himself and an allowance for his dependents.

People with large incomes clearly have a larger surplus over minimum physical subsistence needs than do those with small incomes. Therefore it is logical to assume that they have a larger economic surplus even though their aspirations regarding living standards may also be higher. This is the justification of progression in income tax rates.

However one should note that the ability-to-pay argument really refers only to the higher *average* rate on higher incomes. It does not similarly justify the present American method of securing this higher average rate on large incomes. We levy a flat percentage as a basic tax on all incomes above a certain minimum and then secure progression by imposing a graduated surtax. This tends to discourage a few people from additional effort or entrepreneurial risk-taking which would be socially desirable. They see the last increment to income taxed at a high percentage, one much higher than the average for their incomes as a whole. If the tax laws were written so that only the lower average percentage were seen by the taxpayer, some of those who now are discouraged by the smallness of the net after taxes would put forth the marginal effort they now withhold. Society would

gain, and the government would receive taxes from a larger surplus than would otherwise exist.

Another proposal of similar intent is that the tax on earned incomes combine a progressive normal tax with a regressive surtax. The taxpayer would be allowed to calculate his normal tax on the income received, say, five years before if it was lower than current income. The increase of present income over past income would be treated as a socially desirable increment. Surtax rates would be applied to it as they are at present to the portion of incomes above the minimum. But the rates would decline with successively higher brackets, instead of rising. A tax law of this type would tend to stimulate marginal effort and to increase the marginal surplus taxable by the government.[2]

Wartime experience has clearly shown that many people with relatively fixed incomes may be forced by the tax-collector to live on less than before without causing them to withhold their services or property from community use. Propaganda which causes people to accept increased taxes as a social necessity may be considered to diminish their minimum factor supply prices and to raise their economic surpluses.

15. Summary of Rent and Surplus Concepts.—The foregoing comments on various forms of factor surpluses and their implications were provoked by disagreement with the efforts of some writers to identify one or more of these surpluses with rent. The latter is best seen simply as a payment for the temporary services of a durable good. When such goods are used by the owner instead of being leased, they help to produce output or direct satisfaction. In this case they may be said to yield a rent type of income the amount of which may be imputed by comparison with the rents actually received for similar goods at the same time. Whether rented or used by the owner, the contribution of rentable goods in the production process may be estimated by the marginal productivity or income surplus methods, the latter being usually the most applicable. Specific rent incomes often contain one or more elements of surplus, but in this they do not differ greatly from other factor incomes. The surpluses arise for several reasons, of which probably the most frequent is inelasticity of supply (or long time-elasticity of supply) in the presence of rising demand, but each type and sometimes each case deserves separate analysis. Only thus can understanding be improved or public policy made more helpful.

[2] Cf. Kenneth E. Boulding, "The Concept of Economic Surplus," *American Economic Review*, Volume 35, December, 1945, p. 868.

16. The Relation Between Normal Rent, Interest, and Profits. —It was suggested above that buying and renting durable goods often can be considered a form of business enterprise like buying and selling commodities. The attractiveness of being a property owner-lessor rises and falls with the general level of demand like any other form of business ownership. The norm about which such fluctuations occur was described in terms of the gross rent income which should be received if people are to find that type of business activity desirable. If the normal gross rent is reduced to a net rent by subtracting depreciation and upkeep expense, the remainder may be considered the normal profit of owning-leasing. If desirable, one may therefore say that normal (net) rent is equal to normal profit and that it is not a distinct concept at all.[3]

From normal-profit rent one may also subtract wages of management and a risk premium, if either is present, and the final remainder will be interest on the investment. At that point it may be argued that if investment in durable goods does not yield as much interest through gross rents as can be obtained from straight loans, people will tend to move out of the former field into the latter and become pure creditors instead of realty owners. Or if property ownership comes to yield more than lending, the number of people buying or building will increase. In other words, the long-run normal (net) rent of reproducible goods tends to equal the going rate of interest with due allowance for differences in risk and management effort. This was Alfred Marshall's view when he said that in the long run the "quasi-rents" of buildings equal interest on reproduction cost, but that in the short run they fluctuate like "true rent," which he described as a "producer's surplus" of total revenue above the prime costs of producing goods with the use of a fixed factor.[4]

In time periods too short for supply adjustment through reproduction or contraction it is also possible to have short-run normal rents equal to interest on *acquisition* cost. In fact, whenever specific rents depart from this norm, the market price of the rented goods tends to be bid upward or downward until the normal relationship is restored.

Another way of supporting the argument about the fundamental identity between normal rent and interest is to describe buildings as

[3] But see the last paragraph of this section.

[4] Marshall's producer's surplus concept closely resembles the income surplus concept developed in Section 4 of Chapter 21. The latter, however, is broader and has been presented not as "true rent," but as an estimating device for calculating factor productivity. It may be used by a lessee, a lessor, or a third party either before or after use of the factor. Cf. Alfred Marshall, *Principles of Economics*, 8th ed., London: Macmillan & Co., Ltd., 1930, pp. 412, 424, 657, 832, etc.

capital goods and interest as the return from the use of capital of all kinds. Thus, by the simple choice of certain definitions rather than others, the term *rent* is therefore confined to returns from the use of land. Some would go still further and say that all land has a "capital value" or is purchased with "capital funds" and therefore is a form of "capital" so that it, too, really earns interest, not something different called rent. However this terminological flight departs widely from the popular idea of rent as payment for the services of any durable good. It cannot be justified since it contributes nothing to our understanding of the magnitude and fluctuations of rent amounts, which is our major objective.

From one viewpoint, it seems desirable to identify normal rent with normal profit rather than with the going rate of interest. This approach permits inclusion of the troublesome items of risk premium and wages of management, which are often much larger for lessors than for lenders. Furthermore, renting property involves prior buying or constructing plus upkeep expense. These make it resemble ordinary business enterprises much more closely than it resembles saving and lending.

Probably the best solution would be to abandon any effort to identify normal rent with either interest or profits. It is useful to point out similarities, but normal rent should be kept as a separate term. There are several good reasons for this. First, it is really a gross item, not net. Second, in all except land rents the depreciation element is usually a much greater part of total expense than in ordinary types of business enterprise. Third, the time frame of reference must not be obscured. Comparisons of rent with normal profit require an emphasis upon future reproduction cost and therefore must be limited to long-run normal rent and to reproducible goods. On the other hand, comparisons with the going rate of interest refer to acquisition costs, short-run normals, and to either reproducible or nonreproducible goods.

17. In One Sense Commodity Prices Determine Rents.—If a factor is reproducible and is destroyed by use, its use-price and exchange-price are influenced as much by the cost of producing the factor as by the demand for it. If, on the other hand, the supply of the factor is fixed, demand changes alone are responsible for *changes* in its price. Therefore if rent is given the narrow definition of payment for the use of a factor which is fixed in supply, then rents cannot fluctuate except as the demand for fixed supply factors fluctuates. This will occur when they become more or less profitable to

use, i.e., when the price of the commodity being produced with their use rises or falls.

According to this line of reasoning it then becomes proper to argue that land rents, and building rents in the short run, are different from other factor costs in at least one respect. They do not fluctuate independently of the prices of commodities produced with their use, but are determined by those prices. The line of causation runs in one direction only, from commodity price to factor rent, and not vice versa. For other factors the causation runs in both directions: from commodity price by way of demand for the factor to factor rent, and from factor rent by way of cost of production of the commodity to commodity price.

For instance, the demand for wheat land and for harvesters rises and falls with the income which farmers can get from growing and selling wheat. If the supply of wheat land is fixed, its rent or exchange price rises or falls solely because of changes in the price of wheat. But the price of harvesting machines is influenced both by these demand changes and also by changes in the cost of producing the machines. In other words, the price of harvesters may vary because of forces external to the wheat market, but wheat land may not. In functional terms, the price of harvesters may be considered either an independent variable or a dependent variable as related to the price of wheat, i.e., either a cause or an effect, but the price of land is only a dependent variable. During a war or a prolonged strike the supply of harvesters might become temporarily inexpansible. The harvester price or rent then would tend to resemble land in that its upward fluctuations would depend solely upon increases in the price of wheat.

18. In Another Sense Rents Determine Commodity Prices.— The preceding arguments have considerable merit *if* the special definition of rent is accepted and all the premises are granted. But they do not add much to our knowledge of the reasons for rent-price magnitudes, fluctuations, and effects. Furthermore, the argument about rent thus defined being an effect of commodity prices and not a cause tends to introduce confusion in two major respects.

In the first place, the reasoning is valid only for *fluctuations* in rents. It does not apply to the relative magnitudes of different rents at the same time, nor to the absolute amount of a certain rent at any given time. In such problems *total supply* and *relative productivity,* even if unchanging, are just as important as the intensity of demand. For instance, rents for office building sites are not likely to rise as high in a city like Los Angeles, situated on a broad plain, as on the

peninsula of San Francisco until the population of the former becomes many times greater than that of the latter.

In the second place, the sole dependence argument must specify that the rented fixed factor is highly specialized and not adapted to other uses. An illustration might be found in a waterfall that is useful only to furnish waterpower. In most cases, however, land has more than one profitable use. Much wheat land is really just "farm land" and may be used for growing corn or for pasture. Residence areas may be taken over for stores, and farms for factory sites. The same is true of buildings and most other man-made durable goods whose supply may be fixed in the short period. Whenever these alternative uses do exist, fluctuations in rents do not depend solely upon changes in the price of one commodity but upon several. They occur whenever variations in prospective income surpluses cause bidders for rival uses to raise or lower their maximum bids. In such cases changes in the rent of land available for producing a given commodity, X, become a cause of changes in the price of X.

Another illustration may be found in the case of a person wanting to open a men's haberdashery in the lobby of a fine hotel who may have to pay a very high rent in order to get space away from a florist or a beauty parlor. The haberdasher's total cost of production will be higher than that of competitors located outdoors a block or two away where rents are lower. Therefore he must ask more for his shirts and hats than they do if he is to survive. If he guesses right and if enough people patronize his store who are willing to pay these high prices, we have a case in which one of the determinants of commodity price is clearly space rent. However, it is not the only cause, for it operates on the supply side and the necessary demand conditions must also be present.

There is the further possibility that the haberdasher might have had the good fortune to secure his space on a long lease at a low rent before others realized how good the location was going to be. In this case the enterpriser is not forced by high rents to charge high prices in order to survive, but may choose to do so because he makes greater profits that way. He is in the position of a monopolist choosing the most profitable price at which to sell. His profit income includes a portion which may be attributed to unpaid rent, i.e., to what we have called an opportunity-cost surplus. This possibility should emphasize the fact that an explanation of either commodity prices or factor prices should always include consideration of *both* demand and supply forces regardless of how inconspicuous one of them may happen to be.

It should be apparent that land is not the only factor about which the foregoing propositions about lines of causation may be made. A change in the price of any factor will alter the cost of production of a firm using that factor and thus influence the supply price of its product. A common illustration may be found in the manner in which rising wages cause firms to ask more for the things they produce. On the other hand a rise in the price (or volume of sales) of a commodity will cause the demand for some factors to increase more than for others. The actual price history of a factor in such circumstances is a function of its technical substitutability and its time-elasticity of supply. Only in the latter respect do the peculiar characteristics of land set it apart from other factors, and even then the difference is often in degree rather than in kind.

19. Summary of the Economic Principles and Concepts Regarding Surplus Elements in Factor Incomes.—All factor incomes at times contain surpluses over the cost of inducing people to offer the factor services in the market. The amount of this surplus (or deficit) is a function of:

1. The size of the factor income (based upon the intensity of demand for the factor service relative to the supply of it)
2. The manner in which the factor service supply price is calculated. There are four possibilities:
   (a) Short-run minimum reservation price
   (b) Past cost of production [includes cost of creating the factor source (depreciation), maintaining it (interest and upkeep), and making it available (management)]
   (c) Future cost of reproduction (similarly computed)
   (d) Opportunity-cost of a current transaction of similar nature

These four subtrahends give rise to four types of surplus which may be called:

(a) Lessor's surplus
(b) Residual surplus (accounting profit in factor ownership)
(c) Economic surplus
(d) Opportunity-cost surplus

All of these surpluses differ from the concept of income surplus defined in the preceding chapter as appearing when a factor's contribution to the gross income of a business unit is computed by subtracting from that gross income the total of all nonrent costs. These

surpluses are found in specific factor payments of all kinds, including wages, interest, and profits as well as rent.

In the long run, residual surpluses, the economic surpluses of reproducible goods, and opportunity-cost surpluses tend to disappear or average out. Net rents in the long run tend to equal interest on investment in reproducible goods (when costs of production are deducted). Short-run departures from this norm were called "quasi-rents" by Marshall and were computed as the excess of specific rents over reproduction cost and operating expenses, thus resembling most closely the economic surplus concept given here. Marshall's "true rent" is the specific rent of an object whose reproduction cost is zero and all of whose rent is therefore economic surplus. "Quasi-rents" arise because of temporary fixity of supply when demand changes. "True rents" occur when there is permanent (upward) inflexibility of supply.[5] Both of these concepts need redefinition and subdivision as attempted in this chapter with its four concepts of surplus. This procedure seems more logical than to move in the opposite direction and consolidate them into a general term *rent.*

[5] This impossibility of increasing supply suggests a supply price of infinity, not zero, but Marshall chose to stress the origin of such rented objects as free gifts of nature.

Chapter 23

SPECIFIC INTEREST RATES: AN INSTITUTIONAL APPROACH

1. Problems Connected with Interest.—We must begin again with the problem of definition. This is particularly important because of the variety of meanings which the word *interest* may have. For this book on price economics, a price-type definition will be used. Unless otherwise indicated, interest here means the price paid (and received) for the temporary use of funds. The payer is called the borrower, and the receiver is called the lender. Other terms associated with borrowing and lending will be explained below.

The questions to be asked regarding this price resemble those which have been asked regarding other prices:

1. What are the *causes* of the *absolute* magnitudes of interest rates?
   (a) Who are the borrowers and what governs the demand bids they make?
   (b) Who are the lenders and what governs the supply prices they ask?
2. What is the nature of "credit" and where do lenders get it?
3. What are the *causes* of the *relative* magnitudes of interest rates?
   (a) For different types of loans—same place and time?
   (b) For the same type of loan:
      (1) At different times?
      (2) In different places?
4. What are the *effects* of changes in interest rates?
   (a) Upon the supply of funds?
   (b) Upon the demand for funds?
5. What are the relations between interest rates, roundabout production, and durable capital goods?
   (a) In a capitalistic society?
   (b) In a centrally planned society?
6. What are the strong and the weak points of the theories of pure interest?
   (a) The time preference theory?
   (b) The liquidity preference theory?
   (c) The cost of production theory?

7. What determines the long-run "normal" rate of interest?
   (a) How is it related to the general equilibrium of all prices and quantities?

There are two general approaches which one may take in answering these questions. The first is "institutional." It examines the actual market situations in which interest rates are determined, paid, and received. The borrowers and lenders are seen as actual businessmen, governments, banks, consumers, etc. Typical cases are described. From this information regarding what actually seems to happen in the country we live in, certain generalizations are reached. Major exceptions are noted. The approach is more inductive than deductive.

The second approach is usually called "theoretical." The term really means that deductive logic is employed more than inductive. Premises are stated in very general terms. The different segments of the funds market are not examined as much as the market as a whole. Attention centers on the entire economy or on all possible economies in the search for universal principles. Short-run phenomena are slighted in favor of long-run norms.

The institutional and theoretical approaches cannot be separated completely. Both propound theories, and both make use of observations of the world about us. But the primary emphasis of this chapter, after some preliminary definitions, will be institutional. It will consider the points raised under the first question in the above outline about the *causes* of actual interest rates. The next chapter will continue the same approach in giving answers to the questions about interest rate differences. Chapter 25 will analyze the last four problems dealing chiefly with interest rates as causes.

2. The Borrower Seeks Improved Purchasing Power.—What is it that a person wants when he borrows? Most borrowers would reply that they want "money." Economists would not disagree, but usually use the word *funds* instead, so that *money* may be used in a more precise sense. But what are these funds and why do borrowers want them? If a man borrows from a bank, he usually gets a bank deposit. He can draw checks against this deposit and pay his bills. If paper money or coins are more convenient than checks for certain purposes, he may secure these at the teller's window by giving the bank a check in exchange. When the same man borrows from friends, they may give him either checks or currency. If it is a corporation which is borrowing on its bonds, lenders will pay for the bonds with checks and these will be converted into a bank deposit.

In a few cases, businessmen borrow special drafts drawn by banks on other banks and then use these drafts to pay for goods, as in importing.

In all of these cases the borrower gets something which enables him to buy or to pay debts more easily than he could before he obtained the funds. Another way of describing it is to say that he borrows purchasing power. This phrase has some advantages, for it indicates that he gets the ability to purchase in the future. The borrowing and the buying do not always occur at the same time. Days, weeks, months, or even years may elapse between them. In some cases the borrower finds he does not have to buy, or changes his mind, and returns the funds without having spent them. But he had the power to purchase, if he had chosen to use his funds that way.

At the other extreme is that very common sort of borrowing where a buyer gets goods at once but defers payment. Here the borrowing and spending are simultaneous. The borrower never even sees the funds he borrows. He buys "on credit," "on open book account," etc. The lender and the seller of goods are one. A great deal of business as well as consumer buying is financed in this way. Sometimes the seller-lender himself has to borrow in order to be able to lend to his customers. He may borrow either from banks or from other seller-lenders from whom he buys. A large network of interdependent lending and borrowing of this type pervades our economy. It expands with the volume of business in good times and contracts during periods of recession and depression.

3. Collateral as Well as Interest Often Must Be Pledged by Borrowers Seeking Loans.—The promise to pay interest is often not enough to persuade a potential lender to part with his funds. The borrower also must offer collateral. Many different forms of property may be used such as land, buildings, automobiles, furniture, jewelry, stocks, and bonds. The lender usually requires that the collateral have a market value in excess of the loan. In large transactions the collateral is deposited with a trustee, such as the trust department of a bank. This trustee must turn over title to the lender if the borrower defaults. If the loan and interest are paid in full, the title deeds are returned to the borrower. Pawnbrokers, on the other hand, hold the collateral themselves, giving a receipt to the borrower.

The collateral aspect of loans is very important. Its use indicates the real nature of the borrowing-lending transaction. There is an exchange between borrower and lender. The borrower gives his note

and collateral pledge to the lender in return for funds. The funds are more readily acceptable by most people than either the borrower's secured note or his collateral. That is why the borrower wants to make the exchange. The lender wants the interest. He also wants the return of his principal. If he has confidence in the borrower he may not demand specific collateral. He may be content merely to rely on his legal claim to the general assets and earning power of the borrower. When he does demand collateral, its excess value is merely his protection against a decline in the value of that collateral during the period of the loan. Further comment upon the lender's viewpoint will be found in the next section.

The foregoing concepts of the borrower's viewpoint in a loan transaction may now be summarized. He borrows funds, which may be called the factor source. (Funds are money, bank deposits, drafts, etc., whose full nature is yet to be explained.) The factor service is the way in which these funds may be used as purchasing power, actual or potential. Interest is the price paid for the temporary exchange of things of less wide acceptability for things of more wide acceptability.

4. Lenders Demand Payment for the Service of Providing Funds Which They Create or Accumulate.—Turning now to the lender's viewpoint, we find that he demands payment for providing funds which he has accumulated in the past or creates at the moment. Commercial banks supply deposit credit to borrowers. This credit is really only the promise of the bank to pay people who have a right to demand that payment. Because of the position which commercial banks occupy in our society, their promises to pay are considered better than the promises of borrowing firms or individuals. Borrowers exchange their own promises to pay for the bank's promises to pay. They pay interest for the privilege of getting credit (promises to pay) of greater acceptability than their own. Lenders demand pay for allowing borrowers to use their credit. There are two reasons which will be studied further in the next chapter but can only be mentioned here. The first is the cost of running the bank which supplies the credit. The other is recompense for bad debts. And of course profits should be included as a cost in a capitalistic society.

Some lenders provide borrowers with the funds which they have accumulated. These are held by them in the form of bank deposits which were obtained by depositing currency or checks received from others. When individuals are lenders, we speak of them as lending their "savings." The latter term will be analyzed further in the next

chapter. When savings banks or loan companies are the lenders, we speak of them as lending their assets or their capital, not their savings. Nevertheless, the funds they lend represent the savings of the people who bought stock in the companies, or the reinvested profits of the companies themselves.

Governments may lend funds which they have borrowed, taxed, saved, or created. In "creating" funds they resemble commercial banks. The government orders paper currency to be printed or mints coins. Both represent government promises to pay. They are forms of credit which people are glad to borrow because of their general acceptability. Coins differ from paper currency chiefly in the fact that they cost more to make. Gold and silver cost more than paper and ink.

5. There Are Various Meanings of Interest as a Price.—Interest may be expressed as an amount or as a percentage of the funds loaned. The *amount* of interest paid per unit of time corresponds with wage payments and rents per unit of time. It is more common, however, to speak of the *rate* of interest. This is not merely a time rate, but a rate per dollar loaned per unit of time. It is the amount of interest for a given time period divided by the face value of the loan. Profits similarly may be expressed either as an amount or as a rate per dollar involved.

Several different types of interest rates should be distinguished. The *face rate* of interest (also called the *nominal rate* or *coupon rate*) is the percentage written on the face of the note. Or it is the amount of interest per time period divided by the principal. Thus a $1,000 bond may stipulate that the interest is to be computed at a face rate of 4 per cent or that the amount is to be $40 per year. Coupon bonds have $40 coupons attached, one to be removed each year. The *going rate* of interest for loans of a given type is the average face rate on new issues at the time they are issued. Or it is the average percentage obtained by dividing the interest amount paid on old issues of the same type by their current selling price, if time to maturity is ignored.[1] Since the securities of a given class are highly substitutable in the eyes of current lenders, these two percentages are usually equal. The going rate is also known as the *yield*. If most of the new ten-year government bonds which pay 3 per cent are being sold at par in a given day or week, the yield on old bonds will also be very close to 3 per cent. Thus, a 4 per cent bond would be selling for $133.33,

[1] When the time to maturity is considered, a more difficult computation gives what is known as "yield to maturity."

or thereabouts. In actual practice, the yield on old issues tends to determine the face rate on new issues, rather than vice versa, but they are clearly interdependent.[2]

Sometimes a *real rate* of interest is distinguished, like that of real wages. This is important in periods of rising or falling prices. It may be computed for a one-year note by the following rule. Compute the purchasing power of money at the end of the year as a percentage of what it was at the beginning of the year. Multiply the principal plus interest by this percentage and subtract the principal. Divide the remainder by the principal and the quotient will be the real rate of interest. Thus, if the purchasing power of money falls from 1.00 to 0.91 because prices rise 10 per cent in a given year, the real rate of interest in a 6 per cent loan is negative: $(0.91 \times 106) - 100 = -0.0354 = -3.54$ per cent. The lender has "made" 6 per cent on his investment, but he possesses less purchasing power at the end of the period than at the beginning. Nevertheless, he is better off than if he had left his funds idle during the period.

6. Comparisons Between Interest on Loans and Rent on Capital Goods.—Loans of funds differ from leases of capital goods in three respects. First, loans are stated in terms of money units which are perfectly divisible, but leases involve separate pieces of property like land or buildings. Second, the property owner in both cases expects to get his property back, but the lender does not insist upon the return of the precise dollars loaned, since all dollars are identical. Third, lenders frequently require the borrower to pledge collateral in excess of the value of the loan and often sue in cases of failure to pay interest or principal. Lessors require no collateral, but they may sue when tenants fail to pay rent, damage the property during occupancy, or refuse to vacate at the end of the lease.

7. Funds Are One Form of "Capital."—Funds are one form of "capital," a word with many meanings. In this book funds will refer to that form known as "capital purchasing power" or "liquid capital." They should be distinguished from "capital goods" or "capital equipment." The latter are tangible, durable goods used in making other goods or services. In the chapters on rent such goods were

[2] The illustration given here is only an approximation. It assumes that the $4 yield is capitalized at 3 per cent as though the annual payment extended indefinitely. More precisely, the current selling price of the 4 per cent bond of finite maturity would be the sum of two present worths: the present worth of the principal discounted at 3 per cent for the time to elapse before maturity, plus the present worth of each of the $4 interest payments discounted for its time distance. The latter may be expressed as the present value of a finite annuity.

said to be leased or rented. Occasionally they may be "borrowed" from friends, but when a payment is made for their use, it is called rent, not interest.

Accountants distinguish among several kinds of "capital." They speak of "working capital" to include not only cash on hand and bank deposits, but also government bonds inventories, and other "quick assets." The term "fixed capital" is used to refer to both land and capital goods owned by the firm. The total value of the securities issued by a corporation often is called the "capitalization" of the firm, or merely its "capital."

When loans are made, the promise to pay interest and repay principal is a credit instrument that may be called a "capital claim." The owner of this note or bond has a claim against the earnings of the firm and also against its "capital assets." The term capital claim may be applied also to leases and to shares of stock. All capital claims entitle the owner to income without working. Whether called rent, interest, or profits, they are all forms of property income. They should be contrasted with wages, which are claims against the income and assets of a firm for work performed.

8. There Are Four Major Classes of Borrowers and Lenders.—Because of the price approach with which this study of interest begins, the various market situations in which interest rates are determined by the forces of demand and supply must be classified and their differences explained. Business borrowing is usually the most important type, but in time of war it may be exceeded by the borrowing of governments. Quantitatively less important are the borrowings of consumers and banks. All of these four groups also appear on the lending side of the market, but the order of importance is different. Business firms are probably the major lenders, banks are second, private individuals next, and governments ordinarily are last.[3] Precise statistics on the distribution of demand and supply quantities are not available, but fortunately they are not essential to an understanding of the manner in which interest rates are determined and their fluctuations occur.

9. Business Borrowing.—Businessmen borrow chiefly because they think they see a profit opportunity which they cannot realize unless they possess additional goods to sell. They may get these goods by buying (borrowing) on open book account from producers or middlemen. Or, they may borrow funds from individuals or from

[3] N. H. Jacoby and R. J. Saulnier, *Business Finance and Banking,* New York: National Bureau of Economic Research, 1947, pp. 1–18.

specialized lenders such as commercial banks. With these funds, business borrowers can then "pay cash" for whatever they think they need, whether machinery, materials, land, or services. Funds also may be borrowed to pay taxes or to meet other pressing needs for cash to pay outstanding obligations. This latter motive for borrowing is defensive rather than offensive in the sense that it seeks to preserve the business as a going organization so that profits can be made in the future by other types of spending. Sometimes business firms borrow to refund existing loans at lower rates of interest, but this again may be included within the general objective of maximizing profit.[4]

Most firms do not borrow *all* of the funds used for buying goods, hiring factors of production, or paying tax obligations. They usually have some funds of their own, and a few firms have enough so that they never need to borrow at all. This is the result of managing operations so as to keep either a strong cash position or adequate investment in government bonds and other securities which can be liquidated when the need for funds develops. Small firms may also call upon the savings of owners which have not yet been put to business use but have been kept in the bank for other purposes. Either small or large firms may avoid borrowing by selling a part interest in the enterprise to people or institutions who are willing to accept the risks of ownership and do not confine their investments to loans. Corporations may sell stock instead of borrowing through bond issues. Individual proprietorships may get funds through conversion into partnerships by the sale of a part interest.

Current revenues pay for most current purchases. The need to raise additional funds may indicate that expansion is planned, that sales are falling more rapidly than purchases, that costs are rising while prices remain constant, that prices are falling while costs remain up, or that something else has happened to disturb the planned balance between income and outgo.

