UNIVERSAL
LIBRARY

OU_156090

UNIVERSAL
LIBRARY

OUP—23—4-4-69—5,000.

OSMANIA UNIVERSITY LIBRARY

Call No. 330/F16E Accession No. P.C. 594

Author Fairchild, F.R

Title Economics

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

Economics



THE MACMILLAN COMPANY
NEW YORK · BOSTON · CHICAGO · DALLAS
ATLANTA · SAN FRANCISCO

MACMILLAN AND CO., LIMITED
LONDON · BOMBAY · CALCUTTA · MADRAS
MELBOURNE

THE MACMILLAN COMPANY
OF CANADA, LIMITED
TORONTO

ECONOMICS

by *Fred Rogers Fairchild*
KNOX PROFESSOR OF ECONOMICS

Edgar Stevenson Furniss
PELATIAH PERIT PROFESSOR OF POLITICAL AND SOCIAL SCIENCE

Norman Sydney Buck
PROFESSOR OF ECONOMICS

Yale University

Revised Edition

New York

THE MACMILLAN COMPANY

REVISED EDITION COPYRIGHTED, 1940,
BY THE MACMILLAN COMPANY

ALL RIGHTS RESERVED—NO PART OF THIS BOOK MAY BE REPRODUCED IN ANY FORM WITHOUT PERMISSION IN WRITING FROM THE PUBLISHER, EXCEPT BY A REVIEWER WHO WISHES TO QUOTE BRIEF PASSAGES IN CONNECTION WITH A REVIEW WRITTEN FOR INCLUSION IN MAGAZINE OR NEWSPAPER

Printed in the United States of America

Published May, 1940
Seventh Printing February, 1947

One volume edition originally copyrighted and published, 1937,
By The Macmillan Company

Preface

Many years' experience in teaching economics to beginners at Yale University has taught us the importance of the textbook and has led us to certain definite convictions as to the type of text best calculated to serve the purpose of helping college students to a knowledge of the fundamental facts and principles of the economic world in which they live. That is the sole purpose of this book. It is not written for the edification of the mature student of economics. It is strictly a book for beginners and the teachers of beginners.

Adherence to this purpose has indicated certain rules for our guidance and imposed certain restrictions upon us. As a rule the treatment has been confined to topics upon which the science of economics may fairly be said to have reached definite conclusions. Occasionally, when this rule could not be followed, we have stated — somewhat dogmatically perhaps — what we consider sound conclusions as to subjects upon which there may still be real difference of opinion among economists of authority. In a few cases we have undertaken an impartial summary of considerations upon the two sides of some unsettled current issue, but as a rule we do not believe that the beginning student profits greatly from that type of discussion which merely presents arguments on both sides of a controversial question and leaves him to draw his own conclusions. Such matters belong in the more advanced courses. There is plenty of material to occupy fully the elementary course in economics without penetrating far into those frontier regions where economic science has not yet reached fairly definite conclusions.

This limitation of the field makes easier the strictly scientific treatment to which we have sought to adhere. Our purpose is to present the facts and the principles of economics — to disclose the truth. With the ethical or moral problems which arise upon this economic foundation we are not concerned, except only as prevailing judgments upon such matters may themselves become the subject matter of scientific economic inquiry. Our purpose is to aid the student to discover how things are, not how they ought to be. We have no case to establish, and we offer no propaganda.

Addressed as it is to beginning students, this is essentially an elementary book. We have sought to spare the student the confusion that comes from taking for granted knowledge which he cannot fairly be assumed to possess, and the treatment is correspondingly elementary. We have sought, by giving care to such things as definition of technical terms, sequence of topics, and simplicity of style, to enable the beginning student to follow the discussion without undue difficulty. It is not to be denied that certain parts of the subject of economics are difficult. We have made no attempt to escape such difficulties, either by avoiding the hard topic or glossing it over with a superficial discussion. Rather have we relied upon simple and elementary treatment to remove unnecessary causes of confusion, in the belief that so presented these topics will not prove beyond the capacity of the student.

The plan of the book has dictated the continuous combination of theoretical analysis with historical narrative and discussion of practical problems. It is our conviction that this is a sound pedagogical rule for an elementary treatise. Theory must be illustrated and justified by showing its relation to practical affairs; facts must be made significant and interesting by showing their conformity to general principles. "Theory" and "practice" should, we believe, go always hand in hand.

It is today quite generally recognized that the average college student does not possess that knowledge of the economic environment which is necessary as a foundation for the study of economic principles and problems. A section (Chapters I–V) devoted to the study of the modern economic organization and its development, stressing the coöperative nature of modern society and the importance of price as a motivating force, is therefore made an integral part of this book.

The exercises which will be found at various points do not conform to any rigid rule. They have been introduced only at those points where it was felt that they would contribute substantial aid to the student in his effort to grasp the principles expounded in the preceding text or would give useful training in economic reasoning. These problems are obviously illustrative rather than comprehensive, and the teacher can readily modify or expand them or ask the student himself to construct similar ones.

At the end of the book will be found a classified list of books which are suggested for the student's further reading. This list is of course not comprehensive, and it does not embrace works of a highly advanced or specialized or technical nature. It is intended merely to present for the choice of the student certain books which he will find interesting and profitable reading in connection with his first study of economics.