10. Estimating the Worth of Funds to the Firm.—Since business borrowing is usually a supplementary source of funds, the gains to be derived from borrowing and spending are incremental. Funds are divisible, and therefore it is appropriate to estimate these gains by the marginal factor revenue approach. The income surplus

[4] Some business managers go beyond the risk-taking of ordinary buying and selling of goods with equity or borrowed capital. They speculate with funds and sometimes borrow for that purpose. That is, they may borrow today to increase their cash balances for use on some future date when they expect borrowing to be more expensive or more difficult. This may be precautionary action, may be designed to make a speculative profit through purchasing more cheaply than usual, etc.

method may be used when firms make decisions not about how much to borrow, but whether to borrow this or that. Regardless of the use to which the funds are put, whether to buy machines, materials, labor, or to spend in some other way, it is obvious that a profit-minded manager will arrange the sequence of uses so that the most remunerative comes first. Any particular loan may be devoted entirely to one type of use, but even there the principle of diminishing marginal productivity is likely to apply. As a result an individual firm's demand for funds usually may be portrayed by the conventional type of demand curve sloping downward to the right. The more urgent the total need, the more inelastic will be the demand schedule or the more nearly perpendicular the demand curve.

The quantity borrowed by a firm at any given time will depend chiefly upon the magnitude of the particular project to be financed, the firm's own available funds, and the cost of borrowing. The individual borrower often has to accept an interest rate fixed by forces beyond his control, as will be explained below. His decision in such cases is limited to whether he should borrow and how much. If he has a mental picture of relative marginal factor revenues, he will tend to expand his borrowing so long as the marginal factor revenue exceeds the interest cost of borrowing one unit.[5] When these quantities are equal, he will have borrowed the amount which is most profitable for him to borrow. This general analysis of equilibrium at the point of marginal equality follows the familiar pattern of other explanations of quantities purchased and has logical validity under the assumptions. One should recognize, however, that there are many other variables which are often more important than the interest rate in determining when and how much a firm borrows.

In specific cases the particular needs of the moment may seem so pressing that the borrower does not make any calculations but merely says to himself, for instance, "I've got to get $10,000 and I'll get it even though the bank does charge me 6 per cent." Or a large corporation may be planning to build a new factory with an expected cost of $2,000,000 and realizes that the going rate for mortgage bonds of the type it can offer is 4 per cent. Its decision is whether to float the bond issue and build the factory, not how much to borrow. Optimism about future demand, general profit prospects, and the pressure of technological change are likely to be given more weight in the final decision of the board of directors than any argument that the inter-

[5] With a fixed interest rate, marginal factor cost is constant at an amount equal to that rate times the calculating unit. In business borrowing the calculating unit for such comparisons is often very large, $100, $1,000 or more.

est rate on the bond issue is now ¼ or ½ of 1 per cent higher or lower than it has been in the past.

11. Banks May Supply Funds to Business Borrowers.—There are three major markets for funds in which business borrowers and lenders meet and where the rates of interest and the quantities transferred are determined. The first is the commercial bank which is able to lend because people accept its promises to pay more readily than those of the ordinary firms.[6]

Commercial banks have varying policies in marketing their credit. Most small borrowers are asked to pay a fixed interest rate which is not subject to bargaining and does not change much from year to year. A higher rate is often quoted for unsecured loans than for those where collateral is offered.[7] There is little that resembles price competition in this field. The small businessman usually finds that it is easier to borrow from his own bank where he is known than from another bank. He chooses his own bank because it saves him time and paper work. Large borrowers are in a different category. They bargain with their bankers for the lowest possible rate of interest or the best repayment terms. Often there is much interest competition among banks for very large loans from good companies. This competition is particularly keen as between the banks in one city and those in another. New York banks, for instance, may outbid those of Los Angeles for loans to Southern California borrowers.

A comprehensive explanation of how interest rates are determined on bank loans to business firms must be deferred until other borrowing and lending situations have been described. However, it is clear that banks operate like business enterprises in the sale of their credit. They make many types of loans and determine their charges in many ways. In some cases the cost of the loan is dominant. In others, the major factor is the price (interest) charged by competitors, or the bargaining pressure of large borrowers. Custom also plays a part, particularly with small loans. The general level of rates will tend to rise as business demand for funds rises, and fall as demand falls. There is a significant degree of interdependence between the interest rates which prevail for various classes of loans. Bank loans

[6] Banks are fewer and better known than most business firms. They are supervised by state or federal authorities. They also make their credit available through a deposit and checking system which provides great flexibility and convenience. The ability to make checks payable to specific persons and in any amount makes bank credit in this form often preferable even to government credit in the form of paper currency.

[7] Sometimes interest rates differ because of the quality of the collateral (government bonds as opposed to commercial drafts).

to business firms form part of the credit structure, and their rates tend to rise and fall with the others.[8]

12. Interest Rates on New Loans Are Influenced by the Going Rate on Old Loans as Determined on the Security Exchanges.— A second market where the rate of interest is determined is the securities exchange where buyers' bids and sellers' offers are balanced. Except for the fact that buyers are lenders and sellers are borrowers, the securities exchange is much the same as the commodity exchange. There are speculators on both. The different maturity dates of securities influence their market prices somewhat like the different delivery dates for commodities. But for our present purposes, the major similarity is the fact that on both exchanges no single buyer or seller sets the market price. It is determined by the balancing of aggregate demands against aggregate supplies. Security prices rise and fall with changes in the quantities or the prices in particular collective demand and supply schedules.

The most important difference between commodity and security exchange pricing lies in the interdependence of security prices and security interest rates. The connection may be demonstrated in various ways, of which the following sequence is one simplified illustration:

First, let lenders (or speculators) become more eager to buy bonds.

Second, this will raise the price of bond issues already on the market, i.e., securities with fixed interest and fixed maturity dates.

Third, this will decrease the yield of these fixed interest securities. The new purchasers will get less return on their security investment than those who formerly purchased the securities for a lower price. [If a 4 per cent ($4) bond was selling for $100, it yielded 4 per cent. If now its price is bid up to $105, its current yield is only 3.81 per cent.]

Fourth, because of the interdependence of security prices, the diminished yield on the securities in demand will affect the yields or prices of other securities, including new issues.

Fifth, a new issue of very much the same type (same industry, similar maturity and probable risk) may now expect to find buyers willing to pay $105 for a 4 per cent bond, or $100 for a 3.81 per cent bond. (Because of custom and lender psychology, borrowing firms are more apt to use the former alternative in offering their new bonds than the latter. There may also be a

[8] See Sections 17, 18, and 21 below for further comments on the methods and basis of bank lending.

financial advantage in issuing bonds at interest rates which cause the bonds to sell above par. If, for instance, they sell at 101.5 and are called at 104, the calling firm loses less than if they were sold at 100, or at a discount.)

This simplified sequence shows the way in which the going rate of interest is determined for borrowers and lenders who get together in the securities market. The price of securities is influenced by changes in either the schedule demand for funds or the schedule supply of funds, or in both. The change in interest rates occurs simultaneously with the change in security prices as a joint result in the opposite direction.

In the structure of interest rates there is also interdependence. When 4 per cent bonds rise, more risky bonds paying 5 per cent also tend to rise. More conservative 3 per cent bonds do likewise. New issues are influenced in either their offering rate or their offering price by the going rate (the yield) on old issues of the same type. A more complete explanation of the absolute level of particular interest rates again must be postponed until more factual background has been presented.

13. Business Firms May Borrow from Individual Lenders.— Business firms sometimes market their bonds directly to individual lenders by sales "over the counter." More often the bonds are sold to the investing public through investment bankers who either buy and resell or act as commission sellers.[9] In either case the bonds are sold ultimately to the same group of private and institutional lenders who might otherwise have purchased old issues through the organized bond market. The interest rate paid by the borrowers is likely to be much the same whether the marketing is direct or indirect. There is much fluidity of loanable funds, and it is usually easy for both borrowers and lenders to shift from one security market to another.

From the viewpoint of the large borrower, even the very small differences in interest rates or costs of borrowing by alternative methods may be decisive. The causes of such differences do not concern us here, only the fact that the differences are small. Our main problem is to explain interest rates as prices and the causes and effects of changes in those rates.

Business firms may also borrow directly from government lending agencies like the Reconstruction Finance Corporation, or indirectly

[9] Old securities which are not listed on the organized exchanges are also sold "over the counter" by security dealers. These dealers often aid investment bankers in marketing new issues, or operate in that role themselves.

by borrowing from banks which borrow from these agencies.[10] Interest rates under such circumstances are usually lower than if such funds were borrowed from private lenders. Political considerations are more important than economic considerations.

Occasionally large firms sell their bonds directly to insurance companies who have funds to invest. Bargaining may enter into the determination of the precise price at which the issue is bought by the insurance company, but the yield will closely approximate the going rate prevailing for similar securities sold on the open market.

14. The Proximate Supply of Funds Available to Government Borrowers.—Governments borrow for a variety of reasons. These include construction (public works), destruction (war), relief, relending, and tax anticipation.

The federal government sells most of its securities by public offering. The Treasury Department announces how much it wishes to borrow at such and such terms and then waits for bids. When enough bids are received, the books are closed. If more than enough bids come in during the first day, as has often been the case, the lenders are given only a pro rata share of the total. Bidders include banks, financial institutions, business firms, trusts, and private individuals. Certain classes of securities appeal to one group of lenders more than to others. For instance, very short-term, low-interest "certificates" are purchased almost entirely by the Federal Reserve banks, but some are sought by commercial banks and very large corporations. Ten- or twenty-year bonds with higher interest rates usually attract life insurance companies, savings banks, and private individuals. In short, the proximate supply of funds available to the federal government as borrower includes both (1) credit created for the occasion by commercial banks and (2) credit created by banks and governments at other times which has been accumulated as savings by individuals and business firms.

State and local governments also generally make a public announcement of the amount they want to borrow, but usually sell the whole block of securities for a lump sum to the highest bidder—a bank, other financial institution, or syndicate. If purchased by the latter, the bonds are usually resold to the public in smaller amounts to fit the individual purse.[11] Syndicate members prefer to make their

[10] The RFC gets its funds chiefly from the federal government. (See next section.)

[11] The above account is not intended to be all-inclusive. During the Civil War some federal issues were marketed to syndicates. Banks or other buyers of negotiable bonds may resell at any time.

profit by the margin between the buying and the selling price. Ultimate buyers prefer an interest reward, i.e., they are lenders, not enterprisers.

15. Interest and Non-Interest Considerations in Government Borrowing.—Federal borrowing may differ from state and local borrowing in the matter of interest rates. The federal government usually has some power to influence the rates in its favor, but minor governments do not. Both want to keep the interest cost of borrowing as low as possible. That is one reason why they use so many short-term (low-interest) issues. The national government has another method it can use. Through its influence on central bank policy, it may offset its own increased borrowing by causing the central bank to buy government bonds in the open market. Or the bank may depress interest rates in general by reducing the rediscount rate for commercial paper. Other methods, still more indirect, include reducing the legal reserve requirements of member banks, relaxing quality standards for central bank loans, etc. This topic will receive further discussion in the sections which summarize the forces determining interest rates.

Fiscal policy is often a more important consideration to the national government than the interest cost of borrowing. In wartime, for instance, rather than risk adverse economic and psychological reactions to heavier taxes, additional bond issues are sold, regardless of the size of the interest bill. Or the government may consider its major objective to be holding down prices. This requires absorbing purchasing power in the hands of consumers to prevent them from spending it in an inflationary manner. Securities may then be offered to individuals at high rates of interest to make them attractive, and expensive selling campaigns may be employed.

The amounts borrowed by state and local governments likewise depend more upon what the people want at the time than upon the prevailing interest rate. During depressions when interest rates are low, voters are very cautious. They are apt to disapprove bonds for new schools, court houses, and jails, unless, of course, these can be presented as projects which create employment. But in boom periods when interest rates are usually high, people vote much more willingly for all kinds of public works and the bond issues to finance them.

16. The Proximate Supply of Funds Available to Consumer Borrowers.—Consumers borrow to increase their present consumption of goods or services. In this country most of their borrowing is probably on open book accounts carried by the merchants

from whom they buy. Large purchases are often financed by signing contracts to pay on the installment plan. Merchants, in turn, finance such deferred payment sales either out of their own capital funds or out of funds borrowed from banks or private lenders. Consumers wishing to pay cash (or to quiet insistent creditors) borrow directly from banks, finance companies, insurance companies, building and loan associations, pawnbrokers, etc.

When a consumer wishes to borrow funds for his own use in the market, the place where he seeks a loan depends upon his knowledge of opportunities, his connections, and his influence. Those with checking or savings accounts are apt to think first of their bank as a source of funds. Usually they do not bargain, but pay whatever interest it demands. An alternative is to borrow on insurance policies for convenience, secrecy, or lower interest charges in that way. Members of credit unions, which are co-operative banks, can usually obtain better terms and lower interest rates from such organizations than elsewhere. Poor people, or those already heavily in debt, generally resort to finance companies or pawnbrokers. Interest rates on such loans are high both because of the high cost of making and collecting small loans and because of the urgency of the small borrower's need. The state often has intervened with antiusury laws to try to prevent lenders from taking undue advantage of persons in desperate circumstances.

17. Non-Interest Determinants of the Amount of Consumer Borrowing.—The quantity borrowed by consumers is not influenced very much by the rate of interest they must pay. More important determinants are the occurrence of personal emergency needs, changes in income, and changes in the terms of payment. Although extremely important for particular individuals, personal emergencies do not significantly affect fluctuations in the *total* amount borrowed by consumers at any given time. Therefore, they need not be examined in detail in this study of the reasons for interest rates and their changes.

Income variations are very important in determining the amount of consumer borrowing. When incomes fall, some people try to keep up their levels of consumption by borrowing. Others reduce their borrowing either voluntarily or because lenders will not continue to extend credit. The net effect of a business decline and a rise in unemployment seems to be a reduction in total consumer loans outstanding. On the other hand, an upswing in employment and wage rates tends to stimulate borrowing to buy automobiles, radios, refrigerators, homes, etc. The total volume of debt rises with the boom and falls during the recession.

The amount borrowed by consumers seems to be influenced also by the terms on which the loan can be had. In the purchase of durable consumer goods the size of the down payment is important. If it is raised, installment buying declines; if it falls, buying on credit expands. This is clearly shown by the restrictive effect of action to raise down payments on consumer goods during the recent war. On the other hand, the lenient terms available under FHA loans offered a positive stimulus to home purchase.

The argument that consumer borrowing is not appreciably affected by changes in the level of interest rates must not be taken to mean that *individual* lenders may raise their rates with impunity. On the contrary, the large number of rival lenders gives the careful borrower the opportunity to shop around for the best rate and terms even though that only means hunting the pawnshop which will lend him the most on his watch. Banks compete with finance companies for small consumer loans. They also compete among themselves and with building and loan companies for mortgage loans. They may try to get borrower patronage by offering larger loans on a given piece of real estate than will their rivals. This is more common than the interest-cutting which would be analogous to the price-cutting of merchants. Although rivalry does occur among lenders with regard to interest rates, it is much more significant in determining *who* makes the loans than in determining the total amount borrowed.

18. The Proximate Supply of Funds Available to Banker Borrowers.—Commercial banks borrow as well as lend. The sources of funds open to them include other commercial banks, the central bank, or the national government through some special agency such as the Reconstruction Finance Corporation. Loans from one commercial bank to another are usually relatively small and are generally intended to meet temporary needs for funds. These loans can be concluded quickly and with a minimum of red tape. The interest rate is very low and does not constitute a serious deterrent to borrowing. In times of a monetary crisis, banks may refuse to lend or may impose onerous lending terms rather than try to discourage borrowing by raising the interest rate.

When commercial banks borrow from the central bank of the country, the loan may be obtained by signing a formal note for the funds desired. Less commonly, the central bank agrees to rediscount some of the commercial paper held by the borrowing bank. In either case, the rate of interest charged by the central bank is usually influenced by domestic or international fiscal policies. The profit con-

siderations which govern ordinary banks in their lending are absent or unimportant in central bank decisions. National policies also influence the interest rates charged by the special government lending agencies set up to meet emergency situations such as wars or depressions, or to cater to particular groups such as the farmers.

The proximate supply of funds available to banks as borrowers is chiefly the credit of other banks. In a few cases funds may be borrowed from individual savers or business firms. Government lending agencies like the RFC and HOLC, for instance, may secure part of their funds by selling their own bonds on the open market. Savings banks, of course, receive deposits (borrow) before lending. But they are passive, not active, borrowers. Aside from general advertising, they do not approach potential lender-depositors and ask them for loans. Commercial banks may receive deposits of considerable magnitude from those wanting a place of safekeeping or the check-drawing privilege. Subsequently, some of these funds may be loaned, but they are received, not borrowed, funds. There is no specific solicitation and no interest is paid, except in rare instances.

19. The Ultimate Supply of Funds Offered by Lenders.— Lenders either create or accumulate the funds which they lend. The process of "creating" funds is one of extending to the borrower one's own promise to pay. Borrowers want these promises because they are more widely acceptable than their own. Credit of general acceptability is created chiefly by commercial banks and governments. The major lenders of our economy, business firms, usually accept their own credit immediately after creating it, as when they sell on open book account. Except when they allow drafts to be drawn upon them, they issue no credit instruments of general circulation. Business credit therefore is in a different category from bank notes, checks, and government currency.

Created credit may be either spent or saved. Those who borrow created credit usually spend it. People who think of themselves as savers accumulate credit instruments created by others. When they lend, they are really proximate and not ultimate sources of funds. Nevertheless, their ability to hoard saved funds instead of lending them does give them a significant place in the loan market. By increasing or decreasing the amount they are willing to lend, they influence interest rates.

The remainder of this chapter will develop the themes of the two preceding paragraphs.

20. The Nature of Credit and How It Is "Created."—Etymologically the word *credit* is derived from the Latin verb meaning "to trust." This is also the way in which it is popularly used. *Credit* in this sense is merely *the confidence which one person has in another that the latter will fulfill a promise to pay something in the future.* A store selling jewelry on the installment plan may advertise, "Your credit is good." The merchant is trying to convince people that he will accept their promises to pay in the future if they will sign purchase contracts. But he usually limits his sales to persons of good reputation who have jobs or property.

The merchant himself may find that when he wants to buy goods from a distant firm which he does not know very well, he may be told that his credit is not good. But at the same time, he is known to his banker, and the banker is well known throughout the country. Therefore, he offers to give his own note to the bank in exchange for notes of the bank and he pays for the privilege. If the exchange is merely a transfer of demand notes, as when the merchant writes his own check and gives it in exchange for a cashier's check, the payment is called a fee. But if the merchant gives a note, payable in the future, for bank notes or deposits payable now, the payment which balances the exchange is called interest. Inferior credit plus interest is given for superior credit.[12] In this sense credit is not created, it is merely exchanged.[13]

Another way of using the word *credit* is with reference to obligations to pay. Here it is the opposite of the word *debt.* A promise to pay money is known as a debt when viewed from the debtor's position, and as a credit from the creditor's position. Bank deposits are debts of the bank to depositors, but are often called bank credit. These deposits may arise from the borrowing-lending transaction. The borrower executes a note to the bank and receives in exchange a deposit credit. At the moment of exchange each is in debt to the other. Neither the debts nor the credits existed before. Therefore, in this sense of the word, bank lending may be said to "create" credit in the form of bank deposits.

To distinguish this second connotation of the term, it seems best to use adjectives. Bank lending creates *bank* credit. Business lend-

[12] This is the borrower's viewpoint. The lender requires that interest be paid for several reasons, not merely because he considers the loan risky. These additional causes on the supply side will be elaborated below.

[13] This type of credit may be "created" when a potential borrower, Jones, convinces a formerly reluctant lender, Smith, that a loan will be safe. Jones builds confidence in his ability and intent to pay in the future for what he gets today. Contrary to the usual interpretation, it is here the borrower, not the lender, who "creates" credit.

ing creates *business* or *commercial* credit. Some lending institutions (savings banks, etc.) transfer the credit of others, usually commercial bank credit or government credit (currency). The transfer involves accumulation in various ways and subsequent lending. Of course, every lending transaction requires the cooperation of a willing borrower and hence it takes *two* parties, not just one, to create a credit or a debt (in the second sense of an obligation to pay). It is only customary usage which permits us to forget the borrower and give all "credit" to the lender. (A third meaning of the term!)

The semantic confusion between "debts" and "credits" is a serious one. They are merely different ways of looking at the same transactions. One cannot be created without the other. (The "buying" versus the "selling" of goods furnishes a close analogy.) But the connotation of something "bad" is attached to "debts," while "credits" are "good." For instance, creating government debt by borrowing from the banks is widely thought to be bad. But when banks extend credit to the government they are thought to be performing a beneficial service.

21. Credit Instruments and Their Uses.—Credit instruments furnish evidence of an obligation to pay. They are used to transfer these obligations (credits in the second sense) from one person to another. The most common is the personal check drawn against a bank deposit. It is a draft, or order to pay to some one an amount which the bank owes the drawer. Such drafts against bank credit are widely acceptable because of the good credit (first sense) of the bank. Other forms of drafts may be drawn against banks without deposits having been previously created. In foreign trade, for instance, an importer may make an arrangement with his bank whereby, in return for the importer's promise to pay in the future, the bank allows the exporter to draw a draft against it. Commercial drafts are orders by a seller against a buyer commanding him to pay a debt arising from the purchase of goods. These credit instruments may be sold (transferred to others) if the buyer's credit (first sense) is good.[14]

Another very common form of credit instrument is the note, or direct promise to pay. It takes many forms in addition to the simple I.O.U. Most widely seen, though not often clearly recognized for what they signify, are various forms of national currency. "The United States of America will pay to the bearer on demand . . . dollars." Or, "The Federal Reserve Bank of San Francisco will pay

[14] The quality of the seller's credit is also involved when he indorses the draft.

to the bearer on demand . . . dollars." Formerly, we also saw similar promises of national banks, and still earlier, of state banks.

One important thing about these various types of credit instruments in the analysis of interest rates is that they have different degrees of acceptability. A man may trade his own I.O.U. which has limited acceptability for a bank deposit which permits him to draw checks of greater acceptability. One reason for the inferior status of the I.O.U. is its future date of payment. This is true whether the borrower is a private individual, a business firm, a bank, or even a government. In such cases interest may be thought of as comprising in part a payment which supplements a time credit instrument when it is given in exchange for demand instruments, i.e., funds. On other occasions, people trade bank deposits for checks or for note currency, or vice versa. Both are demand instruments, and therefore no interest payments are involved, but occasionally a fee may be charged. Domestic currency may also be exchanged for foreign currency, usually upon payment of a fee to cover bookkeeping and other expenses involved. There have been occasions when one type of currency became more desirable than another for which it ordinarily exchanged at par. Gold coins at times have sold at a premium over paper currency. Paper currency has also been preferred to bank deposits, as in the panic of 1907. But this premium is not the same as interest, since the transfer is not temporary. There is no time interval involved. If there were, the premium would be an interest payment exacted by the lender because there was a scarcity situation he could exploit.

An interesting anomaly that arises from this analysis occurs when the national government borrows from banks. According to most people's judgment, there is no institution with better credit than the government. Why, then, does the government exchange its credit for the (inferior) credit of banks? And pay interest for the privilege? The first answer might be that the government exchanges a time credit instrument (a bond) for a demand credit instrument (a deposit) against which checks may be drawn. But why does not the government issue its own sight drafts or promises to pay? Surely a government note is as widely acceptable as a bank note or check. An answer to this question might be that there is so much public disapproval of this method of using government credit that the more expensive and roundabout method must be used. In a free enterprise economy, people disapprove of government competition with private banks in the credit creation business. Some contend that governments lack sufficient self-restraint and that interest charges on a

mounting public debt furnish a necessary restraining influence upon government spending. Others deny this argument. A third viewpoint is that governments borrow from banks as a subsidy to keep them alive for the benefit of the community or their stockholders.[15]

22. Determinants of Interest Charges for Commercial Bank Credits, Most of Which Are "Created Credits."—Commercial banks are business firms engaged in the selling of services. The major service which they offer is that of the right to use the credit of the bank for limited periods of time. The bank credit which they lend is chiefly credit which they "create." To a certain extent, however, commercial banks operate like savings banks. That is, they receive funds from stockholders and depositors part of which they may lend to others. These represent accumulated funds as distinct from the funds which the commercial banks themselves create. The interest charges which commercial banks exact from borrowers are payments for services rendered. Banks are organized to seek profit. They incur costs in the creation of the services which they sell to lenders. Therefore, many of the principles which govern the determination of commodity prices are useful here.

Like many other types of businesses, commercial banks also have various sidelines, some of which pay their way and some of which do not. Conspicuous among the latter is the provision of checking services for those who maintain deposits but do not borrow. In recent years fees often have been imposed upon small depositors to cover part or all of the cost of the services performed for them.

Commercial banks differ from other types of business in that they take payment for their product in a variety of ways. Some buyers promise to pay within a few days. Others ask the privilege of waiting several months or even many years. Some buyers offer good collateral; others offer only their general assets and earning power. This variety in the form of payments is equivalent to variety in the type of product sold. Reverting to conventional terminology, we may say that banks make many kinds of loans. Like many other enterprises, banks are therefore faced with the accounting and pricing problems associated with joint costs.

If banks follow customary business practices, their interest charges on loans should cover the direct cost of making those loans plus a share of the general overhead. These direct costs include those of investigation, entering the deposit credit, handling check debits, re-

[15] Cf. Anatol Murad, *The Credit System and Public Debt,* unpublished manuscript, Ch. 8.

ceiving or collecting payment, and a share of the bad debt losses associated with each type of loan. Expressed as a time percentage of the total loaned, these costs will differ. Chattel mortgage loans repaid on the installment plan usually are charged the highest interest rates because of their greater bookkeeping expense and default losses. Short-term loans to the government in amounts of hundreds of thousands of dollars at one time require less clerical effort per dollar loaned than a few thousand dollars borrowed by each of a score of merchants for the same length of time.

23. Non-Cost Factors Influence Many Interest Rates of Commercial Banks.—Bank interest rates, however, cannot be explained completely on a straight cost-of-production basis. The risk premium to be charged for bad debt losses is difficult to estimate, since it varies from borrower to borrower and it changes with the phases of the business cycle. There is also the problem of overhead. In good times, when banks are lending large amounts, the fixed charges per unit loaned should be low (and the risk premium, too), but interest rates are not often reduced as a consequence. In bad times these charges are high, but interest rates tend to fall. One reason is the fact that only the safest loans are made. Another is the decreased willingness of borrowers to pay high rates when the value productivity of the borrowed funds is low. And then there is the troublesome question of how much of the checking service expense should be borne by one class of borrowers as opposed to another or by borrowers in general as opposed to nonborrowing depositors. It should be clear, therefore, that bank interest rates like many other prices cannot be explained in terms of cost of production alone. Nor can cost of production be ignored, since banks must make a profit in order to continue in existence.

One of the other important factors on the supply side is the degree of interest competition. This comes from other commercial banks and sometimes from government agencies. But the most important competition in most lending fields is that offered by those who must save before they can lend. As explained in Section 18, this includes private individuals, industrial firms, insurance companies, and financial institutions such as building and loan companies, personal loan companies, savings banks, etc. All of these engage in direct or indirect lending competition with banks. In some fields of lending, competition is so keen that banks ignore overhead costs in setting the interest rates at which they offer funds. In others they are so free from interest rate competition that they may "charge what the traf-

fic will bear." Monopolistic agreements, at least of the tacit type, have occurred in some cities. Outsiders furnish whatever lending competition exists.

Other determinants of bank interest charges include custom, government regulation or assistance, and bank management estimates of future trends in business. Custom has been strong, for instance, in the rural mortgage loan field, holding interest rates virtually fixed for decades at a time. Government regulation, aside from antiusury laws, has been confined chiefly to the types of permissible loans and to the total quantity that banks may lend. This has had an indirect effect on the interest rates charged by banks, as will be explained later.

24. Limits to the Quantity Commercial Banks Can Lend.— There are several apparent limits upon the amount commercial banks can lend, some effective and some ineffective. The first is based upon the fact that many people prefer currency to checks for certain purposes, such as small purchases, payrolls, bribes, gambling, hoards, etc. The quantity thus demanded tends to rise and fall with the volume of business. Therefore, as lending expands, banks must provide increased quantities of currency. This may be obtained from the central bank by selling government bonds or by borrowing. In modern times central banks rarely limit the amount of currency which they stand ready to provide as needed.[16] They only require that the demanding bank furnish good collateral and abide by certain rules in its lending.

The second limit also is more apparent than significant in modern banking practice. In the United States members of the Federal Reserve System must not lend more than a certain multiple of their reserves kept in the form of deposits in the Reserve Banks.[17] But these reserves may be increased by borrowing as needed. If that is not enough, the Board of Governors may cause the reserve ratio to be reduced, which is the same as increasing the multiple. Or the reserve banks may engage in active purchasing of government securities in

[16] Formerly, when currency was chiefly coin or coin certificates, the quantity available could not be increased rapidly. Banks had to plan to keep their loans within that certain multiple of their currency reserves which experience indicated was safe.

[17] Technically, reserves in the form of deposits in Federal Reserve banks must not be less than a certain percentage of the bank's own demand and time deposits. But the chief cause of deposit expansion by an individual bank is bank lending. Therefore, to speak of a loan : reserve ratio rather than a deposit : reserve ratio is to emphasize the most significant aspect of the problem. For the banking system as a whole this is even more true. An individual bank may receive primary deposits of both government currency and checks on other banks. Only the former remains as a nonloan source of deposits when all banks are grouped together.

the open market, thus raising member bank deposit reserves. However, if the government fiscal policy demands restriction, not expansion, the central banking mechanism may be put into reverse and thus limit the lending capacity of member banks.

A third limit is the cost of lending in relation to the interest income. If potential lenders compete actively for a given amount of borrowing business, they may force down interest yields on loans. This is done by interest competition on new loans and by bidding up the price of outstanding securities on the market. Eventually the yield may get so low that banks will prefer not to take the risks of further lending in some fields. Prospective bad debt losses plus operating expenses look larger than probable interest return. Over a period of years bank profits have not been so large as to indicate a wide margin for interest-cutting.

This leads to the fourth limit, the caution of bank managers. In addition to the danger of negative operating income, banks must also consider possible depreciation of outstanding loans at a time when depositors demand more cash than usual. Under the present system of federal deposit insurance in the United States, this risk is less than it used to be. But the fear of depressions and depositor panics still operates as something of a deterrent to indefinite loan expansion.

The fifth limit also is linked to the third. Although individual banks may compete in the open market and may purchase outstanding loans from some other lender, the total amount borrowed cannot be increased very much by banks in general. All that the banks can do is to make lending more attractive by offering lower interest charges or more lenient terms. If borrowers still fail to demand bank credit, there is not much that the banks can do. This is probably the major limit to bank loans and to lending in general. Under modern banking systems the legal lending ceiling is usually well above the aggregate demand for funds. For individual banks this may not be true, but it certainly holds for the banking system as a whole. Aside from the caution and the profit seeking of those who run our banks, there is no limit to their lending. The significance of this point relative to interest rates will be made clear in a later section.

25. Individual and Business Savers of Funds May Lend, Hoard or Invest.—Individuals obtain cash balances or liquid funds by receiving money income, by receiving gifts, or by selling some less liquid asset such as realty, merchandise, or securities. Most of the funds received from current income are usually spent for current consumption or for things needed in the ordinary course of business.

Some of the cash balances derived from periodic income receipts, however, may be considered to be larger than necessary for current needs. These are often called "savings." Persons with large incomes find it much easier to set aside such cash balances than those who have very little to spend for necessary food, clothing, and shelter. At this point, however, the most important consideration is not the origin of savings, but their disposition. They may be hoarded, loaned, or spent.

The decision which influences the disposition of the savings need have no connection with the decision which brought them into being. This is clearly true in the case of windfalls, like unexpected gifts or inheritances. On the other hand, savings out of current income may occur because of the desire for interest income. In that case both saving and lending spring from the same motive. Although saving with the intention of investing to get income is a very common motive, saving against future contingencies is also frequent, particularly among those with small incomes and among business firms. Contingency saving is apt to lead to hoarding in currency or in checking deposits, since liquidity is a major objective. However, persons with low incomes often consider savings banks or postal savings accounts as good as cash, and the interest is attractive. Such persons combine the contingency saving motive with that of earning interest. Business firms frequently invest part of their contingency reserves in short-term government bonds or commercial paper whenever they think the interest income is worth the slight reduction in liquidity which such lending involves. The final possibility is that of the arrival of the contingency itself. This causes the saved funds to be spent, whether taken from hoards or liquidated from loans. The secular upward trend of total savings seems to reveal, however, that contingencies do not occur with nearly the same frequency or magnitude as do savings. Regardless of the initial motive for saving, the ultimate purpose in most cases seems to be the accumulation of wealth for its own sake or for the income it will bring.

Much of the saving done by individuals is not directly loaned by them. It is used to purchase life insurance, where it accumulates as reserves and is loaned by the insurance companies. Or it may be put into savings banks or building and loan companies. Here, strictly speaking, the saver lends his savings to these firms and then they relend in larger amounts to various borrowers. Some saved funds are also used to contribute equity capital to various lending institutions and that which is not needed for fixed equipment and running expenses may be loaned out. All of these firms which stand be-

tween the individual saver and a loan market into which his funds eventually go may be called savings-transfer institutions. They accumulate funds and lend them under the profit-seeking motives of business firms. In this they differ from individual savers and play a different role in the loan markets.

Lending is only one of the ways of "investing" saved funds to earn income, but it is the one most directly related to the problem of interest rate determination.[18] Instead of buying bonds or notes (lending), a person or institution with surplus funds may buy shares of stock, leases, or rental property. They may also purchase things to hold for speculative gain rather than for interest, rent, or dividend income. The presence of these other alternatives reduces the amount of funds available for lending and therefore affects the interest rate.

The various forms of property income are interdependent. They should all be considered parts of a general structure of rates of return on investment of which interest income is a major element. The explanation of the level of these rates of return has already been given in part. A more complete analysis must wait until psychological and institutional forces influencing supply and demand have been further examined in the next chapter.

26. Descriptive Summary of Loans and Interest Rates.—The institutional approach of the foregoing pages has shown that there are many interest rates to be explained. There are different types of loan contracts, different markets for funds, and different ways in which interest rates are determined. The outline on page 397 summarizes the main features of the complex picture and furnishes a background for some concluding comments.

27. Arbitrary Determinants of Interest Rates: Political Motives of Governments.—In addition to the foregoing statement of the economic principles which determine interest rates, political principles should also be noted. In recent years national governments have come to dominate the activities of central banks in the major loan markets. There have been three chief motives. The first was to regulate the imports and exports of gold by controlling the rate of interest on short-term loans in relation to the rates in other countries. The second was to control the business cycle, particularly to combat depressions by "cheap money" which was supposed to induce bank and business borrowing. The third has been to hold down the cost of borrowing to meet anti-unemployment and anti-enemy expendi-

[18] The word *investing* is here used in the popular sense of purchasing some income-yielding property. Other meanings of the word will be considered later.

Major Types of Loan Markets Where Interest Rates Are Determined

Credit Instrument	Market Where Yield Chiefly Determined Via Price or Interest Rate	Principal Borrowers	Principal Lenders
1. Bonds (long term):	1. Bond exchanges	1. Large business firms	1. Banks
(a) Business	Open market	Governments	Other financial institutions
(b) Government	Over-the-counter		Business firms
			Individuals
2. Drafts and acceptances:	2. Central bank (rediscount rate)	2. Business firms	2. Commercial banks
(a) Commercial	Bank (discount rate)		Large business firms
(b) Bank	Open market		
3. Notes:	3.	3.	3.
(a) Secured by mortgages:	(a)	(a)	(a)
(1) Realty	(1) Banks and other lending institutions	(1) Persons and firms	(1) Banks, building and loan companies, insurance companies, persons
(2) Chattel	(2) Finance companies and banks	(2) Persons	(2) Finance companies and banks
(b) Unsecured notes:	(b)	(b)	(b)
(1) Individual and business	(1) Banks and other lending institutions	(1) Persons and firms	(1) Banks and finance companies
(2) Government (short-term)	(2) Bond exchange and open market	(2) Governments	(2) Banks and large firms
4. Open book accounts	4. Business firms (interest rarely made an open charge)	4. Business and persons as buyers of goods	4. Business firms as sellers of goods

tures which, for various reasons, were not financed by taxes. These three motives appear arbitrary because they follow no clear pattern of precedent. However, they help to explain the level of and the changes in interest rates and therefore must be included in interest theory.

The power of central banks to carry out government policies in this regard is exerted chiefly in three loan markets. Government bonds are bought and sold on the securities exchanges. Commercial paper is also bought and sold over the counter. In addition, the central bank sets the rate at which it will lend to member banks by rediscounting or otherwise. Through various substitution options of borrowers and lenders, central bank activity has significant effects upon many interest rates.

If, for instance, the government wants to keep interest rates from rising, it must keep the prices of bonds from falling. It orders the central bank to buy at a certain price all the government bonds offered at that price or lower. This tends to set the general level of rates for other securities, too. So long as the central bank stands willing to lend funds at a given rate (buy bonds at a given figure), other lenders cannot ask more. Interest rates on other types of loans tend to remain fixed at their normal differential from that for the government bonds which the central bank is buying. Therefore, as long as the central bank continues to lend at a fixed rate, there is a plateau in the general supply curve for funds.

The interest rate level set by the government in this way is largely arbitrary. It is based upon fiscal policies rather than economic costs. But ordinarily it cannot depart very much from the going rate which has been set by the joint action of all the demanders and suppliers of funds. The government cannot depress interest rates too far without hurting some of the other lenders. They have costs of production to meet and some of them are very influential in political circles.

In certain markets, however, the interest rate is fixed by the lender and does not change with every change in the yield on government bonds or commercial paper. Commercial banks, for instance, do not very often change the rates at which they offer to lend to various types of business borrowers. Even a major change in the interest rate of the central bank which may affect commercial bank costs of doing business has little if any effect upon their business loan rates. There is likely to be a plateau in the supply curves of commercial banks that resembles the one created by central banks when they offer to buy securities at a fixed price. Each is established as a matter of policy, the one in a quest for maximum profit in the long

run and the other because of political decisions regarding national or group welfare.

The government may also influence the level of interest rates in its role as a borrower. This is done by controlling the rates which it sets on new issues. If the government wants to depress interest rates, it may offer securities paying lower rates than those prevailing in the market and thus by the power of suggestion (including propaganda) induce lenders to take less. The subjective factor in lender decisions must not be minimized. Wartime bond-selling campaigns emphasize this point. On the other hand, if government officials believe that it would be good fiscal policy to raise the general level of interest rates, they may offer new issues at a higher rate than before. All of this activity on the borrowing side must, of course, be reinforced by the supply-side manipulations previously described.

28. Summary.—Interest is the price paid for the temporary use of funds. Borrowers usually want funds to increase their own purchasing power at that time. Funds, therefore, may consist of anything which is more acceptable to sellers or creditors than that which the borrower can offer before he borrows. In most cases, funds are in the form of promises to pay. They are credit instruments. The borrower exchanges his promise to pay for some one else's promise to pay which has greater acceptability. Inferior credit plus a promise to pay interest is exchanged temporarily for superior credit.

This superior credit may have been "created" or "saved." The creation of credit takes place chiefly in our commercial banking system. Business firms also create credit when they "extend" credit to their customers. Governments create credit when they issue currency. The credit instruments of banks and governments circulate as general purchasing power. When received as income, they may be spent as quickly as obtained. They also may be saved, i.e., they may be accumulated for future spending, for lending, or for the satisfaction of more hoarding. When savers lend, they are influenced chiefly by the subjective costs of opportunities foregone. When the creators of credit lend, the expected costs of doing business are the primary consideration.

The explanation of any particular rate requires an institutional study of the motives and practices of the borrowers and lenders involved. Business firms borrow to increase their profits through expanded purchase of goods and services, including durable capital goods. They lend to stimulate sales or to get interest from idle funds. Private individuals borrow to increase consumption, and

lend to get income from savings. Governments borrow to fight wars, to combat depressions, to finance large construction programs, and to meet temporary deficits. They lend for various economic and political objectives.

The lending of accumulated funds is done by individuals, savings banks, loan companies, etc. Firms engaged in the lending business accumulate funds from stockholders, savers, and their own profits. Although it is not their chief function, commercial banks also lend funds accumulated in this way. In addition, they have a source of funds which other firms lack. They receive funds from people who want the convenience of a checking account. Large firms, for instance, may borrow by issuing bonds and then deposit the proceeds into their checking accounts to be used as needed. Some of these deposits remain unused by the depositors and can be loaned by the bank.

The lending of created funds is done chiefly by commercial banks. When they lend, however, there is no way to distinguish between the two sources of the funds loaned, whether received funds or created funds. Every loan begins by the bank creating a deposit against which checks may be drawn. A brief summary can only state that the amount of funds received from depositors is one of the several determinants of the total amount of loans that an individual commercial bank can make.

The interest rates charged by lending firms of either type depend upon a complex group of forces. These include the costs of making and collecting loans, the cost of interest paid to depositors (rare in commercial banks), the allocation of joint costs, the degree of lending competition, the force of custom, etc. Central banks and governments often arbitrarily fix certain interest rates. Or they may exert pressure indirectly as by the purchase or sale of securities on the open market.

No simple explanation of interest rates can portray all of the complex forces involved in any given loan market. The fundamental motives which govern borrowers and lenders must be understood. But one must know also the institutional setting in which these are given expression.

Chapter 24

DIFFERENCES AMONG INTEREST RATES

1. The Problems of This Chapter.—This chapter examines some specific questions regarding interest rates in modern capitalistic society. The institutional background which one must first understand has been outlined in the preceding chapter. The following explanation of interest rate differences furnishes something in the nature of a bridge to the more "theoretical" chapter which follows.

Relative magnitudes are the general theme. These are examined and explained under the following headings:

1. Why do different interest rates exist at any given time and place for loans of different types?
2. Why do interest rates on the same type of loan:
   (a) Differ from one time to another?
   (b) Differ from one place to another?

2. Differences among Interest Rates at the Same Time: I. Risk.—The reasons why some interest rates are usually higher than others may be found chiefly on the supply side. Demand differences also exist but are generally less important. High reservation prices in the supply schedules of high-interest markets often reflect differences in subjective or objective risk.

From a subjective viewpoint, risk refers to the lenders' fear of loss. Some of the principal may not be returned, interest may not be paid, or the lender may be put to expense to collect what should be regained without effort. This fear of loss is greater in some situations than in others. The experience of lenders in the past indicates that certain types of borrowers are less apt to default than others. Loans to large corporations with a long record of meeting their obligations generally are considered less risky than loans to small firms newly formed. Loans to the federal government are usually thought safer than those to small municipalities. Such is the general opinion of lenders.

Interest rate differences at any given time also result from the differences in expectations among lenders, even though the borrowers are of the same type. One banker may be more optimistic about the

future than another who is also lending to local merchants. And noticeable differences may exist between cities. On the other hand, some lenders may lend at lower rates than others because the thought of possible loss does not worry them, as when government agencies make loans for semipolitical purposes.

From an objective viewpoint, risk is the known experience of loss in the past. It may be expressed as a percentage of the amount loaned. Large banks or banking systems have good records and can make such calculations. Others may read what published studies reveal. But the difference in riskiness between different classes of loans can never be given precise quantitative statement. The only definite *risk premium* that exists is the one revealed by the market itself. On the supply side of that market are many different lenders with varying opinions about the degrees of risk involved in different loans. Each separate lender may change his opinion frequently regarding future uncertainties and the risk class into which a given loan should be put. In short, despite the possibility of obtaining various objective measurements of loan risk, it is not definite like mortality risk, fire risk, etc. The subjective element is dominant.

One of the important causes of differences in risk is the length of the loan. We can predict what is likely to happen in the very near future with greater certainty than we can predict distant events. Therefore, the more remote the repayment date is, the greater the risk to the lender of not being able to collect. The second element of risk on long-term loans arises from the lender's own uncertainty regarding his own needs. He must recognize that unforeseen contingencies may develop before the maturity date of the loan. If these force him to sell his security, the price which he can get will be influenced by the going rate (yield) on such securities at the time of sale. If the going rate at that date exceeds the rate at time of purchase, the security will probably sell at less than its purchase price and the lender will lose.[1] For these two reasons, the risk of lending is generally thought to be proportionate to the length of the loan. The rate of increase in yield diminishes rapidly, however, as the duration of the loan increases. After eight or ten years' maturity is reached, additional time has little effect on interest rates.

3. Differences among Interest Rates at the Same Time: II. Cost. —Some loans cost more per dollar to make and to collect than do others. It is therefore proper that they should yield a higher interest

[1] There are several exceptions which complicate the statement of the problem but do not change the fundamental principle. For instance, as a security approaches maturity, the influence of the going rate on its price diminishes.

rate. Prominent are installment loans on consumer durable goods. If losses on bad debts be counted as costs, then the risk differences described in the preceding section might be included here. Indeed, they often are.

Another cost difference that is sometimes important appears in loans by the national government to business, farmers, exporters, shipbuilders, foreign nations, etc. As indicated in Section 27 of Chapter 23, these loans are often made for policy reasons and therefore the interest rate charged is more or less arbitrary. However, in those cases where the government hopes to break even and cost is a factor, government lending costs may be less than those of institutional lenders. If the loaned funds are borrowed in both cases, the government can usually borrow more cheaply and therefore can lend for less. Certain elements of cost are also lacking or much reduced, such as rent or amortization of buildings, taxes, and normal profits.

4. Differences among Interest Rates at the Same Time: III. Competition.—In some fields there is much more interest rate competition than in others. The degree of monopolistic behavior by individual lenders or by groups is very important. The well-known story of Shylock illustrates the extreme case of monopolistic exploitation of a necessitous borrower. Bond purchases on the exchanges illustrate the other extreme of much competition among many lenders. A recent decision of the Interstate Commerce Commission to require open bidding by investment banker syndicates for railroad securities shows the significance of the problem. Discriminatory pricing by lenders of small amounts is not uncommon since secrecy usually prevails and relending does not occur.

There are also differences in the degree of monopsonistic behavior which should not be overlooked. When the federal government is the major borrower and therefore is the chief dependence of those who have funds to lend, it can depress interest rates in its favor. Because of the interdependence of interest rates previously described, this reduces other interest rates, too. Some are more markedly affected than others (state bonds vs. home mortgages) since they are better substitutes in the eyes of the class of lenders affected.

5. Differences among Interest Rates at the Same Time: IV. Demand.—The urgency of demand and the number of demanders also influence the rate of interest in some markets, although supply forces are usually dominant. An outstanding illustration occurred in the spring and summer of 1929 when call loans on stock collateral were in unusual demand because of the speculative boom. Interest

rates on call loans are usually among the lowest of all because of their short duration and very low risk. But when the demand rose many-fold, call rates rose to 6, 10, 15 per cent and even more, thus proving the influence of demand. The supply of call funds from banks, large firms, and other lenders usually had been sufficient to balance aggregate demand at a very low interest rate in the neighborhood of 1 to 2 per cent. When schedule demand increased, additional supply had to come from other lenders less familiar with that loan market, or from less available funds of the usual lenders. Some lenders also probably felt that call loans were more risky than before.

The call loan illustration has been recounted in some detail because it illustrates the importance of the magnitude of demand whenever the supply of funds is not perfectly elastic. To put it in another way, so long as additional funds are forthcoming at constant marginal cost, demand seems unimportant. But when the marginal supply price rises, as it does in all cases beyond a certain magnitude, then demand quantities are very important. In diagrammatic terms, as the demand curve shifts to the right, it eventually encounters a rising portion of the supply curve and the joint result is a rise in the interest rate.

On the other hand there is the case of government borrowing in time of war whose magnitude and urgency cannot be disputed. Even though the amount borrowed each year increases many times over prewar quantities, the interest rate rarely rises much. In fact, the experience of the United States in the second world war shows that the banking system may be used to expand the supply of funds at progressively *lower* rates, not higher. This seems to show that in the government loan market, the supply curve is horizontal, or nearly so, for a very long range. However, if the government had to rely on loans from private individuals, the increased demand would quickly be reflected in rising interest rates. The collective supply curve of funds from this source rises fairly rapidly.

6. The Interdependence of Interest Rates.—There is much interdependence among the various interest rates. An individual may lend his funds in either one or several different markets. If he does not like the yield on one type of credit instrument, he may purchase or lend on another. He may also leave the loan market entirely and seek his investment return through profits or rent instead of interest. The range of options for bank lenders is not quite so wide. They are restricted by law to certain classes of loans. In some cases banks voluntarily limit their lending, such as loans on long-term realty

mortgages. Business firms also may invest their surplus cash balances in several different fields.

On the borrowing side the possibility of substitution is not so great, although it definitely contributes to the interdependence of interest rates. Governments may borrow on either long-term bonds or short-term notes. Business firms have a similar option. They may also borrow from their suppliers on open book account or by requesting banks to discount drafts on their customers. Individuals have a more limited range of choice, although some kinds of loans may be obtained through more than one type of lending institution. In nearly every case, the possibility of shifting from one market to another, though present, is limited. For instance, when either governments or business firms want to borrow large amounts, they must usually sell long-term bond issues. Individuals without real property must rely on chattel mortgages or unsecured notes.

Because of the possibilities of substitution open to both lenders and borrowers, changes in the interest rates which prevail in one field influence those in others. The connection is close in some relationships, and remote in others. The yields on various types of government bonds tend to rise and fall at the same time, but the spread is not always constant. The general level of government bond yields also influences that for the best grade of utility and industrial bonds. These in turn affect medium-grade bonds and very high-grade preferred stocks. Low-grade bonds are speculative and tend to fluctuate more in keeping with speculative stocks. The different types of negotiable paper sell at rates which tend to move in unison.

Certain fields are relatively isolated, although not by any means completely independent of rates elsewhere. These include rural mortgages, and, to a lesser degree, urban mortgage loans on private dwellings. Pawnshop loans are largely in a field by themselves. In less extreme isolation are small loans made by banks or finance companies to private individuals.

Each loan market has its customary sources of demand and supply. These increase or decrease because of the changed desires of the parties involved. The cross-elasticity of interest rates which attracts or repels individuals at the margin does not remove the elements of difference among markets. The bond exchange still determines security prices and yields by the balancing of aggregate bids and offers. Building and loan companies continue to quote interest rates on mortgage loans in much the same manner that merchants quote commodity prices to customers. Banks either quote rates, as to merchants and home mortgage seekers, bargain, as with large business

borrowers seeking medium-term loans, buy on the bond exchange, or buy on the open market. Pawnshops are notorious bargainers, although the chief issue is usually the size of the loan, not the interest rate.

7. The Average Level of Interdependent Rates Will Be Explained Later.—For these reasons a complete explanation of interest rates must examine each market, not just one. Independence in certain respects does not rule out general interdependence. The *average level* of interest rates, or the minimum rate in a series of rates, constitutes one question. It will be treated in the next chapter. But the question of why there are *individual differences* in interest rates and why these are greater at one time than another should be studied by itself, as we are doing in this chapter. In each specific situation, the forces which play the dominant role are different, or have different weights. No one generalization or theory will suffice to explain the absolute or relative level of all interest rates unless one falls back upon the too simplified abstraction "supply and demand."

Nevertheless, there are certain demand or supply forces which have appealed to various writers as being of major importance in borrowing and lending. These have become the center of different theories of interest. Because of their prominence in the literature of the subject and because they cast additional light upon the funds markets discussed in preceding sections, they will be analyzed in the next chapter.

Again it seems wise to interject a note of caution. Although loan markets are interdependent, they are not completely and perfectly interlocked. They extend like links in a slack chain, or even like a mesh net. The movement of any one influences first the movement of the adjacent ones. If stretched far enough the initial movement may thus reach more remote links, but their motion will not be as great as that which occurs in the first market. Some links will be so remote as not to move at all, or they may have independent motion of their own.

8. Differences in Interest Rates at Different Times: Demand Side.—The chief determinant of short-run fluctuations in interest rates lies on the demand side as it does with fluctuations in other prices. Changes in business profit expectations dominate cyclical variations in interest rates. Other fluctuations may be traced to changes in government demands for funds as the result of wars, depressions, or changes in fiscal policies. Long-run trends may be explained by reference to changes in our banking institutions, or sav-

ings habits, the stock of capital goods per capita, and the rate of technological change.

Another way of subdividing the problem is to study separately the different types of lending. This has already been done in the earlier institutional approach. At this point it is only necessary to recall the cyclical stability of some types of interest rates such as those on home mortgage loans and personal consumption loans. Other interest rates are quite unstable, like those of low-grade securities traded on the organized exchanges. Moderate fluctuations are shown by better-grade securities and by bank loans to large business firms. In this section attention will be devoted chiefly to interest rates on bonds, commercial paper, and term loans by banks.

Business demand for funds rises and falls with changes in the profit prospects of business. When the demand for industrial products rises, the profit return on the least productive funds employed exceeds the cost of borrowing such funds. Entrepreneurs are therefore stimulated to borrow and to use additional funds which will likewise yield a surplus over their interest cost. Equality of interest and incremental factor revenue will eventually be reached (at the margin) as borrowing expands. But in the meantime this increased demand for funds tends to raise the interest rate. The converse is true if profit prospects fall.

9. Special Problems of Changes in the Demand for Funds.— Several points in this simplified presentation deserve further elaboration. In the first place, the additional funds may be secured in other ways than by borrowing. These include the sale of stock and the reinvestment of earnings. But these possibilities are limited. Borrowing is the only option on some occasions. Or it seems the best method. Second, the borrowed funds may be used in many ways, for raw materials, for wages, to pay prior debts, or to improve or increase the stock of mechanical equipment. It is incorrect to associate borrowed funds exclusively with capital goods in the sense of buildings or machinery. The funds presumably are put to use in whatever way is best for the borrower, since he will seek to maximize the difference between his earnings from those funds and the interest cost of obtaining them. The term *marginal efficiency of capital* is most appropriately used with reference to increments of funds obtained in *any* way and put to *any* use. "Capital" here does not refer to buildings and machinery, but to what may be called, for want of a better phrase, "disbursed funds."

In the third place, the increased demand for funds is reflected in

the securities markets by an increased supply of securities. This tends to depress the value of securities of the type being offered and to raise the yield to new purchasers. Stated in another way, lenders do not increase their demand for securities as rapidly as the supply increases. Therefore, the interest rate on each type thus affected tends to rise. Through chain substitution this rise is reflected in other markets not directly affected. However, it is probable that the good business conditions which increase the demand for funds in one market will raise it in several other markets also. One outstanding exception is federal borrowing, which tends to decline in good times. Tax receipts expand and the need diminishes for public works to create employment.

The increased demand for funds is also apparent in the calls upon the banks for short-term and medium-term loans. Here interest rates tend to rise chiefly because of diminished interest-rate competition among bank lenders. In bad times when borrowers are few, there is much interest-cutting to get loans. But when the business tide starts rolling in again, bankers stiffen their backs and become more reluctant to make concessions. That is when the rate starts rising on commercial paper and term loans.

Finally, it is well to remember the important part which government or central bank policy may play here. Through their control over open market operations, rediscount rates, legal reserve ratios, and other ways of influencing key interest rates they may offset to a significant degree the fluctuations in business demand for funds.

10. Differences in Interest Rates at Different Times: Supply Side.—The supply side of the picture is chiefly one of shifts from one type of investment of funds to another. In this it is to be contrasted with the demand side, where the shift is chiefly in the total demand for funds.[2] The total supply of newly saved funds seeking investment each year changes with the national income. But the major change in supply with regard to any one funds market comes from the actions of those old investors who are shifting out of one type of property ownership to seek greater income or safety in another.

Changes in the type of commercial bank lending are quite marked

[2] Reference is to business demand. Government demand may shift to offset somewhat the changes in aggregate business demand, as in depressions. Consumer demand does not change significantly except in the field of loans for installment buying of consumer durables, where interest rates are highly rigid. The process of disinvestment which creates the supply of funds in other markets is, of course, a demand for funds in the market where the securities are sold.

between phases of the business cycle. In boom times commercial banks lend chiefly to business firms at relatively high interest rates. In depressions, banks hesitate to make what few business loans are demanded and seek safety instead by purchasing government securities in spite of their low yields.

During the upswing of the business cycle private investors tend to shift from bonds to stock or to other forms of equity holding. Stocks appear more attractive than bonds because their dividends are rising and can expand beyond the fixed maximum interest on bonds. This brings greater opportunity for appreciation in the value of the shares. Business is good and stocks do not appear as unsafe relative to bonds as they did during the depression.

The sale of outstanding bonds that results from this demand shift to equities is partially offset by two compensating changes in demand. First, some people buy bonds with funds formerly held idle in safety deposit boxes or commercial accounts. As prosperity returns, bonds appear more safe than they seemed during the depression. People may also shift in similar fashion from savings deposits to bonds. The second offset is the increase in national income which gives habitual or potential bond buyers more funds with which to buy. However, this is not as important as it might seem, because bondholders themselves will get no additional interest income (except when defaulted bonds pay up). Only if they also have earned income or stock holdings will they benefit. And persons in either of these categories are apt to choose other ways of disposing of additional income.

There is a similar shift within the different classes of interest-earning securities. Banks, for instance, will be asked to make more short-term loans as business demand improves and therefore will feel able to invest larger amounts in less liquid long-term securities.

The result of these shifts of funds among various investment markets is to cause different security yields to fluctuate in different directions and with different amplitudes. High-grade bond yields tend to rise in boom times and to fall in depressions. The sum of new security offerings plus the liquidation of old issues by restless investors more than offsets the influx of funds from new savings, new bank buying, and favorable investment shifts. On the other hand, many people enter the stock market during booms as new, expanding, or transferring investors. Stock prices are forced very high and yields very low. Low-grade bonds tend to follow stock prices because they are somewhat speculative. Medium-grade bond yields come closer to the stock yield pattern than to that for very high-grade

bonds. In periods of acute panic the desire for maximum liquidity sometimes becomes so great that even the best bonds fall in price and their yields rise.

11. Differences in Interest Rates in Different Places: Demand Side.—The third general problem of interest rate differences is that of variations from one place to another. The first two types of difference are eliminated by the stipulation that the comparison occurs at a given time and relative to loans of a particular type. At the outset the analyst must recognize the importance of institutional differences. Between two countries there may be great differences in the type of banking system, the distribution of income, the incentives to save and to invest, the attitudes of the people, and other variables. These all influence either the supply of funds, the demand for funds, the degree of competition on either side of the market, or the amount and direction of government intervention to influence interest rates. Since these influences cannot be systematized into any kind of formal theory, the analyst must be content to examine the differences which result from purely economic factors.

On the demand side, the determinants of business demand should receive first attention. Since cyclical fluctuations tend to follow parallel courses in different places, this factor in business demand may be excluded by the assumption that we are examining two places at the same time. Or one might take the long-run viewpoint and discuss averages over a period of years. This reduces the problem to the question: "Why is business demand for funds in some places larger relative to the supply than it is in others?" Or more simply, "What makes borrower demand large?"

The first answer is related to the profitableness of employing funds in business. That depends in part upon the number of business firms relative to the demand for their products or services. This in turn is a function of the economic maturity of the region. Newly settled or rapidly growing regions tend to increase in buying power more rapidly than firms are established. The average rate of profit is high. Entrepreneurs seek additional funds and are willing to pay well to get them. Another result of economic immaturity is the absence of many lines of business not yet developed. They offer good profit opportunities in addition to those which are already started and seek to expand.

A second reason why the demand for funds is large in some regions is related to technology. Those economies which make extensive use of capital goods need a much larger investment per unit

of finished product than do those with more primitive technology. A large percentage of this capital equipment is purchased with borrowed funds, since it is easily mortgaged and used as collateral for loans. Industrialized countries also have a larger volume of production per capita. This requires more borrowing to finance firms while waiting for the time when they can sell their products.

The rate of change in technology is also important. When the state of the arts remains little changed for a long period of time, business firms gradually pay off their debts from their earnings. They have less need for borrowed funds. But if capital-using inventions follow one another rapidly and cause a rapid rate of obsolescence of equipment, the demand for loans will remain stronger. An offset to this tendency is found, however, in the way in which modern accounting methods establish depreciation reserves for capital equipment. The more equipment, the larger these reserves, and therefore the larger the amount of funds available within the firm to replace old machines with improved new ones. The need for borrowing in the loan markets therefore is an inverse function of the quantity of capital equipment accumulated and a direct function of its rate of obsolescence through technological change.

12. Differences in Interest Rates in Different Places: Supply Side.—On the supply side of funds markets there are two very important differences between regions: the banking systems and the rates of savings. As indicated earlier, the banking system is of paramount importance. A region that must rely upon savings and their transfer to get its capital funds is obviously much more handicapped than one which has a commercial banking system that can create funds by the mere exchange of credit. Regions also differ in the degree of development of their banking systems. Some of them have banks which provide only short-term credits, whereas other banking systems provide both intermediate and long-term credits. Whenever some of these facilities are lacking, it will be more difficult for borrowers to get funds and interest rates will be higher. Finally, the activity and goals of the central bank may differ greatly in two different places. For instance, it may or may not follow a low-interest policy designed to stimulate borrowing. It may make direct loans to business, or deny them. It may work intimately with a government program of industrialization or socialization, or it may not, etc.

If the banking systems are similar, there may be differences in the rates of saving which characterize the people and firms of two areas. Incomes in excess of bare subsistence are essential to saving. Hence,

regions where there are very few people with large incomes will save less than other regions with the same number of people but a larger percentage in the upper brackets. Or with the same relative distribution of incomes, the region with the highest absolute levels of real income will save the most. On this latter point compare the United States and Britain. On the former, compare Britain with India. Both comparisons must be made, of course, on the basis of percentage of national income saved, not on dollar totals or equivalent.

The rate of saving is also a function of the feeling of security. If people are apprehensive regarding the unfortunate effects of illness, unemployment, old age, etc., upon their planes of living, they will save more than if they are happy-go-lucky like certain primitive tribes, or are protected by comprehensive social security systems as in the most advanced societies. If the typical citizen feels that his family, the state, or God will take care of him, he is likely to be less thrifty than if he believes he must or should take care of himself. Differences in the rate of saving are important because they influence the total amount that can be offered for loan at any given time.

There are also differences in regional culture patterns which influence interest rates in two other ways. First, there are differences in the amount loaned out of current savings or accumulated past savings. For instance, the tradition in favor of hoarding large amounts of currency or coin is stronger in some regions than in others. Second, institutional differences may influence the type of loan which most savers are willing to make. Certain loans are accepted in some places as being "sensible," while others are "too risky." Relative interest rates differ accordingly.

Other factors affect the rate of interest through the extent to which saved funds tend to be loaned. Individual savings may become more abundant than necessary to provide the liquid contingency reserve which is desired. Or, as with wealthy people, funds accumulate without effort and are loaned to get interest income for further accumulation. Also important is the presence of intermediate institutions which are highly developed in some areas. These are the savings banks, insurance companies, building and loan associations, and other firms which receive saved funds in various ways and transfer them to borrowers. These institutions mobilize savings and make interest rates lower than if borrowers had to go directly to savers, often a virtually impossible task.

13. Some Cases of Interregional Interest Differences.—By combining the supply and demand factors described in the two

preceding sections it is possible to explain various regional differences in interest rates. For instance, the relatively low rates in Britain as compared with the United States are the joint result of a relatively advanced banking system, a high rate of saving, a somewhat slow rate of technological obsolescence, and a high degree of economic maturity. A contrast between eastern and western or southern United States reveals that in the Northeast there are more people with large incomes per thousand inhabitants, and therefore more saving. Banking institutions are more highly developed. There are relatively more business firms and lower rates of profit. And so on. A similar comparison could be made between various frontier regions and the areas from which their settlers came.

The interregional differences in interest rates which result from demand and supply differences persist in spite of the mobility of funds. Credit can be transferred from one place to another by mail or telegram with great ease and speed. People wishing to lend their own credit, or that of others which they have saved, want to get the highest possible interest reward. This incentive, combined with the high mobility of funds, prevents interregional differences in interest from being as great as interregional differences in wages. But it does not bring all interest rates to the same level. Suppliers are influenced by subjective estimates of risk. Loans to remote and less well-known borrowers are considered more risky than loans to local borrowers of equal repute. Investigation of distant borrowers takes more time and effort. Collection is more difficult and expensive if default occurs. Interest rate differences measure the opinion of marginal suppliers concerning differences in risk and trouble.

14. Summary.—An examination of the differences which exist among interest rates reveals many of the basic forces in this price field. There is nothing automatic about interest, nothing mechanical which determines what interest rates must be. Like other prices they express the results of human actions based upon human impulses and more or less rational human choice. These motives and these actions do not occur in a vacuum. The demand for funds and the supply of funds both spring from group life. At any given time or place they express the culture pattern of the group, with deviations because of human differences. Borrowing and lending occur also against a background of social institutions such as banks, mortgage companies, deficit financing, corporate securities, organized exchanges, laws, and customs. These influence, limit, and guide but do not completely determine the market activities of borrowers and

lenders. The most important of these relationships are summarized in the following outline:

I. Causes of interest rate differences at any given time and place:
   A. The terms of loan contracts are different as to time, collateral, repayment, etc.
   B. Lenders (suppliers of funds) are different.
      1. They have different ideas regarding the risks of lending.
      2. They have different (objective) costs of lending.
      3. Competition is more keen among some groups of lenders than among others.
   C. Borrowers (demanders of funds) are different.
      1. They have different use-opportunities and urgencies.
      2. They have different abilities to repay.
   D. Funds are not completely mobile on the supply side nor perfectly homogeneous as viewed from the demand side.

II. Yet the various loan markets are interdependent:
   A. Some lenders do shift funds from one type of loan to another. (Some of them also shift back and forth from the lending to the proprietorship position, from bonds to stocks, etc.)
   B. Some borrowers can choose between one type of lender and another. (And sometimes they can obtain funds in other ways than by borrowing, e.g., equity capital, savings, etc.)

III. Causes of interest rate differences at different times:
   A. Changes in profit prospects (demand side).
   B. Variations in concepts of lending risk (supply side).
   C. Changes in government policies (demand or supply action).
   D. The interdependence of interest rates which may make one rate change when another changes.

IV. Causes of interest rate differences at different places, but at the same time:
   A. There are many different markets where borrowers and lenders meet.
   B. There is imperfect mobility of funds from one market to another.
   C. There are differences in the demand and supply situations in the markets where interest rates are determined.
      1. Countries, for instance, differ in banking systems, types of industry, amounts of capital goods accumulated, rates of saving, attitudes toward lending, government fiscal policies, etc.
      2. Regions within a given country often have similar though less extreme differences.

Chapter 25

SOME PROBLEMS OF INTEREST THEORY

1. The Theoretical Approach.—In this chapter interest problems are analyzed on a higher level of abstraction than in the two preceding ones. Those chapters contain much detail about specific borrowing and lending situations in modern "capitalistic" society. The present approach attempts to distill the essence of these institutions. It also uses models of hypothetical economics. This method limits the number of variables so much as to impress some observers as being unrealistic. But it need not be. It is really an application of the *ceteris paribus* method of analysis. The "realistic" variables which are excluded by omission are really just impounded while attention is centered upon the variations in others. Unrealism does not enter unless the partial picture is taken to be a complete one. Generalizations and partializations should not be confused with absolutes.

Some of the questions to be answered in this abstract, "theoretical" way are the following:

1. Why cannot interest rates fall to zero?
2. What are the strong and the weak points of certain theories which explain various things about interest rates?
   (a) The cost of production theory?
   (b) The waiting and time preference theories?
   (c) The liquidity preference theory?
3. What are the effects of interest rates and their changes upon:
   (a) The supply of funds from various sources?
   (b) The demand for funds from various sources?
   (c) The supply of durable capital goods and the length of the production process?
4. Is there a normal or equilibrium rate of interest:
   (a) In the short run?
   (b) In the long run?

2. Minimum Interest Rates Must Pay for Marginal Risk and Trouble.—Interest rates cannot fall to zero except in rare cases be-

cause the marginal lender feels the presence of risk, or trouble, or both. All loans involve time. Between the date of lending and that of repayment, many things may happen. The borrower may suffer unforeseen losses which affect his ability to repay. There is some danger of loss in nearly all cases, but by the requirement of ample collateral, the risk of financial loss from borrower default may be reduced virtually to zero. Another risk cannot be overcome so easily, if at all. It is the risk that the going rate of interest may rise during the life of a fixed interest loan. This depresses the market value of bonds and commercial paper. Hence there is always the risk of loss if the lender's financial needs force him to sell his securities before maturity.

The elimination of "trouble" is also difficult, if not impossible. At the beginning of each loan an agreement must be made regarding the terms. Funds must then be transferred. At the end of the time period they must be collected. These operations take time and effort. The owner of the funds may delegate this work to others and pay them for their services. Or he may perform this work himself. There is trouble of this type whether the loan is for a short or for a long time period. Risk also constitutes a sort of subjective "trouble." Even when experience proves that collateral has been ample, some uncertainty and hesitation may exist in the mind of the lender.

The marginal lender is influenced by his feeling of risk and trouble in lending even though some intramarginal lenders may not be. The supply of funds for lending comes from many different sources. Some lenders may be willing to lend for nothing, or even to pay for the privilege. To others the risk and trouble appear negligible. For instance, governments may make their credit available at zero interest rates, as in certain foreign loans. Lend-lease "loans" during the recent war perhaps were made with the expectation that they would not be repaid fully and therefore reveal negative interest rates. But these cases are exceptional. Most lenders would rather not lend if they cannot get interest or its equivalent. The demand for funds is so great and the various loan markets are so interdependent that most borrowers must solicit funds from these "risk and trouble" lenders. The acquisitive spirit of our society then drives more willing lenders to ask as much interest as the marginal, less willing lenders.

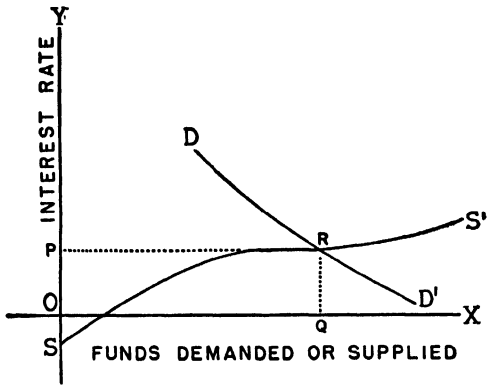
3. The "Cost of Production" Theory of Interest.—The foregoing argument may be called a "cost of production" theory of interest. Like the cost of production theory of commodity prices, it

is useful but not complete. It is primarily a supply-side approach and ignores institutional forces, bargaining strength, and other possible determinants of interest rates. But it suggests two important truths.

First, the creation of funds involves a certain cost of production. This cost may be objective, or subjective, or both. It limits the supply of funds offered at any given interest rate (cf. Sections 7–12 on the interest-elasticity of supply of funds). This makes the total supply cumulative upwards, like other upward-sloping supply curves. The first part of this curve (assuming that a composite supply curve is possible) may be below the *OX* axis. But this is of only academic

FIGURE 69

A Hypothetical Composite Supply Curve for Funds



interest. The important thing is that the total demand for funds at zero interest far exceeds the supply offered at zero interest.[1] Therefore other funds are sought, funds that do have a positive cost of production. The cost of the last supply increment needed to balance demand may be called the marginal cost of funds. The interest rate cannot be less than this figure. The adjoining diagram portrays this hypothetical, composite supply curve for funds. Its shape to the left and right of *R* is highly conjectural, but *R* is surely above *OX*. That is the point of this section about the "cost of production" theory of interest.

Second, the height of each of the many distinct supply curves for funds is also a function of the cost of producing those funds. That is one major reason why interest rates differ on different types of loans made at the same time. Chapter 24 presented this argument

[1] There may be a few exceptions such as intergovernmental loans.

from an institutional approach. Installment loans cost more to make and collect than do others, some loans are better protected by collateral than others, etc. These may be called objective cost factors. Subjective costs will be discussed below in the sections on the waiting and liquidity preference theories of interest.

4. Is There a Minimum Rate of "Pure Interest"?—In their effort to make interest something quite distinct from other prices, some writers have tried to describe "pure interest." Some have called it payment for waiting; others emphasize the loss of liquidity. The foregoing argument suggests that perhaps "payment for trouble" or "cost of production" may be better approaches, but all are supply-side viewpoints and therefore incomplete. However, the "cost of production" approach has the advantage that it is applicable to more situations than the others. For instance, a loan payable on demand involves no involuntary waiting. If it is also secured by ample collateral, the risk element dwindles virtually to zero. These two provisions of collateral and permission to demand payment at any time reduce the usefulness of the liquidity preference theory in explaining "pure interest." But "trouble" always remains as a "cost of production."

Even so, it is difficult to find any *minimum* rate of "pure interest." Rather there are *different amounts* of "trouble" or different "costs of production." And different amounts of objective or subjective risk may be added.

Higher interest rates pay for more trouble or for more risk than do lower interest rates. For instance, a comparison may be made between the yield on long-term bonds of various types. The going rate of interest on "perfectly safe" government bonds is usually lower than that on industrial or utility bonds which are considered more risky. Another comparison may be made between long- and short-term securities issued by the same borrower. Government bonds of short maturity usually yield less than those of longer maturity. This seems to reflect lenders' ideas of differences in risk. Here the subjective feeling of risk is probably associated more closely with lenders' ideas of their own possible need to liquidate at a loss before maturity than with any belief that the government might possibly default. The influence of risk differences is found also in the presence of higher yields on trade acceptances than upon bank acceptances of equal maturity, on United States government bonds as compared with those issued by Peru, etc. Trouble differences have already been shown to reside chiefly in loan collections. There are also differences

in the work which precedes lending as the lender tries to analyze the borrower's ability to pay.

5. The Waiting and Time Preference Theories of "Pure Interest."—Sometimes the supply side of the market in which interest rates are determined is interpreted in terms of the painfulness of waiting for future consumption. Those who supply funds are pictured as people who would not save and lend if they were not rewarded for the postponement of consumption which these acts make necessary. Saver-lenders are described as having positive time preference, at least at the margin of their saving and lending. That is, they prefer to consume now rather than to possess the prospect of consumption in the future when the loan is repaid. Under this analysis of human motivation, interest is said to be paid for the service of waiting, or of abstention. The height of the interest rate is supposed to measure the marginal degree of time preference. Interest rates could never fall to zero because people would rather consume than lend for nothing.

There is obviously a grain of truth in this approach, but it tries to explain too much. The time preference concept is helpful in explaining why some people save, but its chief usefulness lies in explaining why they save as much as they do. People save so long as they have positive time preference less than the interest reward from lending.[2] The limit to their saving is reached when the disutility of saving rises to equality with the utility of the interest which borrowers will pay. A better version of the argument would state that the initial satisfactions from saving are positive but smaller than the initial satisfactions from consumption. Both reveal a scale of diminishing marginal utility. If income is large enough, in most cases the marginal satisfactions from consumption will diminish to the place where they are less than the initial satisfactions from saving and some saving will occur. The equilibrium role of equality of satisfactions at the margin begins to apply to savings as one use of funds just as it does to categories of consumption which represent rival uses of total consumption.[3]

6. Weaknesses of the Time Preference Theory of Interest Rates.—Time preference is not a complete explanation of why people save. It assumes conscious choice, whereas many people save

[2] Zero and negative time preference will also induce saving. They may be considered extreme cases of the deviation of positive time preference below a positive interest rate.

[3] This is an application of the principle of maximizing total utility by equalizing satisfaction at the margin in each category of consumption.

because of habit, the compulsion of a contract, or impulse. Much saving by ultimate consumers in the modern world is done after they have deliberately or impulsively purchased something on the installment plan. There are also other bases of choice in addition to time-of-consumption comparisons. Those with very large incomes, for instance, may save merely because it requires less effort to do that than to find ways of spending their funds.

The theory that interest is payment for waiting also partly explains the fact that interest rates on long-term loans are higher than on short-term loans. As a matter of logic, however, this theory would merely require that the interest *amount* be greater for loans of longer duration, not the interest *rate*. There might be a way out by arguing that the irksomeness of prospective waiting increased progressively as the time of repayment was pushed further off, but that seems like stretching the time preference approach too much. After all, many people never expect to consume what they have saved. If they demand payment for abstention, the amount of that payment per year of abstaining should diminish, not increase with the number of years. The initial act is surely more painful, if painful at all, than the memory of that act as time goes by.

The time preference approach does not explain at all the differences in interest rates between different classes of loans of the same duration. Nor does it shed light on the differences in rates between places unless one were to argue that the culture patterns of time preference differed. The role of demand is also slighted, even though the valid argument is presented that borrowers are people with high positive time preference.

Another serious weakness lies in the fact that it does not explain divergences between the rate of saving and the rate of lending. Saving-lending is usually considered a simultaneous process although, by a twist in the traditional line of reasoning, time preference is sometimes described as preference for present liquidity rather than for present consumption. This approach tends to confuse the liquidity preference argument to be discussed in the next section with the risk and trouble of converting illiquid notes.

The major difficulty with the time preference approach is that it elevates a minor part of the picture into first place. The lending which takes place out of saved funds is secondary and inferior. The dominant forces in the loan market are those which govern the creators, not the savers of credit. Neither banks nor governments have time preference or subjective costs of waiting. Business firms also are motivated in their lending chiefly by objective costs.

The correct theory of the cause of *positive* interest rates in modern society must be sought in the trouble and risk approach. It has the added advantage of being clearly applicable to all types of situations in which funds are supplied, whether from private individuals who save, from institutional savers, or from those who lend their own credit, such as commercial banks.

7. Liquidity Preference Theories of Supply and Demand for Funds.—The liquidity preference approach to the supply side also has some truth in it. It is very useful in explaining *changes* in interest rates, but it does not help much in answering the question at hand regarding the existence or nature of "pure interest."

The liquidity preference approach separates the acts of saving and lending. Saving occurs for contingencies, for convenience, or for prospective interest reward, as explained in the preceding chapter. The saved funds are usually held in the form of government credit (paper currency) or bank credit (demand deposits), which are widely acceptable and therefore can be spent easily. When loaned, the savings are transformed into a note which is not so easily given in exchange for goods or in payment of debts. The saver, by the act of lending, sacrifices liquidity.

From this approach, interest becomes the payment which borrowers are willing to make to obtain liquidity and which lenders exact as the price of giving up liquidity. It is logical, therefore, that the more liquid the borrower's note, the less the lender will demand, and vice versa. The lowest rates are those paid on short-term notes which mature quickly and on notes issued by well-known borrowers of good reputation.

Examined more carefully, the liquidity of an asset may be measured in terms of the cost or loss involved in a prompt exchange of it for the most widely acceptable form of credit, namely, currency. In this case that means the difference between the amount loaned and the amount that can be realized by sale or discount of the note given by the borrower. If there has been no change in the prevailing interest rates, the cost of conversion will be merely that of the labor or trouble involved in making the exchange. A rise in interest rates, as explained earlier, will bring some additional loss of principal. When stated in these terms, the liquidity preference explanation of positive interest rates takes on many aspects of the trouble and risk theory expounded on pages 415–416.

8. Weaknesses of the Liquidity Preference Explanation of Interest.—The chief shortcoming of the liquidity preference approach

is that it does not describe the role of commercial banks and other important lending institutions. Commercial banks are in the business of lending funds. *They usually want to become "illiquid"* because that is the way they make their profit. In times of business uncertainty, it is true that banks may shift their investments so as to become *more* liquid, but to become completely so would mean that they had gone out of the lending business entirely. There again, the trouble and risk theory explains the different interest rates which banks charge much better than the liquidity preference approach.[4]

Certain other lenders such as insurance companies, savings banks, and mortgage loan companies likewise operate under motives which do not fit the liquidity preference pattern. For many of them there are no foreseeable contingencies which might make them want to liquidate their loans before maturity. There is therefore no risk of loss through sale of securities whose value has depreciated because of increased interest rates. The management need consider only the risk of borrower default and the costs of making and collecting. Some wealthy private individual lenders are in the same category. Liquidity preference does not exist for them. They want their funds continually invested and earning something. They are glad to get what they can, no matter how small it may be. The same is also true of many lenders of moderate means. Even speculators who want their assets liquid on certain occasions pending more advantageous investing opportunities do not require an interest reward to induce them to enter the market. They seek to make their income chiefly by buying low and selling high, or by selling short and buying back at lower prices. Their motivation is clearly different from that of potential lenders.

It is difficult to contend, therefore, that the reason for interest rates being greater than zero is the positive liquidity preference of those with funds to lend. Trouble and risk offer a superior explanation and also are better adapted to explaining the differences in interest rates which exist at any given time.

The chief merit of the liquidity preference approach would seem to be in its separation of the motives of saving and lending. Its major usefulness in dealing with current problems lies in the matter of *changes* in interest rates. In economic dynamics, much of the emphasis shifts from the supply to the demand side. Borrowers usually are persons who want funds to spend. They want temporary

[4] This is not to deny that in the management of bank portfolios, the question of liquidity plays an important part. For instance, commercial loans eligible for rediscount at federal reserve banks are preferred to those which are ineligible, even though the trouble and the default risks are the same.

liquidity so that they can buy the particular objects of their desire. This is perhaps a form of liquidity preference. However described, its changes are a major cause of changes in interest rates.

9. The Interest-Elasticity of Supply of Funds: General Aspects.—The effects of interest rate changes upon the supply of funds are best studied by separating the sources. These include (1) commercial banks, (2) institutional savers, and (3) private savers, both individual and business. For each category there should be a further subdivision into short-run and long-run effects. Nor can the different classes of borrowers be overlooked in this supply-side problem. Very often what appears to be an increase or a decrease in the funds offered is merely a shift from one type of offering to another with no change in the total. Therefore, one constantly must remember that there are very many supply curves of funds and very many elasticity functions. The interdependence that exists among them merely makes the problem more complex.

10. The Interest-Elasticity of Supply of Funds from Commercial Banks.—Commercial banks lend in various ways as described in Chapter 23. When they lend at a fixed rate of interest, the elasticity of supply is infinite, at least for a certain range of lending. The banks supply whatever funds are demanded. Interest rates are set by the lending banks individually or in concert with others of the community or region. The question of supply-elasticity is not nearly so important as why banks change these asking rates of interest, and that has been answered above. (See Chapter 23, Sections 11 and 22.)

At times these infinitely elastic supply curves are deceptive. Banks put other restraints upon lending, either qualitative or quantitative. If there is a quantity limit, then the curve turns upward abruptly at that point and drops from infinite to zero elasticity. Where the restraint is qualitative, a two-dimensional supply curve would be utterly unrealistic.

Banks also lend by purchasing securities on the open market. This type of lending is hardly a function of the interest yield, and the total does not rise and fall appreciably because of changes in that yield. Therefore, the interest-elasticity for the total volume of this lending is virtually zero. Shifting occurs, however, between particular types of loans, as when one short-term rate becomes more attractive than another.

The same sort of observation may be made for bank lending to large borrowers where interest rate competition and bargaining fre-

quently appear. There is little evidence that banks lend more at any given time because these rates have risen than they would have loaned if the rates had remained constant or fallen. It is true that the amount loaned has a positive correlation with the interest rate, but only because they are both a function of general business activity and optimism, not because interest rates are a determinant of bankers' decisions to expand or to contract the quantity of funds thus loaned.

Although what has been said about commercial bank lending thus far applies particularly to individual banks, it also holds for the system as a whole. Reductions in legal reserve ratios, increases in central bank gold reserves, relaxation of rediscount rules, and similar institutional developments have all accompanied the secular uptrend of the volume of bank lending.[5] Falling interest rates have certainly not diminished the total volume of funds offered by the banking systems of such countries as the United States and the United Kingdom.

11. The Interest-Elasticity of Funds Supplied by Other Lending Institutions.—The quantity of funds offered in various loan markets by savings institutions, life insurance companies, etc., is primarily a function of the quantity of savings. It is secondarily a function of the liquidity preference of the managers of the institutions. The question at issue is, therefore, "What effect, if any, does the rate of interest have upon either the volume of savings entering such institutions or the liquidity preference decisions of executives?" The savings question will be considered more thoroughly in a later section dealing with the savings of individuals. At this point the only comment that need be made is that some of the loanable funds of these lending institutions are derived from their reinvested profits, although most of such funds go into buildings and equipment. Insofar as higher interest rates mean larger earnings, there might be a long-run connection here, but it is of little significance at best.

The desire of managers of these institutions is to keep as fully loaned as possible with due regard to safety. Therefore they feel impelled to keep their liquid funds at a minimum and will lend approximately the same percentage of available funds whether the interest rate is high or low. A *change* in interest rates, however, as distinct from the *level* of those rates, may provoke abnormal entry into or exit from the loan market. When managers become apprehensive regarding safety, they may reduce their current rate of lending. A new feeling of optimism will produce an opposite result. As has been pointed out frequently, however, such changes in attitude

[5] The major determinant has been the rising level of general business activity.

are usually more the result of the general trend of business than specific changes in the interest rates that prevail. Furthermore, many of these institutions lend at rates of interest which they themselves determine as a matter of business policy. They do not take the initiative in making loans. They merely announce that they are willing to lend to those who want to borrow. Passive lenders of this type have infinite elasticity of supply in the usual sense. From another viewpoint they might be said to have zero elasticity of supply, since a general rise in interest rates does not cause them to offer more funds than before.

In the long run if interest rates decline, many institutional lenders will reduce the amount they pay the persons from whom they get their funds. This is particularly true of savings banks, insurance companies, and building and loan associations. If interest rates rise, the long-run effect may be reversed. Depositors and policy holders will be paid more in one way or another. If there is insufficient competition among such financial institutions, much of the gain from increased interest rates will become profits. In that category it may be reinvested by the firms or distributed as dividends. This brings us again to the question of the reactions of individual income receivers and holders of funds.

12. The Interest-Elasticity of Funds Loaned by Private Savers. —In discussing the problem of individual reaction to interest rate levels and changes, it is particularly important to keep the short-run and the long-run pictures separate. In the short run the important question is that of liquidity preference, which governs the percentage loaned out of savings. In the long run the amount of current savings is more significant. There may be also secular trends in habits of thrift akin to changes in the cultural pattern of liquidity preference.

Without repeating previous arguments too much, it may be stated that the level of interest rates has little to do with the rate of lending so long as interest is high enough to cover the expected risk of loss. If individuals believe that they can get their funds back at any time without loss, they are willing to lend at very low rates of interest. Savings banks experienced no decline in total deposits when their interest rates fell to 1 per cent as compared with 3 to 4 per cent twenty years earlier. A similar observation might be made about building and loan depositors, particularly since the federal government established its deposit insurance program. This guarantee, however, is not for immediate payment of the full amount loaned, and the memory of large depression losses in some instances still

makes it necessary for such companies to pay higher rates than savings banks, whose record has been better.

So far as bonds are concerned, the level of interest rates has been brought much lower in recent years than was previously thought possible. Yet the rate of investing in these securities does not seem to have been adversely affected. One reason is the fact that the possible yield on alternative forms of investment has declined, too. People seem to prefer some interest to no interest. A very large percentage of saved funds is valued by their owners for potential income yield, not for use in foreseeable future contingencies. Therefore securities are purchased with an expectation of holding until maturity. This eliminates one of the major risks involved. Another major risk, seen by economists, that of a change in the general price level, is not usually considered a risk by investors and therefore does not influence their decisions.

Some lenders do at times become alarmed regarding the third big risk, that of a borrower default, but this is usually a result of major business declines or security market panics. People sell securities when they lose confidence. This raises interest rates. The adverse shift in estimates of risk constricts the rate of lending. This is the true causal sequence. It is the reverse of the usual elasticity approach. It is not a reaction of potential lenders to a drop in the rate of interest below their margins of liquidity preference. In formal terminology, the change in supply is a restriction in the schedule sense, not in the market sense.

13. The Interest-Elasticity of the Amount of Current Saving. —This topic must be divided into two parts. The first assumes that incomes are constant. The question becomes one of the influence of interest changes upon the percentage saved out of current income. This is the problem of incentives to saving in the sense of nonspending of money income. Does a rising interest rate, for instance, cause most people to want to save more out of their (unchanged) incomes? If so, then there is a positive interest-elasticity of saving. Then if lending rises as savings rise, there exists by this indirect route a positive interest-elasticity of lending. The supply curve of loanable funds available from individuals slopes upward to the right. The truth of this proposition will be appraised in the remainder of this section after a digression on the second division of the general topic.

Another part of the problem assumes that incomes may change. Do interest changes influence the level of money incomes? Authorities differ as to the correct answer. Some business cycle analysts,

for instance, contend that governments can stimulate business recovery from a depression by reducing interest rates. Others argue that of itself that action is not enough. Conditions must be just right for such manipulation to turn the tide. Or other action to stimulate business borrowing must take place at the same time.[6] At the opposite extreme stand those who say either that these other activities would do the trick without interest rate changes, or even that the upturn would come more rapidly by itself if only the government would do nothing at all. We cannot explore this controversy further at this point.

There is much less dispute about the high positive correlation between the level of people's incomes and the amount they save. When people get more pay, they have more income out of which to save. Carefully collected statistics also show that the *percentage* saved rises as incomes rise. For these two reasons, therefore, it is important to know whether interest rate changes influence those changes in the business cycle which have such a strong bearing upon average incomes.

Returning now to the first question regarding the possible direct connection between interest rates and saving, we must examine four motives for saving: (1) the expectation of interest reward, (2) the expectation of producing other future income, (3) the anticipation of contingencies, and (4) the reduction of effort.

14. Saving Motivated by a Desire to Increase Income or Wealth.—Saving may be stimulated by the hope of financial gain. Most people like to get interest income if they can. But saving does not seem to be stimulated very much by *increases* in interest rates, nor discouraged by decreases. This is probably due to the fact that other motives dominate saving and the hope of interest reward is a very minor incentive. It is like eating because you are hungry, but being thankful at the same time that the food is tasty.

At the extreme high and low ends of possible interest rate variations, the interest reward motive might be important. If the rate jumped from 6 to 60 per cent per year, some would surely save who otherwise would consume. Similarly, if the rate which saver-lenders could get dropped from 4 per cent to ¼ of 1 per cent without any change in the risk involved, some savers would undoubtedly choose to spend instead. One must not, however, infer from these extreme and hypothetical cases that the supply is similarly elastic for changes between 6 and 7 or between 4 and 3 per cent. Within the customary

[6] These other activities may include deficit financing of armaments or public works, devaluation of the currency, tariff changes, etc.

range, the small elasticity of individual supply curves of loanable funds that does exist is probably due more to liquidity motives than to the motive of interest gain through saving-lending.

One should note also that variations in the amount of funds saved out of current income do not change the total amount of funds in existence. Changes in total funds result from changes in total credit granted or created for production or consumption. Changes in saving merely influence the amount that individual savers (persons or firms) may lend. Another way of looking at it is to say that the velocity of circulation of funds is changed, not the total amount. It takes longer for funds to go through the received-saved-loaned-spent cycle than merely to be received as income and then spent.

Some people save to get a profit-type income by investing their savings in stock shares or directly in production goods. This desire for financial gain through the use of savings may also be seen in the activities of speculators and gamblers. The reward for the successful use of funds in these ways is usually called profit. But profit yields and interest yields generally rise and fall together. Therefore there is some positive correlation between interest rate changes and saving for profit-type income, even though the causal connection is weak.

Savings for business use do not enter the loan market. Their effect upon the rate of interest is indirect. Their presence diminishes the amount which businessmen need borrow. Reducing the demand for funds has the same restraining effect upon interest rates as increasing the supply. But our major attention here is upon interest rates as causes, not effects. If profit rates are raised by boom times, and interest rates too, then they stimulate saving which becomes available for use or lending, thus restraining the rise in interest and profit rates. The situation is somewhat analogous to that of commodity prices during a boom. Rising demand raises prices. These stimulate an expansion in supply. The increased supply does not usually offset all of the increase in demand, but only part of it. The price increase is slowed, not reversed, at least not at first.

15. Saving to Meet Contingencies or to Reduce Effort.—Most small savers think of themselves as saving "for a rainy day," for a vacation, for their old age, or for some similar contingency. This type of saving is little influenced by changes in the rate of interest. An infrequent exception may occur when an individual calculates very carefully how much he must save per year in order to accumulate a specified total on a certain date. Cases of this type have become in-

creasingly rare in this day of installment buying and salary loans. Other contingency savers have no definite amount as a goal or do not compute interest accumulations carefully. For them saving becomes a habit of putting away a fixed amount each payday, or of living as cheaply as possible and saving the rest.

For a few people with large incomes, saving may be easier than spending. When consumption patterns are fixed by habit and current income mounts, people sometimes save because it requires less effort than deciding what to buy. Lending is also more arduous than mere accumulation. Giving away surplus funds seems contrary to the habits of a lifetime and may involve painful choice. Sometimes such a person hires an investment adviser and lets him do the work of discovering investment opportunities. Many trust funds operate in similar fashion. Their budgeted outlay proves to be less than their income. The surplus is reinvested or held as a bank balance rather than spent.

Business firms which accumulate large cash reserves usually have a long-range plan of expansion or a loss prospect in mind. But occasionally stockholders get the idea that managerial habit is stronger than judgment, stronger than the wails of people starved for dividends.

In all these cases a rise in interest yields has no direct influence upon the amount saved out of a given income level. It may, however, raise that income or occur at the same time that there is a rise in the general level of incomes.

16. Collective Saving by Governments of Planned Economies. —The personal motive of saving for future income has its counterpart in the motives of leaders of socialist or fascist states. When a private individual saves, he diminishes his present consumption because of some greater pleasure. This higher satisfaction may be the thought of larger income, or wealth, or power, in the future. When government planners decide to forego present consumption for future gain, they often have similar objectives in mind.

The economic planner must make two important estimates. First, he must make a technician's decision regarding the amount of future output which will result from given present inputs diverted from present consumption. Second, he must make a social welfare decision regarding the desirability of this output in relation to other possible uses of the available inputs. These alternative uses include production of other things in the future, production of things in the less distant future, aesthetic enjoyments, etc.

A social welfare decision requires a scale of values. This is likely to be based upon some philosophy of welfare for the group, or even for the planner. In a democratic planned economy a vote may be taken in which the majority decides about each major project. Even in a capitalistic economy, voters are frequently called upon to approve or disapprove important public works like schools, bridges, dams, sewage systems, etc. The scale of values involved in such votes is not often clearly formulated. Affirmative votes may be cast by people who think they do not pay taxes, whose tax payments are expected to be less than the benefits, whose ideas are very hazy about the burden of higher prices, or who are swayed by emotional appeals and half-truths. In an authoritarian economy where saving-investment decisions are made by a few elected or self-appointed leaders, the scale of values is usually better understood. But it may change as conditions change, like the Communist "party line." Or top priority objectives may appear to outsiders to be evil and menacing, like the Nazi's belligerent slogan, "Guns, not butter."

Given a scale of values, good or bad, clear or vague, a choice can be made. The general principle that is applied can be stated formally as attempting to equalize the marginal utilities of expected goods which differ in time dimension as well as in cost, kind, quantity, and quality. The solution of such a problem in an exact way is impossible. The best that one can expect is a rough approximation.

17. Use of Funds to Increase Consumption in Distant Future vs. in Near Future.—With these cautions in mind a numerical illustration may be offered. If a planner today appraises 100 bushels of additional wheat at the end of one year at, say, $100 on a relative scale, and at the same time values an equal amount of extra wheat ten years from now at $80, he will choose the former. But the effort required to yield this year's 100-bushel increment may produce very much more wheat ten years hence if used to develop an irrigation project in the meantime. Then the planner should compare a possible 150-bushel return after waiting ten years compared to a 100-bushel return after waiting only one growing season. If the 150 distant dated bushels appear to be worth $110 in the planner's scale of values, he should choose them in preference to the 100 bushels worth only $100.[7]

This means diverting labor and materials from one form of employment to another. Granted the assumptions of the preceding

[7] The comparison should really be between two *streams* of inputs and outputs extending indefinitely into the future, but using only two dated amounts simplifies the presentation and does not distort the argument. Cf. Chapter 10, Section 5.

paragraph, this will be done in time choices only if there are *both* positive time preference and technical productivity. The diverted factors are "saved" in the sense that their product is not currently used. They are employed to make presently nonconsumable things which are expected to yield more consumable things in the long run than could be obtained by using the factors directly for the same time period. This is the so-called "roundabout method of production." It is described as being technically more efficient than "direct production," although the comparison in modern times is really between more and less roundaboutness, rather than between some and none.

There is the further possibility that present use may be sacrificed for some future use which is not commensurable with that given up. If the illustration is in terms of bushels of wheat, now and in the future, a comparison can be made which reveals a measurable gain (or loss). But if present skyscrapers are sacrificed to secure the building of a bridge which ultimately speeds the flow of automobiles across a bay, there is no common denominator. Soviet planners sacrificed food, clothing, and shelter during the First Five Year Plan to secure blast furnaces, copper mines, hydroelectric projects, and the like. Many of the latter are more easily connected with a self-sufficiency program or with armaments than with an increased per capita output and consumption of basic necessities, especially when trade opportunities are considered. It seems more logical to interpret such action as a choice of the more attractive of two alternative present satisfactions than as a manifestation of the "superior productivity of roundabout production." [8]

18. The Interest-Elasticity of Demand for Funds in the Short Run When Rates Fall.—The chief reason for examining the interest-elasticity of demand for funds in the short run is the power of the government to influence that rate. In modern banking systems a virtually unlimited amount of funds can be loaned without prior saving. The central bank of a nation cannot compel people to bor-

[8] It is conceivable, however, that future output might be less than at present, as with a natural resource subject to depletion. If this product seems highly desirable for the future, an economic planner might use current factors to create the means for insuring some future supply rather than none, or more rather than less, even though the "more" is less than could be obtained for an equivalent expenditure to secure current output. Such is the reasoning behind many conservation programs and certain aspects of "national defense" planning. When marginal increments of future oil, coal, steel, or other goods are preferred to present marginal increments of equal size, this must be called *negative* time preference. The common assumption of positive time preference should not blind us to the presence of these important exceptions.

row, but it can bid up the price of outstanding securities by purchasing them on the market. This decreases the yield, as explained above, and brings lower interest rates for those who might want to borrow. Or the process may be reversed. The effect of both a rise and a fall in interest rates should be considered.

When fiscal policies bring a fall in interest rates, what effect does that have upon the volume of borrowing? The answer should be divided into several parts according to the potential borrowers involved and the time allowed. A reduction in interest rates on government securities and on commercial paper will not cause a proportionate decline in interest charges on consumption loans. Even if there is some reduction, consumer borrowing is not likely to be affected appreciably. The volume of such loans is dependent almost entirely upon other things, such as the current income of consumers, repayment terms, etc., described in Chapter 23, Section 17. This conclusion holds for both the short run and the long run.

Government borrowing is likewise only incidentally a function of interest rates. Therefore, national government demand for funds will have very low interest-elasticity. If the low-interest policy is born of a war situation when the government wants to borrow large amounts, it is true that the total interest bill will be less. But one cannot argue with conviction that if interest charges had not been reduced the government would have borrowed less and taxed more. Surely it would not have spent less. In antidepression borrowing and spending there is less unity of purpose. Opponents do point with alarm at the interest cost of borrowing, but their main concern seems to be (or to have been) with the size of the public debt. There are probably very few legislators whose votes swing one way or the other because of any change in the interest rate on government bonds. Arguments which hold for the short run are also probably true in the long run. There seems to be no historical evidence that governments have become more profligate during secular downtrends in interest rates.

The crucial question remains, "What effect do reduced interest rates have upon business borrowing?" Here the fundamental criterion is their effect upon prospective profits. In a dynamic society such as ours, profit expectations depend upon a host of cost and revenue considerations of which interest is only one. The phase of the business cycle is very important. When the future trend of sales points downward, businessmen are unlikely to increase their short-term borrowing even if interest rates should drop to zero. At such times entrepreneurs prefer to pay off their prior loans and to reduce their

indebtedness, not to increase it. Long-term borrowing will also decline in most cases, because a business recession is a time in which businessmen are usually gloomy about the distant future, too. This decline in schedule demand is not offset by the fall in interest rates which generally occurs during such periods. During the upswing, expectations of future earnings rise so rapidly that they offset any deterrent effect of rising interest rates. The record of net new issues of securities tends to follow the ups and downs of the interest rate, not the reverse. It is probable that, in the short run, factors other than the interest rate have greater potency in determining the amount borrowed by business.

On the other hand, a fall in the rate of interest may prove the crucial determinant of the level of business borrowing at times when there is no definite trend in sales. These are the trough and the summit phases of the cycle. Here there is no general consensus regarding the trend of other variables, and therefore changes in the interest rate may yield more than their usual influence.

19. The Interest-Elasticity of Demand for Funds in the Short Run When Rates Rise.—The probable effect of a *rise* in interest rates remains to be studied. Consumer and government demand is quite inelastic in the face of such change for much the same reasons as given above. Business demand will also be inelastic in periods of decided upward or downward trends in earnings. Nevertheless, business borrowing seems to be more easily discouraged than encouraged. Its demand is less inelastic when interest rates rise than when they fall. A very sharp increase may spread panic, but a sharp decrease does not seem to create confidence.

The elasticity of borrower demand, like that in other demand schedules, is a function of the rate of change of the independent variable. (Cf. Chapter 4, Sections 3 and 4.) A large and abrupt change in the interest rate will obviously have more effect on the quantity demanded than a smaller change or the same change achieved more gradually. This is important in two respects. Abrupt increases in interest yield occur more frequently than abrupt decreases. That is, large, panicky declines of the securities markets occur more frequently than equally large, optimistic rises. In the second place, central bank selling of securities becomes contagious more quickly than central bank buying.

This leads to the important observation that the market activity of central banks is not always sufficient to cause a decisive rise or fall in interest rates. The total volume of buying and selling of securities

is so great that central bank participation is rarely more than a small fraction of the total. Usually it relies for its influence upon "tipping the scales" and setting a trend which will be copied by others until the desired effect is achieved. At other times the central bank may *force* certain commercial banks to offer securities for sale. This may occur when two conditions are present. First, commercial bank depositors buy the securities offered by the central bank. Second, these member banks do not possess excess reserves upon which they can draw. Therefore they must sell securities to be able to honor the buyers' checks tendered by the central bank. Furthermore, the central bank may raise reserve requirements at the same time, or quite independently. This may force member banks to sell securities to increase their reserves.

On the other hand, the speculative bearishness of private investors may start a downward trend that central bank activity is impotent to stop. Speculative booms are only slightly less difficult to control. Finally, one must remember that, regardless of how it starts, a marked trend in either direction will influence business chiefly through its effect upon forecasts of changes in sales, commodity prices, and other variables, *not* through the rate of interest.

20. Interest Rates in the Long Run.—If there is any such thing as a "long-run normal rate of interest," it will depend upon the position and the slope of the long-run demand and supply curves. The elasticity of demand will be a function of diminishing marginal productivity or utility. This includes, but is not limited to, the increase in the stock of producer's goods which is associated with lengthening the period of roundabout production. The elasticity of supply will be a function of motives for not saving, i.e., for not decreasing present consumption. These are usually grouped under the heading of positive time preference. One could also refer to the increasing marginal utility of funds resulting from increased subtractions from expendable balances.

The position, as distinguished from the slope, of the long-run demand curve for funds is chiefly a function of the level of business activity and of government deficit spending. Also important are the existing quantity of capital goods, their cost, and their physical productivity. The latter determinant is related to inventions and the state of the arts. Cultural factors which stimulate or restrict borrowing for enterprise or for consumption are very influential. Consumption, of course, includes government spending as well as individual.

The position of the long-run supply curve is dependent chiefly upon the number of financial institutions and their powers to create credit. Also of major importance are the level of income and its distribution among the people. Other forces include attitudes toward saving, laws which stimulate spending or saving, the incidence of taxation, the opportunities for installment buying, social security legislation, etc.

A stable equilibrium rate of interest in this long-run sense must do more than merely equate a long-run supply quantity of funds with a long-run demand quantity. It must also be neutral in its effect upon the positions of the two curves. That is, for the rate to remain stable it must not induce any shift in the demand or the supply curves in the cycle of events which follows the borrowing-lending which occurs at that rate.[9]

Complete equilibrium requires a balance among a whole host of variables. Labor and capital goods are sometimes substitutes for one another, and therefore the equilibrium interest rate must not disturb wage rates. Rent and profits are ways of getting property income which are substitutes for interest income from lending, and they, too, must be brought into the picture. The prices of commodities influence the revenue derived from using funds in business enterprise and must be considered by borrowers. When one looks far enough, he sees that to describe interest rate equilibrium in the hypothetical or long-run sense requires a survey of the entire economy and a statement of its many interrelationships. However, for most practical purposes of explanation, prediction, or control, partial equilibrium analysis or short-run determinants are enough.

21. Summary.—Interest theory is complicated by the large number of different sources of demand and supply, most of which are themselves interdependent. Different motives are involved and the dominance of certain motives changes from time to time. The short-run, day-to-day explanation of interest rates is very different from that for the long run. Analysis on any general plane must either shorten the list of independent variables or take refuge in a not very helpful statement that all prices, including interest rates, interact to determine a general equilibrium. Nevertheless, a review of the institutional material of Chapters 23–24 suggests certain dominant relationships which can be summarized profitably.

Pure interest, if it exists, is variously described as a function of

[9] Or the shifts must offset one another.

the following restraints upon supply, preferably the first, but having some connection with the others:

1. The minimum payment for the risk and trouble of the marginal lender on the shortest and safest of loans
2. The positive time preference of the marginal saver-lender
3. The liquidity preference of the marginal holder of funds

The interest-elasticity of *supply* of funds in the short run is chiefly a function of:

1. The way in which changes in interest rates influence expectations regarding future changes in:
   (a) The interest rates, especially relative to their effects upon possible speculative gains or losses in the security markets
   (b) The profit rates for investors in equities
   (c) The safety of principal against borrower defaults
2. The type of lender:
   (a) Institutional lenders: very little supply-elasticity, except that when loans are offered at a fixed rate, supply seems to have infinite elasticity at that rate
   (b) Individual lenders from funds saved out of current income: some elasticity, especially at very high and very low rates

The interest-elasticity of *demand* for funds in the short run is chiefly a function of:

1. The way in which changes in the interest rate influence expectations regarding future changes in:
   (a) The trend of security prices
   (b) The profit rates for business firms
2. The stage of the business cycle: very little elasticity of demand except at the top and the bottom, where no pronounced trends exist
3. The type of borrower: elasticity of demand is significant only in the case of business borrowing

In the long-run picture the level of interest rates becomes the major issue. Dynamic forces associated with changes in expectations can be eliminated. The positions of the demand and supply curves are as significant as their slopes. Major determinants of any "long-run normal" rate must include:

1. On the demand side:
   (a) The marginal productivity of additions to business resources, particularly durable capital goods. (This is a

function of (1) the total quantity of such goods in use and (2) the state of the arts.)

(b) The marginal utility of consumption spending by individuals and governments. (The decisions of the latter are best seen in a socialist economy.)

2. On the supply side:

(a) The financial institutions of the region and their policies

(b) The marginal propensity to consume, which is another way of describing the desire not to save out of current income

(c) The marginal liquidity preference for other than speculative purposes, particularly the desire to hold cash balances against contingencies

(d) The level and the distribution of individual incomes in relation to subsistence standards of living

Chapter 26

PROFITS: AN INSTITUTIONAL APPROACH

1. Statement of the Problem.—Like the first three distributive shares, profits may be analyzed from either an institutional or a theoretical approach. The institutional approach of the present chapter treats profits as an effect. It examines the causes of the relative and absolute magnitudes of profits and the causes of profit changes. The next chapter is concerned with theoretical problems. Most of them treat profits as a cause.

The specific questions to be answered in this chapter include the following:

1. What are the various meanings of the word *profits?*
   - (a) Which of them is used in this chapter?
   - (b) How do profits differ from wages, rent, or interest?
2. What problems are involved in determining:
   - (a) Profit amounts?
     - (1) Inventory valuation
     - (2) Depreciation
   - (b) Profit rates?
     - (1) Computation of profit amounts
     - (2) Choice of base
     - (3) Valuation of base
3. What are the causes of differences in profit rates:
   - (a) Among firms in the same industry?
   - (b) For one firm from one time period to another?
   - (c) Among industries?
   - (d) Among regions?

2. "Accounting Profits" Are the Chief Concern of This Chapter.—There are several possible ways to define profits. The approach which fits best into the pattern of this volume is to begin with a definition which adheres as closely as possible to the connotation accepted by the typical "man in the street." From this viewpoint profits are the net income of business units figured by some type of accounting procedure which subtracts outgo from income.[1] The

[1] A more technical description of the accounting process will be given below as one which estimates the net increase in the value of the assets of a firm during a given period of time.

business unit is best known as a "firm." Its characteristic activity is buying and selling in the hope of taking in more than is paid out. From these elementary concepts we may develop others of importance.

In the first place, firms are owned by one or more individuals (or other firms). Sometimes the owner and the firm are so closely connected that there is nothing to be gained by trying to make a distinction between them. Such is the case of the small firm operated as a sole proprietorship. At other times the owner is clearly different from the firm, as in the case of the corporation. Partnerships lie in between. Generally speaking, it is wise to distinguish the owner from the firm he owns. *The firm "makes" profits; the owner merely receives profits from the firm.* He often gets or takes less than the total profits made. Sometimes he receives more than the firm made during the period. This disparity between profits made by the firm and profits received by the owners is most clearly seen in the case of large corporations. The total dividends to stockholders in a given year are usually smaller than profits, but are sometimes larger.

3. Accounting "Profits" May Be Negative, i.e., May Be Losses. —A second important group of observations about profits deals with their residual nature. Profits are not "paid" by one person to another. There is no contract agreed upon in advance to pay a certain amount, as in the case of nearly all wage, rent, and interest payments. Profits are uncertain until *after* a firm has done something. They are calculated by a process of subtraction. They are found by accountants to exist (or not to exist as the case may be). No one pays profits to a firm. They are a residual form of income and differ markedly in that respect from the contractual incomes popularly known as wages, rent, and interest. The significance of this contrast will be developed in many of the sections which follow.

A third distinct feature of profits which separates this form of return from the other distributive shares is the fact that profits are sometimes negative. Cases of negative wages, interest, and rent conceivably may be found by very careful search, but they are very, very rare. Negative accounting profits occur in a substantial minority of cases. Their occurrence is so common that they have been given a name of their own, i.e., *losses.* The financial statement of firms which deals with their income and expenditure record over a period of time is called the "profit and loss" statement.

The possibility of such loss is implicit in the nature of business activity. The simplest accounting approach implies in effect that expenditures are made out of income. The total spent is subtracted

from the total income and a negative remainder may occur. In fact, however, buying usually precedes selling. Expenditures are made out of funds on hand and are subsequently recouped, if possible, by sales. Following the logic of the sequence of business activities, the subtraction might be reversed. Income might be subtracted from expenditures. Then the remainder, if positive, would indicate a loss. This inversion is useful in that it indicates how a firm may finance its losses. It shows how the dividing-up process known as distribution theory starts with something more than the total income of the period. The total to be distributed is that income plus the net worth of the firm. When the firm contracts to pay out more than the income it realizes during that period, it must draw upon its net worth. If the firm becomes obligated to pay more than its income plus its net worth, the firm is clearly bankrupt. It has undertaken to divide up more than it possesses.

In concluding those introductory remarks, it may be well to point out again the concept of profits which is chosen for the initial approach to the subject. Profits are defined as *accounting* profits. Whenever the term is used without a qualifying adjective, this is the meaning to be understood. In later sections we shall refer to "normal profits" in an effort to draw a parallel with the concept of "normal price." There will also be reference to "pure profits" in the discussion of deviations from "normal profits." It is very important not to confuse the concept of pure profits with that of accounting profits, and the adjectives should be used copiously. The former stresses a unique *type* of distributive income. The latter emphasizes a unique *way* of *receiving* income. Pure profits are supposed to be devoid of all wage, rent, or interest elements due to owners of business units. The concept of accounting profits avoids that issue by separating the firm from its owners and using a functional approach to explain what the firm and its accountants do to "make" profits. Various concepts of profits are useful and will be employed in the following analysis. No one concept should be singled out as "right" while all others are declared to be "wrong." All that is required of the economist is that he make his meaning clear and significant. He should define his terms so as to expedite that task.

4. Problems Involved in Determining the Amount of Accounting Profits.—Two problems of accounting theory are of special interest to economists. Both involve value estimates. The first is that of inventories. A firm usually enters an accounting period with some goods on hand and leaves that period with a different quantity.

Inventories of manufacturers may include raw materials, goods in process, and finished products. Some of these goods may be held for several months or longer. When they are held so long that prices change in the meantime, a valuation problem emerges. For instance, if raw materials are purchased at one price and are not used until after a later date on which the firm must calculate its profit or loss for the period, the accountant must decide whether to carry the inventory at original cost, at replacement cost, or at the price the goods would bring if sold. He is likely to be conservative and choose the lowest of the three figures. If he chooses one of the others, he will be able to show a larger profit for the period.

The second problem deals with estimates of depreciation. This is a valuation process at the outset, because a value must be placed upon the asset being depreciated. The usual procedure is to use the original cost of the good, but sometimes replacement or reproduction cost seems more appropriate. Depreciation, however, involves a more crucial decision than this one. It raises the question of how rapidly to depreciate an asset, and according to what formula. If a building has a probable lifetime of profitable use which is approximately 50 years, the annual depreciation may be calculated on a straight-line basis as 1/50th of the cost each year. More conservative practice would call for its depreciation more rapidly than this, as by 1/25th each year for an assumed life of 25 years. Or the depreciation could begin at a larger fraction and decline each year until the building was fully depreciated. There are a host of possible depreciation formulas. Some depart from the time approach and use a formula which considers the amount of use following one pattern of cost accounting. Others allow for special depreciation charges whenever there is evidence of unexpected obsolescence or sudden decline in demand for the product of the equipment.

5. Valuation Problems Connected with the Rate of Profits.— Valuation decisions enter into the determination of the *rate* of profits as well as their amount. The rate of profits is the percentage calculated by dividing the amount of profits by some other amount. This amount is usually the net worth of the enterprise, but it may be anything else that the calculator chooses. Sometimes it is the original purchase cost of the shares of stock held by an individual. Or it may be the amount he put into a sole proprietorship or partnership. A different approach uses as a divisor the amount which the individual could realize by *selling* his share of ownership in the firm. This may differ from the initial investment for a host of reasons including

stock market gyrations, reinvestment of earnings, changes in technology, etc. From the viewpoint of the individual the numerator of the fraction should not be the total profits of the concern, but only that part which that individual receives, or to which he feels entitled.

The importance of estimates in calculations of the rate of profits can be further appreciated by considering the problem of valuation when net worth is used as the divisor. The net worth of a corporation, for instance, is the difference between total assets and total liabilities to persons other than the shareholders. Assets may be valued in various ways. The problem of inventory valuation has already been considered. More important quantitatively in most cases is the valuation of fixed assets. These may be carried in the balance sheet at original cost, reproduction cost, or present market value. If either of the latter alternatives is chosen, the personal bias of the valuator can be very important. Even original cost may be a larger or smaller amount, depending upon the carefulness with which the initial investment is made. This is revealed by public utility commission studies to determine "prudent investment" as compared with the "original cost" figures submitted by the firm.

Often there are also included among assets such intangible items as goodwill, trade-marks, value of patents originating within the firm, etc. The valuations required to make such entries often work backward from earnings by capitalization. This introduces circular reasoning when they are subsequently counted as part of the assets used to determine the current rate of profits on net worth.

Enough has been said to demonstrate the ambiguities and uncertainties involved in profit estimates, particularly when profits are expressed as a rate of return upon net worth. Additional items might be mentioned, but the point should have been well established by those already cited. If the owner does the estimating, he may get one figure; if an outsider performs the calculations, he may get another. The result may depend upon the skill or bias of the accountants used. Sometimes the purpose of the estimate is of major importance. When the rate of profit is being determined for tax purposes, efforts will be made to keep it low. When new capital funds are being sought, there is a temptation to exaggerate by one device or another. Public utility commissions have recognized this problem and at times have made a serious effort to secure the adoption of "uniform" accounting systems and methods. But they cannot standardize the accountants who apply the rules.

We are forced to the conclusion that there is no good answer to the question, "What is *the* rate of profits?" in any particular case. What

one should do is to tell how the calculations were performed and what assumptions were used in the various valuation processes. And when comparisons are made among the rates of profits of different firms or industries, one must examine accounting procedures to see whether they are similar. If they are not, some highly erroneous conclusions may be reached.

6. Profits May Include Work Income as Well as Property Income.—The rate-of-profits concept wrongly implies that profits are always a form of property income. When the profit amount expressed in dollars is divided by a chosen valuation expressed likewise in dollars, there is the clear implication that the profits are derived from the property thus valued. This is often far from the case, as will be shown later in sections dealing with determinants of the amount of profits. In small firms the residual income usually contains a large share properly attributable to the work efforts of the owner-manager. In large firms, particularly of the corporate type of organization, *dividends to nonworking owners* become clearly a property type of income. But these dividends must be distinguished from the *profits of the corporation* itself. And one must not allow the prominence of large corporations to lead him to make generalizations about profits which exclude the host of cases involving partnerships, sole proprietorships, and small corporations.

Confusion will be avoided as suggested above, if profits are thought of as *a way of receiving income,* whether work income, property income, or both. An individual in a capitalistic society has the option of seeking income in various ways. If he has no property, he usually is forced to choose the contractual method and goes to work for some one else. If the individual has property, his options broaden. He may let some one else use that property for a contractual rent or interest payment as the case may be. At the same time he may work for contractual wages. If he chooses the alternative of seeking income residually, he has several further options. He may purchase shares of stock or a silent interest in a partnership. His return is then likely to be in the nature of interest residually received. Or he may combine an "investment" of his time and effort with his investment of property. That is, he may work in and manage or help to manage the business unit which seeks profit income. The individual's return will then combine work and property returns. The third possibility, residual work income without residual property income, is very rare, but it is conceivable. A firm might have zero net worth appearing on its balance sheet and still show an accounting profit at

the end of the year. According to strict logic, this income should not be associated with any property of the owners of the firm, but must be considered a result of the activities of the firm itself. These activities may be identified with the decisions of management, the efforts of labor, the passive receipt of chance gains, ownership supervision, or other things as the case may be. Various possibilities will be discussed below.

7. Differences in Profit Rates Among Firms in the Same Industry.—The nature and origin of accounting profits can be further illumined by examining differences among firms. For simplicity, it seems best to begin with firms in the same industry. Differences in size may be eliminated by using the concept of profit rates instead of profit amounts, remembering the problems associated with their calculation.

In the light of Sections 2, 3, and 4 above, the first observation must be that differences in accounting methods are of primary importance. Depreciation and valuation procedures may differ widely among firms in the same industry. This is particularly true for small business firms, although there are some notorious instances involving large corporations.

Among small firms there are also differences in the amount of labor contributed by the owner. If he hires a manager or mechanic to do the work he might do himself, profits will be less than if he does that work himself. For income tax purposes, any wages a sole proprietor (or partner) pays to himself cannot be counted as an expense to the firm. They are classified as profits presumably because the owner does not enter into a contract with himself for certain services. The nature of sole proprietorships is such that the owner must get residually whatever income he receives.

A similar observation might be made regarding those profit differences which stem from differences in the length of the owner's workweek. One complaint of white farmers in California against those of Japanese descent was that the latter worked too hard and too long! They made an unfairly large profit, saved too much, and thus got ahead too rapidly.

8. Efficiency of Management as a Source of Profit Differences.—Profit differences are often due to differences in the efficiency with which rival firms are managed. However, this topic must be approached with care. It is not efficiency in management *per se* which produces large profits, but *relatively greater* efficiency. If all the competing firms are equally well managed, their profits may be small

because of competitive bidding up of purchase prices or cutting of selling prices. In fact, it is hard to measure managerial efficiency except in terms of profits. People have a tendency to reason backward from large profits to superior efficiency, although this inference may prove unwarranted when the acts of management are examined in detail.

A second caution might be given to the effect that attempts to infer relative efficiency from relative profits must start with the assumption either that the owners are the managers in each case or that the managers are paid alike. Since firms being compared are usually of different sizes and are organized in different ways, this assumption is rarely in keeping with the facts. The point is that hired managers of equal efficiency may be overpaid by one board of directors and underpaid by another, with obvious effects upon the residue called profits. If there are no hired managers, this discrepancy will not arise.

With these qualifications in mind, we may now approach the question regarding the nature of this attribute of a firm called "efficiency." It may be divided into three main parts: buying, selling, and operating. A fourth might deal with planning and cover such things as the choice of time and place for an enterprise, for new investment, etc., akin to the production planning type of competition described in Chapter 11, Section 8.

9. Efficiency Differences in Buying and in Selling.—Successful management on the buying side includes bargaining and discovery. Most materials are purchased by most firms at prices which are not subject to their control. But there are exceptions. Large buyers are often able to get price concessions not available to small buyers. Sometimes the shrewdness of the purchasing agent makes the difference. Success in discovery of willing sellers also contributes to profit opportunities, particularly when materials are scarce. In the factor markets, some firms get advantages by discovering rental opportunities and negotiating long-term leases at favorable terms. Or they may borrow at a time and place where lower interest rates are obtainable.

Probably the most important differences in buying efficiency, however, occur in the bargaining between management and labor. Here the problem is to get a more favorable contract than your competitor and hold your workers to it. Where there is industry-wide collective bargaining, no such differential is possible. The problem then becomes one of striving for the most advantageous terms possible for the group as a whole.

When bargaining occurs between managers of business firms and

those from whom they purchase, two effects are determined simultaneously. For instance, if organized workers succeed in getting a wage increase, management at the same time loses potential profit. Or if a lessor gets less of the gross income to be distributed, the lessee retains more. Changes in bargaining power are probably more important than changes in bargaining skill. But often superior skill is revealed in successful maneuvering for positions of power. The constant dickering between management and labor about who shall have the right to decide certain terms of employment is a case in point. By increasing its power in this field, labor is often able to increase its wage bill at management's expense. Each side may also plan how it can get outside support, as from other unions, from a trade association, or from the government. If the term *bargaining* is extended to include all these kinds of activity, efficiency of management becomes a wide phrase, indeed.

On the selling side there are also differences among the managers of firms which significantly influence the profits made. Bargaining is sometimes important, particularly when one firm sells to another rather than to the ultimate consumer. The chief difference, however, lies in the ability to promote sales inexpensively. Most firms have unused capacity most of the time and could increase their profit if they could raise their rate of sales. This requires selling efforts which succeed in attracting buyers without costing too much. It is the selling expense problem which was introduced above in Chapter 6, Section 1. Ingenuity in discovering effective sales appeals or astuteness in hiring advertising agencies or salesmen who can invent them are the conditions of success.

10. Efficiency Differences in Production Cost or Output and in Planning.—Profit differences also arise from differences in operating costs. This topic may be defined so as to eliminate those advantages which spring from fortunate purchases or bargaining successes. That management is most efficient which succeeds in devising ways of combining factors and materials so as to get more output per dollar's worth of inputs than do other managements. Success may come from inventing new machines, new processes, use of by-products, etc. It also requires either that competitors be kept ignorant of these inventions, or that the efficient management keep ahead by inventing something new while rivals are copying the old. Production efficiency is obviously an important source of profits in many cases. But it is only one source and should not be elevated into a complete theory of profits.

Finally there may be noted that group of efficiency attributes grouped under the head of planning. This includes superior foresight in knowing what to produce, when to produce it, and where. In order to expand at the right time, enterprises must make their plans far in advance. This requires good judgment of probable future actions by competitors, trend of the business cycle, legislation, consumer preferences, and the like.

This section should be concluded with the warning given at the outset: one must not infer high efficiency from high profits. Efficiency as defined here has covered many aspects of business, but there are other sources of profits, too. Some have been presented and others will be discussed shortly. And if *hired* management is paid in proportion to its efficiency, its contribution to profits will be nil. This point is so important that it will be analyzed further in the next section.

11. Underpayment of Factors as a Source of Profits.—Relative superiority in managerial efficiency will be a sure source of profits if the owners are the managers and take their pay residually. That is obvious. But if the managers are hired, another variable enters the picture, the pay given them for their services. This pay will be determined presumably by bargaining between the owners and the managers. If the owners can get a manager for less than he contributes through his efficiency to the income of the firm, there will be a profit residue from that manager's service. This is often the case for several reasons. A manager sometimes improves in skill at a given task more rapidly than his salary is raised. He is satisfied with the pay received, and his family likes the community. No one else has offered him enough more to warrant making a change. So he stays, and his employer profits.

Many such managers yearn for the day when they can have a business of their own and pocket the profits they feel they are making for their employers. Others work for such large-scale enterprises that such a hope seems futile. If not content with their pay, they merely seek to bargain for a higher salary. Then it is up to the owner, or whomever he hires to bargain for him (another form of manager!), to decide whether to pay the increase or to let the man go and hire a replacement who will probably need some training before becoming of equal worth.

The foregoing is not to imply that all managers are underpaid in this sense, but only that some of them surely are. Some also are overpaid, and the firm suffers losses accordingly. And again it must

be said that neither profits nor losses are *prima facie* evidence of underpayment nor overpayment. There are so many possible causes that no one can be singled out by reasoning backward from the effect.

Other factors may be underpaid or overpaid, too. This was explained in the case of land during the discussion of various concepts of surplus (see Chapter 22, Section 6). Long-term loan contracts may give rise to opportunity-cost surpluses in relation to interest payments. And workers below the managerial level should not be ignored. They may also be underpaid, although probably not as often as they claim to be. When they are well organized, they bring constant pressure to bear to raise their pay to the level of their productivity. They often judge this productivity by the profits being made by the firm, although the profits might be ascribed by some one else to a different cause, such as the underpayment of a different factor. At times workers may demand a pay increase which is definitely beyond the capacity of the firm to pay and stay out of the red. These overpayment situations do not occur nearly so often as employers claim in their negotiations, but some cases are not to be denied.

None of this discussion is intended to imply any ethical judgment regarding underpayment or overpayment. The question of economic justice is far too complex for that. On the other hand, there are some occasions where one side or the other takes advantage of an unusually strong position to "drive a hard bargain." This may be sanctioned by the ethics of the competitive system, may be appraised as shortsighted, may be condemned as "exploitation," etc., depending upon the viewpoint of the person rendering judgment. No sweeping generalization seems possible. The presence of profits, even of large profits, in a particular case does not warrant the assumption that some factor hired by the firm was unjustly underpaid.

A final comment regarding underpayment takes us back to some points made in Section 6 of this chapter. In the corporate form of business enterprise an owner may be also a hired manager, or an employee may own some shares of stock. Differences in profit rates between corporations, particularly small ones, may arise from differences in the amount of salary paid for managerial labor contributed by the principal owner. It is also probable that there are differences in the degree of underpayment of owner-managers of various firms. One set of owner-managers may decide to overpay themselves, while another prefers underpayment so that profits will be larger. Sometimes the desire to reduce total tax burdens stimulates the payment of high executive salaries.

12. Differences in Capital Structure Affect Equity Profits.—Firms often differ in their capital structure, and this affects the rate of profits on the owner's equity. A firm whose owners supply all of the funds will be in a different profit position from one which borrows much of the funds invested. Consider, for instance, the following hypothetical cases:

Firm A		Firm B	
Capital		*Capital*	
10,000 shares common stock	$1,000,000	4% mortgage bonds	$ 400,000
		6,000 shares common stock	600,000
		Total	$1,000,000
Earnings in 1946	100,000	*Earnings in 1946*	100,000
Earnings per share	$10	Interest on bonds	16,000
Rate of profits per share	10%	Remainder	$ 84,000
		Earnings per share	$14
		Rate of profits per share	14%

Firm B is said to have much greater "leverage" in good times. When net income before bond interest rises, the net per share of common stock will rise more rapidly than in Firm A. If earnings had been $200,000, a similar computation would show Firm A getting 20 per cent per share, and Firm B making 31 per cent for its shareholders. On a declining market the picture is not so rosy for Firm B. If earnings should drop to $16,000, all of this amount would have to go to the bondholders and stockholders would get nothing. But in Firm A, common stock profits would still show 1.6 per cent. The picture might be elaborated still further by showing other types of capitalization, such as those involving bonds and preferred stock. The fundamental argument would not change. Different capital structures often cause differences in profits when the latter is figured with reference to common stock. If the total capitalization of the firm is used as the divisor in calculating profit rates, then the explanation of interfirm differences must be found in the other causes cited above.

13. Profits May Result from Lucky Events.—Sometimes the profit or loss of a business firm results from occurrences which are clearly fortuitous. This appears most obviously in farming, where capricious storms may destroy the crops of one region while bene-

fiting those of another. Locusts, boll weevils, beetles, and other pests may invade an area beyond the power of any farm manager to foresee or to control. Similarly, fire may destroy the plant of the major customer or supplier of a given factory. This may bring losses which no farsighted manager could avoid by insurance. It may also bring gains to rivals which are clearly windfalls. A war may bring losses to thousands of firms whose management could not prevent the war nor take action to avoid its harmful effects. On the other hand the same war may bring undreamed-of profits to other firms whose managers did not seek it. Scores of other lucky or unlucky events might be cited.

In addition to these clear-cut cases, there are many borderline situations. Here it is difficult to state whether the management deserves credit for the profits or blame for the losses. Certainly the management's own appraisal of its performance is suspect. A neutral judge would be best. However, no segregation of chance gains and losses is necessary in a residual theory of profits. The firm gets what is left regardless of the origin of that surplus. If the owners want to know whether the management of their firm is doing a good job under the circumstances, they can render their own decision or hire an efficiency expert to write a report. The appraisal will have its chief influence upon the amount of profits through its effect upon the salaries paid the managers. If the judgment is favorable, managers may be given or may bargain for more pay. Without any change in the operation of the firm, this will tend to reduce the profit residue. If luck is given the credit and not management, a salary increase is not likely to occur, and profits will be larger. A similar argument could be constructed on the other side of the picture for unlucky or inferior managers.

Luck is put toward the end of our list of the causes of profits in order that its importance should not be exaggerated. Yet it should not be overlooked. Every firm "takes a chance." That is the essence of putting oneself at the end of the line to take what is left. The remainder may be large or it may be small, depending upon the various determinants explained above. At the beginning of a venture period that return is uncertain. It cannot be predicted with precision. But fairly close estimates often can be made. These are based upon *probabilities* which must be distinguished from possibilities. A person is not said to be lucky or unlucky when the probable happens. But if the improbable occurs, which is always possible, we describe it as good luck or bad luck as the case may be. Therefore, the minor element of uncertainty in every probability must not be elevated into

a theory which says that profits are essentially a chance return.[2] The latter phrase should be restricted to its usual connotation of a highly uncertain event. Luck may be present in every profit or loss situation, but it is not a complete explanation. Usually it is one of the least important determinants of the size of the residual return which we call profits.

14. Monopoly as a Cause of Profits.—Monopoly profits, so called, may result from the actions of a firm's management in charging prices which are abnormally higher than the cost of supplying the goods sold. This was explained in Chapter 8, Sections 12 and 13. The ability to charge high prices of this type for more than a very short period of time is described as a monopoly position in the market. This favorable position is one which others cannot attain quickly, or without considerable expense. It is obtained by the fortunate firm either by design or by chance. For instance, management may seek it through collaboration to secure agreements which restrict competition and exclude potential entrants. Or management may strive to obtain cost-reducing processes or equipment which are patented or kept secret as long as possible. Sometimes chance discoveries or unforeseen changes in demand or supply bring a favorable position which is not easily duplicated by other firms.

It seems reasonable to conclude, therefore, that monopoly as a cause of profits is not as fundamental as the concepts given earlier to explain profits in general. These include, particularly, the relative managerial efficiency factor, underpayment, and chance. When seen in the light of their causes, monopoly profits are not significantly different from other forms of profit, including those popularly called competitive. They become a distinct category only if defined very narrowly as those profits resulting from antisocial action to restrain some form of competition in an unethical or illegal way. And even these activities are managerial. All of which supports the approach of this chapter, which explains profits chiefly as a residual remuneration to a firm for the otherwise unpaid work performed by its employees and owners plus pay for the otherwise unpaid services of the property it owns and controls.

15. Owner Profits Compared with Firm Profits.—The business unit known as the firm makes profits in the first instance, but its rate of profits often differs from the rate of profits received by the owner. There are two main reasons for this difference. The first springs from

[2] One must remember that "profits" in this chapter means "accounting profits." On the concept of "pure profit" and its relation to chance gains, see the next chapter.

the interposition of the firm as an entity between the owner and the profit-making activities of the firm's employees. This is most readily seen in the corporate form of business enterprise. The directors of the corporation may distribute as dividends only part of the profits of the firm. The owner gets a second residue, as it were. First, the expenses of the firm are subtracted from gross revenue. Second, the firm sets aside that part of net revenue which it wants to keep for expansion or other reasons. Only the remainder is paid out as dividends. However, the owner's equity is increased because of retained profits. Some owners calculate their profits on the basis of reported net earnings per share instead of by the dividends they actually receive.

The second difference arises from the way in which each rate of profits is computed. For the firm the divisor is the net worth. For the individual owner the divisor usually is either the purchase cost of the shares he holds or their present market value. Either of these divisors will be different from the book value, which is based on net worth and sometimes they differ by several hundred per cent. Profit rates vary accordingly. This is important because owner profit rates have a greater effect than firm profit rates upon decisions to remain an owner, to become a new owner, or to expand ownership. This topic will be treated in the discussion of normal and abnormal profits below (Chapter 27, Sections 7–11).

16. Changes in Profit Rate of a Firm from Year to Year.—The profit rate of a firm may change from year to year for much the same reasons that have been advanced for the differences in profit rates among firms at any given time. Emphasis might be placed, however, on "lucky" changes in demand as the major variable in most cases. These include the cyclical ups and downs of business in general and of the particular industry of which the firm is a part. There also may be some counter-cyclical irregularities in demand affecting the individual firm, but not induced by the actions of its management. This last qualification is important if we are to segregate the chance developments from those stimulated by management. Among the latter would be changes in demand as affected by selling policies. It is obviously impossible to determine just where to draw the line in many cases such as that between sought and unsought increases in demand. But that difficulty does not destroy the advantage of recognizing that different causes of demand change do exist.

A second important variable from year to year is differences in cost of production. In recent times the negotiated rate of wages is

and has become of major significance. The annual or other renewal of union labor contracts involves the setting of terms which may make the difference between large profit and small, or between profit and loss. Of a similar nature, but usually of smaller importance and less frequent occurrence, are negotiations for leases and for refunding loans. Management skills are important here. When labor wrests concessions from management, the taxonomist is confronted with the problem of deciding whether to ascribe the reduction in profits to overpayment of labor or overpayment of management. But again the classification is unimportant unless one is combating a monistic explanation of profit.

17. The Role of Inventions in Causing Changes in the Rate of Profits.—Costs of production also differ significantly with changes in the processes or equipment used. Technological improvements are a constant objective of management. Often their discovery, though consciously sought, should be ascribed to chance. Their introduction involves another complication. If the invention is patentable and is productive of an ascribable increment in profit through reduction in cost, its value may be determined by capitalization. If management pays this value amount for such an invention, there will be no gain unless in the particular firm the invention will yield more than the return estimated by the seller. Thus an invention may be appraised by its seller on the basis of an expected earnings rate of $10,000 per year for eighteen years. If the buyer-user can make it yield $11,000 per year, he will gain by purchasing it at the seller's figure.[8]

A different situation exists when the invention is developed by someone in the employ of the given firm, but is patented by that firm. Several new problems emerge. Should the invention be carried on the company's books at an estimated capitalized value, at a nominal $1, or at an estimated cost of development? The decision will influence the amount of net worth and therefore the rate of profits computed on that base. The invention itself will raise profits per share and therefore potential dividends. Stockholders usually are not concerned with the accounting decision regarding balance sheet entries, but there are some exceptions. The management may use increased net worth as grounds for an increase in the number of shares outstanding, thus diluting the return per share. Or an invention might be carried on the books at a high figure and depreciated over a

[8] Assuming, of course, that the seller does not use a lower discount rate than the buyer. Cf. Chapter 19, Section 6, and Chapter 21, Section 14.

seventeen-year period, thus reducing current dividend payments. Public utility shareholders may be glad to have net worth padded. It reduces the rate of profits without altering the absolute amount and may prove helpful in efforts to increase rates.

A few other causes of profit differences from year to year may be singled out for special mention. Accounting practice may change, particularly with reference to depreciation. Worker efficiency often is influenced by wage adjustments, union organization, the prevailing volume of unemployment, etc. Strikes hurt earnings, even if "won" by the management. Government action may grant or withhold subsidies, alter price control, revise tariffs, or change taxes. Key management officers may be changed through death, new hiring, promotions, and resignations. Volume of output may be curtailed at times by scarcity of essential raw materials.

18. Differences in Average Profit Rate Among Industries.— Further light may be shed upon the nature and source of profits by an examination of reasons for the differences that exist at any given time among the average profit rates of different industries. The list of determinants should include all of those previously explained for differences among firms in the same industry and differences from year to year for a given firm. For instance, changes in the intensity of schedule demand occur at different times for different industries. At the close of the second world war, the demand for construction materials rose and that for military airplanes declined. On the supply side, cycles of material costs also may differ in their peaks. This was particularly noticeable in the period of irregular decontrol of prices following the war. Labor gains at the expense of firms may come at different times in different industries. And of course, there are differences in accounting procedures in various industries.

The most important argument to be developed in Sections 18 and 19 is that the average rate of profits in an industry is a function of the productive capacity of that industry in relation to the demand for the goods it produces.[4] Total productive capacity is made up of two components, the number of firms and their several capacities. The determinants of each of these may be examined briefly. Ease of entry is an important determinant of the number of firms. The nature of the industry and its monopolistic practices suggest themselves as the major factors influencing ease of entry. Some firms, such as corner groceries, are easily established because they require little capital to

[4] The troublesome question of what constitutes an industry need not be discussed at this point. A commonsense criterion of high substitutability of firm products will be implied.

achieve a scale of operation which has relatively low unit costs. Others, such as automobile assembly plants, require large-scale production with huge capital investment. In some cases the lack of requisite skills limits entry, as in the manufacture of chemicals. In other cases potential entrants are held back by inability to secure the use of certain patents, materials, or sites. At times fear of retaliation by existing firms may be a deterrent. These and other points regarding the problem of entry were discussed in connection with the classification of forms of business activity in Chapter 11, Section 11.

19. Industry Profit Differences Are Affected by Firm Expansion Programs.—In addition to aids and obstacles to entry, current capacity is a function also of prior expansion of existing firms. Here the determinants are closely interrelated. Expansion will tend to occur when profit rates are high. Profits will be high when firms are operating at or near capacity. When operations approach capacity, firms begin to plan for expansion of that capacity. Expansion also is planned when it seems likely to permit production of larger volume at unit costs lower than would be possible by overtaxing present capacity. (Compare Chapter 13, Section 7.)

Occasionally capacity expansion is induced by technological change. New plants are erected using the new equipment or methods, but the old ones are not immediately scrapped. Sometimes this is also what happens when there is entry of new firms. The erection of new plants in new places to secure labor, material, or marketing economies is of the same pattern.

Some industries characteristically plan farther ahead than others. They expand capacity in order to be ready for an expected increase in demand, as in the case of electric utilities. Or they may anticipate declining supply from other sources, as in oil well development. Some try to get ahead of rivals by early development of choice but limited sites. In a few cases overcapacity with attendant low profits has been the result of vertical development of firms seeking to control sales outlets below them in the production-marketing process, such as gasoline service stations. Efforts to control raw materials are probably less frequent and do not often cause excess capacity. They are more apt to represent attempts to avoid paying high prices to cartels which themselves sprang from excess capacity situations.

Finally, there is the question of the availability of funds. When profits have been high in the past, funds for future expansion are relatively easy to obtain. If not present in the reserves of the firms themselves, they may be secured on good terms from the general

public because of the attractive earnings record of the firm or industry.

In short, the profit differences among industries in a given year are a function of many variables. Chief among them are demand and investment fluctuations which differ in amplitude and in time of occurrence. The demand fluctuations relate to the year in question. The investment fluctuations relate chiefly to years prior to the year in question. A third important determinant is the extent of monopolistic activity which raises selling prices, resists increased buying prices, and restrains entry. It is often argued that profits are payment for taking risk, i.e., for assuming the chance of loss. This concept will be examined further in the next chapter, but at this point one must note that it does not serve as a good explanation of interindustry differences in profit rates in any given year. For instance, there is a high mortality percentage for beginners both in retailing and farming. This indicates the high risk of entering such enterprises. But those industries usually have a very low rate of return relative to other industries. The argument of a correlation between risk and the rate of profits must be defended, if at all, on other grounds.

20. Differences in the Average Profit Rate in Different Regions. —A final approach to the institutional explanation of profits may be made by examining causes of differences in the average rate of profits of firms in a given year when the firms of two regions are compared. For instance, the average rate of profits of New England firms is probably less than that of those in California or Alaska. The reasons are complex and have their roots in the history, the culture pattern, and the natural resources of the two areas.

The explanation of differences in average profit rates between two regions should begin with an explanation of the differences in interest rates. The average rate of profits and the average rate of interest tend to be high or low together because they represent substitute uses of funds. (We ignore for the moment the possible wage element in managerial profits.) The explanation of interest rate differences was given in Chapter 24, Sections 11 and 12. Some specific determinants of profit rate differences should be given special mention.

If the producers of a certain good in two different regions sell that good in the same market with the same transportation costs, then those producers will have the greater profits who have the lower costs. The cost of producing a certain good may be lower in one region than another because the former has lower factor prices or

greater factor efficiency. Lower factor prices result from a relatively greater abundance of the cheap factors in one region than in another. The influence of cheap land and cheap labor has long been felt in international trade. For some products, low cost results from abundant capital goods.

Abundance of the factors combined by business management would not of itself produce high profits. Enterprise funds and ability must also be relatively more scarce in one region than in another. It is differences in the multiple ratios among the several factors in two regions which bring about differences in the profits made by producers of those goods which enter into international problems. It is not merely greater abundance of one or all of them in one region as compared to another.

This argument is part of the theory of comparative costs in international trade and need not be given further elaboration here. It holds most clearly for single products from two regions. The low cost and high profit region may be either one of two types. It may be one which has high profits in general as compared to another region. Or it may be a region in which the given industry is new and in which the number of competing firms has not yet become large.

Differences in the average level of profits may also arise from differences in the general culture patterns of two areas. One may be highly competitive and the other may have many obstacles of entry to protect existing firms. Inventions may be frequent and technological change rapid in one area and slow in another. Accounting practices may differ. Tax structures are not often the same and occasionally may be quite different. One may have more government ownership and control than the other. Finally, the regions may differ in their stage of the business cycle. One may be near the peak of its boom while another has not yet reached it or has gone beyond it.

21. **Summary.**—The concept of profits most commonly used in this chapter is that of the accounting profits of firms. These are the net income which accountants compute by subtracting costs from gross income. They may be expressed as an amount or as a rate. Profits may also refer to the funds received by individual owners of a firm. In either case they are a form of residual income which in most cases is derived from property, not from work.

The *magnitude* of the accounting profits of a *firm* is a function of:

1. The valuation processes used by accountants, especially regarding depreciation and inventories

2. The amount of uncompensated funds, labor, or durable goods contributed to the enterprise by the owner
3. The relative efficiency of the firm in comparison with competitors. This includes efficiency in production, buying, selling, and planning.
4. The amount paid to factors in relation to what they contribute to the efficiency of the firm. (Underpayments may be based upon the factor owner's bargaining weakness, ignorance, immobility, etc.)
5. The capital structure: the ratio of stock to bonds, etc.
6. Windfall gains not attributable to the actions of any factor
7. Monopoly position: usually a combination of management efficiency and luck

The *rate* of profits of an *equity owner* is a function of:

1. The amount of profits which the firm distributes to the owners
2. The purchase price of the owner's equity share; or its present market value
3. The difference between the purchase price and the sales price of shares the owner buys and sells ("speculative" gain or loss)

The profits of a firm *differ from year to year* because of changes in all the variables described above, but particularly because of changes in:

1. The demand for its products
2. Its costs of production: both factor unit costs and efficiency
3. Its accounting practices
4. Taxes

Differences in the average profit rates of *various industries* are a result of most of the above determinants for individual firms, but especially:

1. Different demand intensities for the products
2. Different aggregate capacities to meet the respective demands
   (a) The number of firms: past stimuli to and opportunities for entry
   (b) The size of firms: past stimuli to and opportunities for expansion

Interregional differences in average profit rates result from both general and particular relationships:

1. The current demand for and supply of funds for *all* productive uses, whether as loan capital or as equity capital

2. The relative costs of production of a certain good depending on factor costs and factor efficiencies, especially
   (a) The abundance of equity capital in that industry
   (b) The abundance of other factors in that region
3. Different culture patterns regarding business

Chapter 27

PROFITS: SOME THEORETICAL AND SOCIAL PROBLEMS

1. Problems of Profit Theory.—Up to this point we have been concerned chiefly with the institutional causes of profit, profit differences, and profit changes. In this chapter our attention shifts from profits as an effect to profits as a cause. Institutional explanations are supplemented by more general theories. The questions to be answered include the following:

1. What effect does the rate of profits have upon the supply of the factor which receives profit?
   (a) Upon the individual firm, its capacity and net worth
   (b) Upon the number and total capacity of firms in an industry
   (c) Upon individuals as equity owners
2. Why must the opportunity to earn profits exist in a free enterprise economy?
3. Why does the average rate of profits for industry as a whole tend to be positive?
4. Can a normal rate of profits be defined comparable to the concept of normal prices, normal wages, etc.?
5. What part does risk play in explaining the rate of profits in general or in particular cases?
6. What effect does the taxation of profits have upon the economy as a whole?

2. The Profit-Elasticity of Supply.—When profits are considered as a cause, the first problem to be considered is the way in which profits affect the supply of the factor which receives profit. This may be called the problem of the profit-elasticity of supply. It resembles the problems discussed in connection with the price-elasticity of supply of commodities, the wage-elasticity of the supply of labor, etc. In these other fields the effect of prices upon demand was also considered, but the nature of profits as a residual share makes the demand approach difficult and perhaps unimportant.

The profit-elasticity of supply of the profit-receiving factor can-

not be analyzed until that factor is defined. According to the argument of the preceding chapter, this factor seems best thought of as either (1) a firm or (2) an owner of an equity in a firm. Both receive profits as a residual form of income. The solution of the supply-elasticity problem requires a reasonable explanation of (1) the capacity and net worth of a given firm, (2) the number and total capacity of firms in a given industry, and (3) the entry of individuals into the equity ownership position and their exit from the equity ownership position.

3. The Profit-Elasticity of Individual Firm Capacity.—The profit-elasticity of supply relative to the individual firm must be examined in relation to the changes in the firm which are apt to be induced by changes in the rate of profits to net worth. These include particularly (1) the capacity of the firm, (2) the percentage of net worth to total capitalization, (3) the output as a percentage of capacity, and (4) the cost of production.

4. Firm Capacity Expansion When Profits Rise.—The capacity of a firm tends to vary directly with the rate of profits. The management is much more likely to plan expansion when earnings are good than when they are poor. Funds for expansion are obtainable more readily when the firm's rate of profits has recently increased than when it has decreased. Three reasons may be noted. First, large profits permit large retention and reinvestment of earnings. Second, they stimulate the allocation of previously accumulated cash reserves to an expansion type of investment. Third, additional funds may be obtained more easily from outside the firm. This is true whether the funds are sought from present owners, from potential new owners, or from lenders. The correlation between changes in the rate of profits of firms and changes in capacity is probably positive, but not high. It is highest when profits are increasing.

The minimum profit rate which a firm must earn if it is to attract *loan* capital for expansion cannot be defined in any general way. Potential lenders demand ample security in terms of either the amount of profits or the collateral pledged. The amount of profits, as distinguished from the rate on net worth, indicates the margin of protection afforded the lender. The degree of safety is sometimes measured by expressing the total amount of profits as a multiple of the total amount of interest payable on the loan. Another test is the ratio between the value of the collateral and the amount of the loan. It is true that a high current rate of profits makes lenders less exacting in their safety margin demands or collateral requirements. But

there is no generally accepted minimum rate which they will demand as a condition for lending to a given company.

The same argument holds for equity capital. It is the peculiar nature of residual income that it may fluctuate widely and cannot be predicted with great certainty more than a short time in advance. Prospective earnings in future years may often be a greater inducement to equity investment than high earnings now. If this were not so, few firms would be able to get started or to grow from small beginnings. Many people buy stock chiefly for expected dividends and appreciation, factors which need bear no close relation to the present profit rate on net worth. In the light of these facts it is difficult to be sure how much correlation exists between a firm's rate of profits and its ability to attract new equity capital. Even though the evidence probably justifies a cautious affirmative answer, there is no definable minimum profit rate which must be exceeded at any given time before a firm can get equity capital. Certainly there is no normal rate below which existing equity capital will be withdrawn. Further comments on this point must be deferred until Section 10 where the discussion shifts from the rate of profits of the firm to the rate of profits of the stockholder.

5. Firm Capacity Contraction When Profits Fall.—When profits fall, but do not become negative, *capacity* to produce is not often diminished. *Output* may decline as firms shift to produce other things, as described above. Output also may be reduced by voluntary curtailment following the $MC = MR$ principle described in Chapters 8 and 10. But output is not diminished because of reduced capacity. The most that one can say is that falling profits generally restrain increases in capacity. Firm owners are reluctant to add to their investment in equipment. Lenders are not so willing to lend to the few brash owners who still want to expand. On the other hand, one should note that depreciation charges permit replacement of equipment. In periods of rapid technical progress, the replacement equipment may be more efficient or cheaper than the old. Under such circumstances, firms showing as little as zero profits *may* increase capacity. When a firm sells a part of its plant which it can no longer profitably operate, its capacity to produce is diminished. The proceeds from the sale, however, may be used to meet pressing obligations. This may help the firm to avoid still further contraction.

6. The Profit-Elasticity of Other Determinants of the Profit of the Firm.—In addition to the correlation between profits and capacity, a second supply-elasticity relationship should be noted. It is the

manner in which the rate of profits on net worth is influenced by the percentage of net worth to total capitalization. If a firm borrows a large part of the capital funds invested in fixed plant, the rate of profits on the remaining capital contributed by owners will fluctuate widely with changes in earnings. The amount borrowed in this way is chiefly a function of the type of business and the arbitrary decisions of management. In the case of railroads and public utilities, usually there is much real property against which mortgage bonds are relatively easy to sell. In other cases, promoters, directors, or top management officials may decide to seek funds in the bond market or from banks rather than from potential investors in equity capital. Various motives may be present, such as the desire of present stockholders to retain full control. The correlation between such decisions and the current rate of profit of the firm is probably so small as to be negligible.

A third correlation in this group is that between profits and output, capacity being held constant. The output of a firm as a percentage of its capacity is influenced by the firm's rate of profit when the firm produces to stock rather than to order. In such case a high profit rate often induces management to expand output in hopes of an increased rate of sales in the future. This profit-elasticity of supply then operates in the usual manner to reduce the actual or potential future price of the product and tends to force the profit rate down again. But there is no definable rate of profit which is generally considered "high." This would require the presence of some normal rate to serve as a measuring stick, but this rate seems impossible to discover or to define. A mere upturn in the rate of profits after a decline is likely to stimulate such expansion, regardless of the profit rate at which the trend is reversed.

The fourth possible source of a positive profit-elasticity of supply (mentioned in Section 3) is changes induced in the cost of production. If a rising profit rate generally stimulates greater efforts to improve efficiency and to reduce costs, then a positive correlation exists. If a falling rate more commonly provides such a stimulus, the correlation is negative. If the efforts of management in this direction are continuous and persistent regardless of the profit rate, as is most generally the case, there is no correlation at all. The first alternative seems less probable than the second or third. If the correct generalization is in favor of a slight negative correlation, the supply effect offsets some of the slight positive correlations noted above. Whichever conclusion is supported, there is no denying the argument that there is no "normal rate" of profit above which addi-

tional efforts are forthcoming and below which they are reduced. In an acquisitive society, the labor element of the ownership position tends constantly to press for more profits, regardless of the profit level currently attained.

7. The Profit-Elasticity of Supply of Firms: The Problems of Entry and Exit.—Our next questions relate to the possible connection between the rate of profits being made by the firms already in an industry and the probable entry of new firms or exit of existing firms. If a positive correlation exists, as seems likely, then a rise in the general profit rate will set in motion forces which subsequently will tend to reduce that rate.[1]

Entry into the position of business ownership is based upon choice and the ability to execute that choice. The amount of prospective profits influences the decision. The choice is not merely whether to enter or not, but what to enter, and what to do if one does not. The three major alternatives may be outlined as follows:

1. Whether to put one's factors to use or to keep them idle
2. Whether to put them to use in a contractual (interest or wage) position or in a residual (profit) position
3. In what industry and what firm to take the ownership position

The basic choice is that between keeping one's capital and labor idle or putting them to use in one way or another. When people enter the profit-seeking position, they demonstrate a preference for it over the other two alternatives. This preference is undoubtedly influenced by profit prospects. The higher the expected profit, the more capital and labor will be put to work in an ownership capacity. But there is no definable minimum rate which must be exceeded before this will occur. The trend of profits is more important than the amount, or rate. Nor can one be sure when entry occurs that the incoming factors come out of idleness rather than from other equity positions or from contractual employments. The labor element of the ownership position is also difficult to pin down. Probably more time is diverted from "idleness" to owner-management work when profits are falling than when they are rising. If this be true, labor inputs have an inverse relationship at the same time that capital inputs have a direct relationship to the rate of profits.

The second choice is that between seeking residual profits and lending at interest (or working for wages). It involves a compari-

[1] The subsidiary question of major importance is whether the correlation is positive for the entire range of profit rates from the highest down to zero or whether it becomes negative at some minimum rate before zero is reached. If there is some such minimum rate, how may it be defined? See Sections 12 and 13 below.

son chiefly between the loans that can be made and the business opportunities that appear open. There is a definite tendency for investors to favor ownership when profits are relatively high and to favor loans when profits are relatively low. The potential entrant who expects to contribute only capital differs from the one who expects to contribute management labor also. The pure investor of funds is likely to be influenced by calculations of relative interest and profit returns more than is the potential owner-manager. Those who buy stock in new enterprises often are lured by tales of large profits made by a few firms of that type in the past. These profit seekers rarely demand to see careful statistics proving that the current average profit rate in the industry exceeds the interest rate obtainable from bonds or realty mortgages.

8. Choosing What Industry to Enter.—For a potential entrant the attractiveness of industry A as compared with other industries is definitely a function of the rate of profit being made by the firms in each. The comparison may be in terms of averages, or of the number of firms in a given profit bracket (compare the bulk-line-cost analysis of Chapter 10, Section 10). The low average profit of an industry may be offset by the conspicuous success of a few firms. This is the lottery spirit bolstered by the chance to bet on one's own ability as well as one's luck. Another important basis of choice among industries is ease of entry, which will be explained shortly.

His ability to execute the decision to enter a particular industry may often determine the choice of the potential entrant. Ability to enter is a function of (1) the amount of funds which the potential entrant owns, can borrow, or can induce others to supply in the equity position, (2) the skill and energy which the potential entrant can bring to the organization of the enterprise, (3) legal obstacles to entry, and (4) economic obstacles imposed by existing firms which want to prevent entry of competitors. The prospect of high profits may make it easier to raise funds. Economic obstacles to entry, and sometimes legal obstacles, may be overcome by sufficient outlays, but these raise the cost of production, which must be recouped through sales. Very high profits sometimes must be in prospect to induce potential entrants to incur such expense. In general, entry will be more likely at high rates of profits than at low, although this must not be taken to imply that other determinants can be ignored.

9. The Profit Causes and Conditions of Exit.—The chief type of exit influenced by profits is that which occurs when a producer shifts his facilities from the production of one commodity to the pro-

duction of another. This may be caused by falling profits in one line, or by rising profits in a substitutable alternative. An obvious illustration may be found in the case of farmers who shift from one crop to another when changes occur in the relative profitability of different crops as market prices rise and fall. Manufacturing equipment is sometimes so unspecialized that it, too, can be used for more than one type of product. During the second world war, many factories "converted" from peacetime products to munitions, and then after the war they "reconverted." The cost of such shifts obviously differs with the type of building and machinery involved.

In cases where the cost of conversion is so great that there would be no benefit from shifting facilities to another product, losses may not force exit but only recapitalization. Corporations with large interest charges on bonds or other loans often "go through the wringer." The creditors take over the business and oust the former stockholders. When the bondholders or banks become owners, their former fixed interest claims become contingent profit claims.[2] Costs are thus reduced and the firm continues in operation with its productive capacity unimpaired. It may even save so much on fixed charges as to be able to put aside or to borrow funds needed for new equipment.

To summarize, there is a much higher correlation between rising profits and entry than between falling profits and exit. Entry and exit occur in several ways. Basic factors such as land, labor, and capital funds may enter or leave a particular occupation. The alternatives to that occupation may be idleness or some other use. The same is true of specialized factors like buildings, equipment, administrative organization, etc. The more specialized the factors, the less the profit-elasticity of capacity governed by entry or exit. This is especially true for exit when profits fall. The profit-elasticity of output and capacity for existing firms was discussed in Sections 4 and 5.

10. The Profit-Elasticity of Supply of Individual Equity Owners.—The third major subdivision of the problem of the profit-elasticity of supply of the profit-receiving factor deals with particular individuals who may be induced to enter or to leave the equity ownership position. It follows logically after the question of the creation or the dissolution of producing units discussed in the preceding

[2] There are many variants. Entire replacement of stockholders need not occur. Owners of preferred stock may be given some common stock. Bondholders of various classes need not become stockholders, they may merely move down into positions of poorer priority in claims upon earnings, as by accepting income bonds in exchange for first mortgage bonds.

section. It is related also to the problem of expansion of firm capacity analyzed in relation to profits in Section 4.

The reactions of individuals to changes in profit rates depend primarily upon their expectations of yields on shares bought or sold. These yields depend proximately upon the price of the shares and the expected dividends. Ultimately they depend upon the amount of profits earned by the firm. In addition to yield expectations, individuals who buy or sell equities are influenced by expected appreciation or depreciation. Other subjective factors also may be important in certain cases, but do not give rise to significant generalizations.

The fundamental questions to be examined are two. First, is there a positive correlation between the current yield on equity shares and the entry or exit of people from the ownership position? This must be analyzed in terms both of firms in general and of particular firms. Second, is there some definable minimum profit rate below which there will be exit and above which there will be entry? If such a rate can be defined, it might be called the normal rate of profits to owners, since it would keep ownership constant.

When profits in the sense of stock yields rise, there is a definite inflow of funds into the equity ownership position. This is shown by the rise in the aggregate value of all shares traded on the stock markets. The profit-elasticity of supply of equity funds is so great that the average profit yield rises very little as dividends increase. At times the current yield actually declines because of large speculative bidding for securities, although not all of that bidding can be traced to realized increases in dividend yields. Much of it is based upon the *expectation* of a *further* rise in dividends. Other demand comes from the expectation of a rise in the prices of stock shares themselves. Motives are so mixed and vary so much from case to case that the generalization about the positive correlation between profit yields and entry must not be allowed to obscure the presence of other independent variables with which positive correlation also exists.

In addition to this purchase of outstanding stock shares at rising prices, individuals also enter the equity ownership position by buying new issues and by starting new companies. These new firms may be either of the proprietorship, the partnership, or the corporate form of organization. New security issues will attract new equity funds if their prospective yield is higher than the yield that could be obtained on similar shares in the recent past. Again other determinants are present, but they need not be elaborated. The amount of equity capital going into new companies at a time of rising profits will tend

to exceed that going out. One reason is the optimism engendered by the profit trend. Another is the lower rate of bankruptcies during periods of general profit improvement.

11. Other Alternatives Exist for Potential Takers of the Equity Position.—The amount of funds which will enter the equity position under the stimulus of a rising profit level is limited by other opportunities for the use of funds. Most important among these alternatives is that of loan contracts. If many people try to sell their bonds, for instance, in order to buy stock, the interest yield on bonds may rise to the place where it appears so attractive to potential stockholders that they may prefer the contractual to the residual position. Individual preferences differ, but the aggregate response of all individuals is the focus of attention here. Other alternatives that may gain in relative attractiveness as one's equity investment increases include (1) increasing safety hoards of liquid funds, (2) reducing debt balances due others, and (3) consumption. These last three are mentioned to emphasize three other sources from which equity-seeking funds may be drawn: hoards, borrowings, and current saving out of income.

Despite the presence of these other opportunities for using funds, there seems to be no definable minimum profit rate for equity owners below which there will be exit and above which there will be entry. The most that one can say is that loans in general are a substitute for equities in general for investors in general. Substitution between loans and equities on the basis of relative yields does occur. The process tends to depress the current yields of the securities being bought and to raise the yields of those being sold. By a sort of arbitrage the range of interest rates tends to be kept fairly close to that of profit rates.

This brings up the question of the difference between current yields and what may be called investment yields. Current dividends constitute the numerator of the percentage fraction in each case, but the denominators are different. In the first, the divisor is the current price; in the second it is the past acquisition cost. Investment profit yields are often very large, particularly when stocks are held for a long time in growing companies. They may go as high as several thousand per cent per year on the original purchase price. Interest yields, similarly figured, do not go so high.[8] At the other extreme stand both bondholders and stockholders who have lost everything.

[8] There may be a few very rare cases where defaulted bonds have been bought at a few cents on the dollar and have later resumed interest payments, but even there a maximum interest rate holds down possible yields.

Lenders, however, do not get wiped out so frequently as equity owners, if only because lenders have a prior claim against assets when times are bad.

12. The Concept of Normal Profits in Equilibrium Theory.— In keeping with the pattern of price theory followed in this volume, normal profits may be defined as that rate of profits which tends to perpetuate itself unchanged. It stimulates no shift in supply or demand schedules. In profit theory the supply side is most amenable to analysis. Here we note that normal profits must be defined in three parts as suggested in the preceding sections of this chapter. The first deals with a given firm where the normal rate of return on net worth stimulates no change in capacity, output, or capitalization. The second concerns an entire industry, and normal profits provide no incentive to net entry or exit. The rate of profits involved in the industry approach may be either that of particular firms, the average for the industry, or that of equity owners in the firms. The third concerns enterprise in general relative to individuals in general. In this third case a normal rate of profits on equity ownership will stimulate no net entry nor exit in the aggregate.

The current profit rate should be seen as *only one* factor in the estimates of future profit-type gain which influence human action. The *opportunity* to make certain *expected* profits is what attracts or repels. These include accounting profit, dividends, and appreciation. There is much less certainty in predictions of this type of financial gain or income in the future than in the case of contractual incomes. In fact, the connection between current profit rates and future profit gains is so slight that one is tempted to omit it altogether. Nevertheless, expectations do spring in part from observations of recent conditions including the current rate of profits. Therefore, the remainder of the discussion will concentrate on that rate.

People with funds may use them in at least four other ways in addition to the quest for profits. For simplicity, funds are taken as representative of factors in general. The argument would not be changed if labor time and tangible property were included as things that people may use in profit-seeking in addition to funds. On the income side, there are lending opportunities. Non-income-producing uses include hoarding, spending for consumption, and paying debts. Normal profits exist when the pull toward the equity ownership position exerted by profit prospects for people in general just balances the pulls in these other four directions. This equilibrium position in the short run is best seen as a balancing of these five positive desires.

This approach seems preferable to a balance between profit prospects and an abstract negative force labeled "risk." Each possible use of funds has disadvantages, or risks. Each use represents the foregoing of an opportunity to put the funds to some other use with its own positive and negative aspects. When these opposing forces are enumerated as in the preceding paragraph, their differences and their influences become clear. They may be seen also as rivals among themselves, not only as rivals for the equity-position use of funds. For instance, the potential spender may choose to lend, to hoard, or to do something else with his funds.

We are here concerned with people in general as users of funds. They differ among themselves. For each particular use individuals will have different intensities of preference. For instance, a cumulative schedule for profit receivers would include many intramarginal people whose desire for profits was less intense than those at the margin. The aggregate demand and supply schedules for equity funds determine profit rates and profit prospects. Intramarginal people and intramarginal investments by any one individual receive a seller's surplus of the type described in Chapter 22, Section 3, and elsewhere.

13. A Definition of Normal Profits for Short-Run Situations. —A normal rate of profits may now be defined in summary as one which for its particular situation offsets the four forces which tend to pull funds away from investment in equity shares. Short-run profit equilibrium is thus defined in terms of at least five variables. These seem to be the only ones that need to be mentioned in a statement of the formal conditions of equilibrium unless the motives of firm managers be given further analysis. The long-run conditions of stability in a dynamic society, however, involve many more variables, such as all of the other things that influence profit expectations whether of the income or the appreciation type. One must not confuse the conditions of equilibrium in a hypothetical static state with the conditions needed for stability in our modern economy.

The price of equity shares should remain fairly constant in any situation in which normal profits exist. Their price stability under normal returns is probably less than that for bonds at times of normal interest or for tangible goods earning normal rent. The difference springs chiefly from the greater importance of expectations in the profit field. These enter even the definition of normal profits. Appreciation possibilities also are given more attention in the market for equity shares than for bonds, land, or buildings.

14. Pure Profits as the Excess Above Business Profits (Opportunity-Costs).—Some writers use the theory of opportunity-costs to derive a concept known as "business profits." This is described as the amount which an investor could obtain in the contractual markets for his funds, labor, and tangible property. He is then said to be unwilling to take ownership risks unless he receives at least as much in profits as the sum of these possible interest, wage, and rent returns.

The statistics of profits on net worth are then examined to derive so-called average profit rates to compare with average interest rates. When the former are found to be larger than the latter, the difference is said to represent "pure profits." This excess is claimed to prove that investors in general require additional recompense for taking their returns the risky, residual way instead of the safer, contractual way.

Our argument has been that current expectations are the most important thing to examine, not long-run averages. We want to explain the interdependence of interest and profit rates and the shift of people from one position to another. This involves the study of the various motives of potential equity owners, as explained in the preceding section on normal profits. It does not require any statistical average rates of profit on net worth. Attention focuses rather upon current yields on the cost of equity shares and the appreciation in income and market value which investors in general hope to realize. This avoids all the statistical difficulties which beset the analyst who tries to measure an arithmetical difference between some sort of average profit rate and an average of interest rates.

In cases where current profit yields exceed current interest yields, one may argue that investors attracted to this industry desire security against uncertain or diminished income more than they desire the possibility of increased income. This preference varies from industry to industry and from time to time. Where growth possibilities exist, profit yields tend to be less than interest yields. For mature industries or recession periods, the reverse is more often true.

15. Profits as Payment for Taking Risk.—"Pure profits" should not be made synonymous with payment for assuming risk. This identification is a major weakness of the theory which uses pure profits as a term to describe the excess of accounting profits above business profits. In the first place, risk begins the moment one takes the equity position. Risk does not remain absent until after a certain opportunity-cost level of profits has been reached.

In the second place, all those dependent upon a business enterprise

for their income take some risk except, perhaps, those who receive payment in advance or at the moment of sale. The risk is in proportion to two things: (1) the legal sequence of claims against the firm's income or assets and (2) the length of time that elapses between the performance of service (including the delivery of goods) and the date of payment.

The equity holder is at the end of both the legal and temporal lists. Therefore, his position is particularly risky, but the difference is one of degree, not of kind. Interest contracts, for instance, are at higher rates for some types of loans than for others. These interest differences are usually ascribed in part to differences in risk as seen by lenders. An individual seeking income from his property balances his hopes against his fears in choosing which position to occupy. If his hopes prove the stronger, he chooses the tail-end position with its greater risk. If his fears win out, he chooses contractual income with its lesser risk. Or he may hedge by occupying both positions at the same time. For instance, he may sell his labor contractually and lend some of his funds. The rest of his funds may go into preferred or common stocks, the former being considered less risky than the latter because of precedence in the distribution of firm income. A particular individual's choice is a matter of temperament, experience, egotism, and opportunities. The generalization, however, must be that *one* restraining force which limits the flow of funds and labor into the residual position is its *relatively greater* risk.

A third difficulty with the theory which associates profits exclusively with the assumption of risk is that profits should then be proportionate to the risk assumed. The more risky the enterprise, the higher should be its average rate of profits to attract timid funds. An examination of profit statistics for a variety of industries, however, reveals no such correlation. Logic also denies the truth of any monistic explanation. In Chapter 26 a whole series of causes of profit differences were explained. In an earlier part of this chapter the profit-elasticity of the supply of equity funds has been examined. Taken together, these two approaches to the explanation of profits furnish a far better understanding of past and present profit rates than any argument about either the uniqueness or the primacy of risk in the determination of profit rates.

Only in the very difficult case of attempted comparison between the general level of profits for all enterprise over a long period of time and the general level of interest rates for all loans does the risk theory have possible merit. If it can be demonstrated that such an average level of profit rates to *owner*-investors is greater than the

average level of interest rates to *lender*-investors, then some reason must be sought. It might be found in a general margin of preference among investors for the greater certainty of contractual interest as compared with the greater opportunity for gain or loss from residual profit. Statisticians are still examining the evidence and have reached no consensus regarding it. In the meantime, there is little to be gained by *a priori* arguments about general investor aversion to risk. There is much to be lost if it diverts attention from the specific causes of individual and average profits and the changes therein.

16. The Motives of Equity Entry: Risks Deter and Profits Attract.—Business enterprise resembles the so-called "games of skill and chance" in which the luck factor is very important but where skill is also useful. Business enterprise, that is, consists of something more than a lottery gamble. Potential enterprisers can bet not only upon their luck but also upon their ability. People are induced to play games of chance and skill by the chance for rewards. It is difficult to state just how large the rewards must be, but the possibility of large winnings must be present. If participants *never* obtained much more than they could get by using their time and funds in other ways, the game would not prove nearly so attractive. But if there are a few big prizes known to have been won in the past, entrants will be attracted. Like many commercial games of skill and chance, the average rate to the player may be less than zero. The loss to the players represents the "take" to the owners of the game. In the field of business enterprise society gets the "take," if any.

Proponents of the risk theory of profits are often inclined to suggest that there is no "take." They hold that, on the contrary, potential entrepreneurs have to be paid by society to play the enterprise game. These risk theorists argue from averages rather than from particular cases. They contend that if individuals do not expect to earn from enterprise on the average more than they could get by lending or lending-working, they will sell their services contractually. To bolster this argument they sometimes try to marshal figures to prove that on the average, in the long run, profits exceed interest. There are many statistical pitfalls in such comparisons, such as the problem of allowing for the losses of those who vanish from the industrial scene entirely during the period being surveyed. The *a priori* argument itself seems to misconstrue the motives of the players in the enterprise game. Since statistical measurement is so difficult and when employed by different people yields such different results, there

seems little likelihood that it can prove the arbiter in the controversy. People will still advance either the risk theory or the game-of-skill-and-chance theory according to their appraisals of human motivation.

The special rewards from this game are of two main types: (1) very large income and (2) appreciation. Since most people invest money with the objective of increasing their wealth and not their consumption, the second type is very important. Equity funds are attracted particularly by rising stock prices. Most of the funds thus attracted do not go into business expansion. They go into inflated values of old shares. This represents no gain to society. It often means genuine harm because a speculative boom in equity shares may hasten and magnify business collapse.

17. Tax Policy in Relation to Profit Theories.—Taxes on profit income probably discourage somewhat the quest for that income. In order to analyze this problem, however, we must ask what are the places in which taxes on profits are in addition to, or are greater than, the taxes on other forms of income. Notorious in this respect are corporation taxes and excess profits taxes. The funds frightened away from the equity position may go into (1) new loans, (2) bidding up the price of old loan securities, (3) hoards, and (4) spending. Since funds are accumulated primarily to get more funds, neither hoarding nor spending is likely to be increased by any particular *level* of profits taxation. One must not confuse the effect of upward changes in profits taxation with the effect of a general level of such taxes long sustained. It is obvious that only *changes* in tax rates would lead to a bidding up of the prices of old securities. Therefore we are left with only the first alternative in the above list.

Funds discouraged from taking the equity position may go into the loan position. This will have the effect of reducing the prevailing interest rate. In the long run this result seems unlikely to do any particular harm to the economy and if there is a downward secular trend, good may even result. In earlier chapters the argument was advanced that we do not today need to worry about too little saving but too much; therefore, declining interest rates are not a menace.

18. General Social Objectives in Relation to Profits Taxation.—The appropriate tax policy for our country can be decided only by examining a large group of factors. Among these are the various objectives which we may seek. Assuming that the question of the amount of total government spending may be ignored in the present discussion, the question then becomes *who* should bear *how much* of the tax burden in order that *we* shall get the things that *we* want.

When a person uses the first person plural in this sense, he often refers to the first person singular. Therefore, the following comments must be understood as one person's appraisal of what he thinks society does or should want.

There are three possible goals which to many people seem desirable. First, efficient use of available resources; second, increasing efficiency in their use, often called technical progress; and third, maximum individual freedom compatible with a large measure of success in achieving the first two objectives. Let us examine the possible effects of profits taxes upon enterprise in relation to these three criteria.

19. Possible Effects of Profits Taxation upon Changes in the Business Cycle.—An efficient use of available resources implies continuous and full use. This introduces the cyclical problem. According to the theories presented earlier in this volume, the end of the business upswing may be attributed to a lack of balance between the demand for goods and the supply of goods. This is sometimes explained as resulting from a decline in the rate of consumer spending or business spending. On other occasions, stress is laid upon a rise in the rate of production. When the flow of goods to market exceeds, for any appreciable length of time, the flow of purchasing power demanding goods, unsold goods will accumulate and profit prospects will darken. A given *level* of profits taxation seems unlikely to cause *changes* either in the rate of spending or in the rate of production. An *increase* in taxes upon profits is likely to diminish profit prospects. This in turn may decrease business spending more than it diminishes the flow of finished goods to the market because much business spending is preparatory to future production. On the other hand, a *reduction* in profits taxes may stimulate business spending. The actual effect of such changes upon the trend of the business cycle will depend, of course, upon the amplitude of the change, the time at which it occurs, its effect upon consumer spending, and a host of other factors. The point to be stressed here is that one must not confuse the effects of a *change* in profits taxes with the effects of a given *level* of such taxes.

Does the level of profits taxation have any effect at all upon cyclical fluctuations? Statistical studies by the TNEC experts and others seem to indicate that a major proportion of large incomes is derived from profits. Large incomes provide most of the savings. Therefore, heavy taxes on profits should reduce the rate of saving and increase at least somewhat the rate of spending. This should tend

to prolong the upswing, particularly in comparison with its duration if equivalent tax revenue were to be derived from sales taxes.

The other side of the picture is the argument that profits taxation reduces the amount of investing which takes place out of currently saved funds. Reasons for doubting the significance of this contention have already been presented. If a high level of profits taxation induces spending rather than saving, that is to the good. If it stimulates hoarding saved funds, that would be bad. This possibility should be examined further.

There seems to be no major connection between profits taxes and the motives for hoarding. These include the desire for precautionary balances and the accumulation of funds for future large sum spending, lending, or giving. Hoarding for speculation might be reduced by the prospect of heavy taxes on speculative profits. But on the other side is the offsetting effect of a high level of profits taxation in stimulating people to lend rather than to buy equities. This will tend to reduce interest rates and may stimulate a slight increase in hoarding among those who are deterred from lending by the trouble and risk involved.

There is no apparent connection between profits taxation and operation of a business unit below capacity. The rate of output is determined by the relations between revenue and cost functions, in which profits do not appear.

20. Possible Effects of Profits Taxation upon Economic Progress and Individual Freedom.—Heavy taxation of profits might remove some of the incentives for industrial research aimed at methods of reducing costs. Much depends, however, upon the structure of profits taxation. If progressive rates are used or if flat rates do not go too high, business firms should still feel an incentive to increase profits before taxation. The prizes in the enterprise game are most effective if they can become large. A downward change in the possible maximum, such as might be caused by an increase in profits taxation, would probably discourage some types of research and betterment expenditures. The adverse effects, however, are likely to wear off and the business community would adjust itself to the lower maxima in such a way as to be stimulated to approximately the same type and amount of entrepreneurial expenditure or investment as before. To a certain extent high profits taxation may stimulate research expenditures as it has stimulated advertising and other outlays which are only indirectly connected with current revenue.

The third objective mentioned above was the preservation and

expansion of the areas of individual freedom. This is more closely related to the structure of profits taxation than to its average level. If new firms are given special exemption from taxation or credits for early losses, they will be stimulated. Small firms should probably be given special treatment, although there must be some debate regarding where the line should be drawn between large and small firms. If society chooses to stimulate new firms and small firms, tax favoritism is one of the best devices. Tax revenues not obtained from such sources must, of course, come from someone else, presumably the larger and the more well-established firms. Another reason for stimulating small enterprises in this manner is the indirect stimulus which it may give both to technical progress and to price competition.

21. **Summary.**—The amplitude of possible fluctuations in the various rates of profit is restrained by the profit-elasticity of supply. Rising profits induce expansion of firm capacity, an increase in the number of firms, and a rise in the demand for equity shares. Each of these results tends to reduce the profits below that rate whose height stimulated the expansion. Falling profits do not produce a comparable contraction in the number of firms or in firm capacity, although they do diminish the demand for equity shares.

Normal profits are those which exert no influence to alter supply factors and therefore tend to remain unchanged. Expectations of future changes in the value of equity shares and in the rate of dividends are probably more important than the current yield in influencing the supply of equity funds. Equilibrium exists when the profit inducements balance four other incentives, chief among which are contractual income opportunities and the desire to avoid risk. If business profits are defined as being equal to opportunity-costs for the owners' factors, then pure profits or losses may be described. This approach tends to identify profits with payment for risk, an argument that has many weaknesses. Its grain of truth should be used with great care.

The average level of taxes on profits affects somewhat the volume of economic activity. Usually, it is not so important as the tax rate structure and the timing of changes in the tax rates.

INDEX

Advertising, 118–122
Average factor revenue, 271

Balancing of aggregates, 80, 85
Banks,
  central, 386–387, 398, 431–434
  commercial,
    costs and interest rates of, 391–392
    as creators of credit, 387, 391
    cyclical changes in lending by, 408–409
    influence of, on interest rates, 423–424
    as lenders to business, 380
    limits to lending by, 393–394
  savings, 387, 389
Bargaining,
  and prices of commodities, 182–185
  and rents, 336–337
  and wages, 253, 268
Bargaining range, 183
Basing point system, 132–135
Borrowing,
  business, 377–380, 382
  consumer, 384–386
  government, 383–384
  non-interest determinants in, 384–385
  for purchasing power, 372–373
Bulk-line cost, 152–154
Buyer's surplus, 184–185

Capacity,
  of a firm, 106–107, 203–208, 455, 461–462
  of an industry, 454–455
Capital (assets, claims, equipment, fixed, goods, liquid, working), 376–377
Capitalization, 346–347
Cobweb theorem, 226
Collateral, 373
Competition,
  cost, 167
  direct, 165–166
  indirect, 167
  marketing, 165
  monopolistic, 111–112, 164–165, 196–197
  predatory, 166
  price, 165
  product, 164, 177
  production planning, 168
  purchasing, 166
  pure, 109–111, 162, 178, 180, 196–197, 247
  unfair, 167
Competitive behavior, see *Competition*
Constant returns, 205
Consumer's surplus, 184–185
Cost,
  acquisition, 365
  decreasing, 205–209
    monopoly conditions inevitable under, 208–209
  increasing, 204
  opportunity, 151, 358–359, 367–369, 471
  reproduction, 355, 357–358, 365–366
  selling, 118–123
Cost schedules of individual firms,
  average fixed cost, 98
  average total cost, 98
  average variable cost, 98, 101–102
  marginal cost, 99–102
  total fixed cost, 96
  total total cost, 97
  total variable cost, 96
Credit,
  created, 387–388, 391
  instruments, 389–390
  nature of, 388

Demand,
  concepts, summary of, 15, 27
  curves, 17, 24, 46, 48
    arithmetic, 49
    logarithmic, 49–50
    statistical, 29–30, 233–237
  derived, 245–246
  elasticity of,
    cross, 19–20, 66–67, 122, 125–128, 177, 245
    income, 23–24, 69–71
    logarithmic axes in estimates of, 49–50
    price, 39–56, 71–73, 233, 245
  joint, 245–246
  law of, 28
  market, 15
  schedule, 15
    causes of changes in, 17–24
    diagrams of increase in, 24
  schedules,
    collective, 30–32
    discontinuous, 36
    instantaneous, 32
    speculative, 36
    statistical, 29–30, 233–237
    successive, 32–34
    time differences in, 33

Differential surplus (rent), see *Surplus, differential*
Discounting,
  of future revenues and costs, 144–147
  in capitalizing, 346–348
Discrimination, see *Price discrimination*
Diseconomies,
  external, 211
  internal, 211
Distribution theory, 241
Dividends, 443, 467
Dumping, 173, 189
Durability of goods,
  and price fluctuations, 220–222
  and rents, 333, 346–349

Economic parasites, 313–316
Economic rent, 355, 357–358
Economics,
  definition of, 3–4
  principles of, 11–12
  purposes of, 9–10
  vocabulary of, 8–9
Economies,
  external, 209
  internal, 209
Elasticity, see *Demand elasticity* and *Price elasticity*
Elasticity coefficient, 234
Entry of equity owners, 464–469, 473–474
Entry of firms,
  ease of, 454
  excessive, 190
  obstructed by monopolists, 170
  price policies of competitors and, 128–130
  when prices are above normal, 219–220, 227
  when profits are above normal, 150–152
Equilibrium,
  of the consumer, 60–61, 419, 469–470
  of the economy, 274–275, 435
  of the firm, 148–150, 156–158, 469
    see also *Substitution of factors*
  of the industry, 150–158, 469
  of interest rates, 434–435
  price, 108, 115–116, 152–158
  quantity, 108
  rent, 344–345, 365
  wage theories, 269–293
Exit of firms,
  when prices fall, 217–219
  when profits are below normal, 150–152
Expectations,
  in attracting equity investment, 469–473
  in borrowing, 432–433
  in estimating rents, 334–336
  in production planning, 144–148
  theory of, 147–148
Extensive margin, 342

Factors of production, 242–243
  services and sources, 243–244
Featherbedding, 264, 289
Funds, 372–375
  accumulated by lender, 374–375, 387
  created by lender, 375, 387–393

Goods, inferior, 22, 37, 69–73

Hoarding, 393, 395, 474

Income consumption curve, 65, 69–71
Income effect, 71–73
Income surplus, 334–336, 344, 356, 368, 378
Indifference curves,
  outlay, 57–60
  utility, 60–63
Intensive margin, 339, 342
Interest,
  amount of, 375
  competition for loans and rate of, 403
  cost determinants of rate of, 391–392, 402–403, 416–418
  elasticity of demand and rate of, 431–434
  elasticity of supply and rate of, 423–426
  going rate of, 375, 381–382
  government influence on rate of, 383–384, 433
  interdependence among rates of, 381–382, 404–406
  liquidity preference and rate of, 421–423
  long-run normal, 434–435
  non-cost determinants of rate of, 392–393
  political motives of governments and, 396–399
  pure, 418–419
  rate of (face, nominal, coupon, going, real), 375–376
  and risk, 401–402, 415–416, 418, 421
  and time preference, 419–421
  and trouble, 415–416, 418, 421
  and waiting, 419–421
  yield, 375
Interest differences,
  at different times,
    demand causes, 406–408
    supply causes, 408–410
  at same time,
    demand causes, 403–404
    supply causes, 401–403
  in different places,
    demand causes, 410–411
    supply causes, 411–413
Inventions, 209–210
  and real wages, 301–302
  and unemployment, 325–327

Labor,
concepts of, 243, 250–252
demand for,
elasticity of, 262–263
equilibrium of, 270–274
supply of, 250–252, 255–257, 283–285
elasticity of, 258, 266–267, 281
equilibrium of, 275–279
Law of proportionality, 297–299
Lending, 415–426
bank, 380, 391–394, 422–424
government, 386–387
individual, 382, 425–426
institutional lenders, 422, 424–425
saving and, 394–396, 422
Leverage, 449
Liquidity preference, 21
and rate of interest, 421–423
Loans, see *Lending*
Long run concept,
moderately long run, 196–203, 209, 211
very long run, 203–211
Luck, and profits, 449–451, 452, 473

Marginal concepts,
cost, 99–102
factor cost, 263, 272
factor product, 271–272
factor revenue, 263, 271–272, 378
lenders, 416
product cost, 99–102, 272
product revenue, 271
revenue, 42–47
value product, 143–144
Marginal efficiency of capital, 407
Marginal productivity, 246, 297
and wages, 271–272, 285–287, 297, 307–309
Marginal rate of substitution, diminishing, 62–63
Marginal utility,
diminishing, 60–62
of money, 69
Markup methods, 89–92
Maximum efficiency, 105
Maximum profit,
equilibrium of the firm and, 148–150, 156–158
price and,
when quantity is constant, 94–96
when quantity varies inversely with price, 111–113
special cases, 112–115
quantity and, when selling price is constant, 109–111
Maximum total utility, 63, 469–470
Monopolistic behavior, 160–161, 164, 168–174
anti-buyer, 170–173
anti-competitive, 169, 178
anti-entry, 170, 190
in lending funds, 403
Monopolistic behavior—(*Continued*)
price discrimination, 135–138, 172
in renting land, 332–333, 337–338
in selling labor, 289
supernormal profits and, 173–175
Monopoly, 162
bilateral, 162, 183
collusive, 163, 178
labor, 289
pure, 180–183, 247
see also *Cost, decreasing*
Monopsonistic competition, 185–186, 403
Monopsony, 162, 180, 183

Oligopoly, 126–128, 162, 178, 180, 182, 247
Oligopsony, 162
Opportunity costs, 151, 358–359, 367–369, 471
Outlay, total, 40, 63–64
Ownership price, 244

Parasites, see *Economic parasites*
Physical cost curves and schedules, 104
Planning curves, 206–208
Price acceptance, 179–180
Price changes, long-run effects of, 124–131
on buyer attitudes, 125–126
on entry and exit, 128–130
on prices of competitors, 126–128
on public regulation, 130–131
Price-consumption curve, 66–69
Price determination methods, 83–92, 162–163
Price discrimination, 132–138
in the basing point system, 132–135, 188
limits to, 137–138
monopolistic, 172
requisites for, 135–136
Price-elasticity of demand, 39, 67
diagrams of, 43–54
average, 51
height of *MR* curve, 46
point, 51, 54
price-consumption curve type, 66
slope of *AR* curve (logarithmic), 50
total revenue areas, 49
elastic situation, 41
elasticity of unity, 41
inelastic situation, 41
methods of computing,
ratio between percentage changes in price and quantity, 52
rise or fall in *TR*, 41–42
Price,
first law of, 28
normal, 152, 155–158
stability, 186–188
Prices,
ownership, 244

Prices—(*Continued*)
subjective, 5, 29
transaction, 5, 79–80, 244
use, 243–245
Producer's surplus, 185, 365
Product differentiation, 123, 177
Product surplus, 339–341
Production function, 259
Production policies, 140–141
how much to produce, 141
how to meet demand changes, 141
how to produce, 142–144
what to produce, 140
when to produce, 144–148
Profit motive, 84
Profits,
accounting, 355, 361, 438–443
business, 471
differences in
among firms, 440–451
among industries, 454–457
at different places, 456–457
at different times, 452–454
dividends and, 439, 451–452
efficiency and, 444–447
elasticity of supply and, 460–469
equity and, 449, 462, 466–469
inventions and, 453–454
lucky, 449–451
monopoly, 451
normal, 150–157, 440
for equity capital, 469–471
under monopolistic competition, 157
under pure competition, 158
see also *Bulk-line cost*
owner, 451–452
pure, 471–473
rate of, 441–443
regional, 456–457
risk and, 471–474
social problems and, 473–477
tax policies and, 474–477
Progress and individual freedom, 476–477
Propensity to consume, 20
Proportionality, law of, 297–299
Purchasing power, 243, 372

Quantitative economics, 229–237
Quasi-rents, 365, 370

Rent,
definitions of, 330–331
demand bids, influence on, 333–336
determined by commodity prices, 366
determines commodity prices, 367
as differential surplus, 342
economic, 355
as *ex ante* income surplus, 335
fluctuations in, 348–351
imputed, 331, 364
Rent—(*Continued*)
normal,
long-run, 344–345, 365
percentage, 345–347
short-run, 345–347
as product surplus, 339–344
quasi-, 365, 370
resemblance to interest, 365, 376
resemblance to profit, 365
Ricardian theory of, 339–344
as royalties, 351–352
supply prices and, 331, 337
true, 365
Resale price maintenance, 88
Repeated use good, 221
Reproduction cost, 355, 357–358, 365
Reservation price, 77, 79, 83, 85
Returns,
decreasing, 102, 205
diminishing, 102–104, 205, 296–299
Ricardian rent theory of, 339–342
increasing, 102, 205
Revenue,
average, 44, 47
factor, 263, 271–272, 378
marginal (product), 42–49, 143–144, 271
total, 40–44, 48–49
Rigid prices, 171–172, 186–188
Risk,
factor in interest rates, 401–402, 415–416, 418
objective, 402
subjective, 401, 409
Risk premium,
element in interest, 402, 418
element in profits, 471–474
Royalties, 351–352

Saving,
collective, 429
distinct from lending, 422
influenced by interest rate, 426–427
and lending, 394–396
liquidity preference and, 421–423
motives for, 394–396, 427–429
propensity to consume and, 419, 433
as source of loans, 425
time preference and, 419–421
Seller's surplus, 184–185, 355
Selling cost, 118–123
Service prices, 241
Substitution,
effect, 71–73
of factors, 142–144, 254, 274–275, 289–290, 301–302, 325–327
of goods, 19–20, 66–67, 125–126
Subtractional costs, 214
Superior goods, 22, 37, 69–73
Supply,
income-elasticity of, 247
long-run, 192–194

Supply—(*Continued*)
in market sense, 78
market time, 194–196
moderately long-run, 196–203
price, 77, 79, 83–93
price-elasticity of,
cross, 247
directional, 199–200
time, 192–194
quantity of, 76, 81
schedule, 77
in schedule sense, 78–79
short-run, 194–196
very long-run, 203–208
Supply price situations,
customary price, 86–87
imitative price, 88–89
legal maxima or minima, 87–88
markup added to cost, 89–92
maximum profit calculations, 93–116
no reservation price, 85–86
resale price maintenance, 88
Surplus,
buyer's, 184–185
consumer's, 184–185
differential, 339–344
economic, 355, 357–358, 361–364
elements and tax policy, 361–364
elements in interest, 360
elements in profits, 360–361
elements in wages, 359–360
income, 334–336, 344, 356, 368, 378
lessor's, 355–357
opportunity cost, 355, 358–359, 367–369
producer's, 185, 365
product, 339–344
residual, 355, 356–357
seller's, 184–185, 355

Tangents,
rule for finding *MR* at a point, 45–46
steepness of, 50, 55
Taxation,
of personal incomes, 363–364
of profits, 362, 474–476
of rent income (theory of Henry George), 361–362
of surpluses, 361–364
Theory of expectations, 147–148
Time-elasticity,
of demand for labor, 262–263
of supply of goods, 192–194, 220, 223–225
of supply of labor, 258
Time preference, 419–421, 431
Time trend, 231–233
Transactions price, 5, 79–80, 244

Underpayments, 447–448
Unearned incomes, 314–315
Unemployment, 319–329
cyclical, 328–329
due to declining demand for product, 320–321, 327–328
individual causes of, 319–320
invention and, 325–327
prices of substitute factors and, 323, 324, 325, 327–328
rigid wages and, 321–323
wage cuts and, 321–323
wage increases and, 259–263, 324
Unions, labor,
effect of, on wages, 254–255, 261–262, 289–291, 303–306
increase efficiency, 290, 305–306
indirect effects of, 305–306
objectives of, 291
oppose use of substitutes, 289–291, 299, 305
Use prices, 243–245

Vacancies, 338–339
Valuation, 441–443

Wage concepts,
cuts, 264–267
equilibrium wage rates, 269–293
increases, 256–263
minimum wages, 255
money wages, set by
bargaining, 253–254, 268
employer, 252–253
government, 255
organized workers, 254, 261–264, 289–291
non-wage advantages, 256, 263–264
overtime, 257
wage differences due to
bargaining, 287–289
cost of production, 285
demand for labor, 282–283
individual factors, 291–292
marginal productivity, 285–287
skill, 284–285, 287
supply of labor, 283–285
Wages, real,
capital accumulation and, 299–301
determinants of, 242
differences in, between regions, 309–310
economic parasites and, 313–316
income distribution and, 316
inventions and, 301–303
population (labor supply) and, 298, 306–308
statistics of, 308–309
unearned incomes and, 314–317
union strength and, 303–306
war and, 311–312
war preparation and, 314