These in brief are the more important principles which have guided us in harmony with our aim to give to those students who may not pursue the study of economics further a broad knowledge and understanding of the economic world of today and to others a firm foundation for their further study of economics and allied subjects. It is in such matters as are here noted that our contribution, if any, is to be found.

While we have had in mind predominantly the needs of college students, we are nevertheless bold enough to believe

that the requirements thus imposed upon us will have made this book useful also to those more mature readers who, for any reason whatever, may have need of a short elementary exposition of the fundamental facts and principles of economics in the light of present-day knowledge.

The present book is a shorter edition of our two-volume *Elementary Economics*. It has been written in response to the demand for an elementary textbook in economics similar in character to the longer work but of length suitable for use in half-year courses or in those full-year courses which use a general foundation text followed by selected special books. As regards purpose, balance of topics, plan, and style, the two books are generally alike.

The first edition of this book appeared three years ago (1937), at a time when there was observable an inclination to view the extraordinary events of the last decade as in some way casting suspicion upon the generally accepted principles of economics, with the inference that these principles must now be thrown in the discard and replaced by a "new economics." The prosperous years of 1922–1929 were regarded by not a few persons as having ushered in a "new era," with poverty abolished and the human race lifted to a level of prosperity such as the mind of man had not before conceived. Economic science, casting doubts upon the validity of this roseate dream, was to be put aside in favor of the "new economics" of perpetual prosperity. More recently the sad experiences of adversity have brought forth another "new economics," the economics of depression, rationalizing for example permanent unemployment and permanent government relief on something like the present scale. It is our belief that a less superficial view would disclose the prevailing body of economic principles as generally competent to explain the economic events of recent years, whether of prosperity or of depression. We are furthermore of the opinion that only in the light of such fundamental economic principles may there be any true analysis and judgment of the "New

Deal" and other recent developments. It is in this spirit that the present revision has been prepared. While the formulation of basic economic principles has undergone little change, these principles find fresh illustration in a wealth of new facts and are called upon to elucidate the economic problems, new and old, which confront the world today.

The following are some of the more important changes in the present edition. The whole section on price has been reorganized and largely rewritten in order to incorporate the concept of imperfect competition as an integral part of the theory of price. Imperfect competition and pure competition, as well as monopoly, have been merged in a unified treatment, which we believe not only is an advance in theoretical accuracy but serves also to make the treatment more realistic and practical. In the section on money and banking, some clarification in the statement of principles has been found possible, and the material has been rearranged in such way as to avoid repetitions and merge the recent complicated changes in American monetary and banking legislation into the treatment of the subject as a whole, in a manner that was scarcely possible three years ago. Recognition has been given to recent advances in the theory and practice of "managed money," stabilization funds, etc. The chapter on nationalism and the tariff has been enlarged by the introduction of some discussion of recent official restrictions upon international trade and the problem of reciprocal trade regulation, including Secretary Hull's trade agreement program. The discussion of labor problems has been revised to take account of recent governmental policies affecting wages, hours, conditions of work, and the relations of employers and employees, including a realistic discussion of social security legislation. A substantial amount of new material has been added to the chapters on radical reform proposals, including an impartial description and appraisal of socialism, Russian communism, and Italian and German Fascism, combining historical and theoretical treatment. Finally the entire book has been

brought up to date with respect to statistics and all statements of fact.

We make grateful acknowledgment of the invaluable assistance which we have derived from our colleagues in the Economics Department of Yale University. Though the actual task of authorship has fallen to the three undersigned, this book represents, especially in its first edition, the combined contributions, direct and indirect, of a large group of teachers who for many years have been closely associated in the task of teaching elementary economics to the students of Yale University. To all of these, past as well as present colleagues, we acknowledge our debt.

In particular we are pleased to record our indebtedness to Professor Irving Fisher, whose influence upon the authors, his former students, has been profound; likewise to Professor Clive Day, who has given valuable suggestions in the field of economic history, to Professor Winthrop M. Daniels for assistance in the preparation of the chapters on railroads, to Professor Ray B. Westerfield for contribution of valuable material on modern banking practice and on the subject of risk and insurance, to Dr. Benjamin P. Whitaker for aid in gathering material illustrating the principles of price, to Professors A. J. Nichol and R. M. Bissell, Jr., who contributed much to the last two revisions of the chapters on price, to Mr. C. H. Whelden, Jr., for assembling materials on the "New Deal" and other recent developments in applied economics, to Professor Ralph C. Jones and Mr. J. H. G. Pierson for suggestions in the fields of accounting and monetary theory respectively, and to Professor I. R. Barnes for advice on the public regulation of business.

We are indebted also for helpful criticism and suggestions from a host of teachers who have used this book in other institutions, among whom special acknowledgment is due to Lieutenant Colonel Herman Beukema of the United States Military Academy and to Professor Clair W. Swonger of the University of New Hampshire, who has given the exercises a

thorough overhauling for this edition. General acknowledgment is finally due the authors of the standard works on many subjects embraced within the broad field of general economics, upon whose writings we have leaned beyond the possibility of specific acknowledgment in every case.

FRED ROGERS FAIRCHILD
EDGAR STEVENSON FURNISS
NORMAN SYDNEY BUCK

NEW HAVEN, CONNECTICUT
April, 1940

Contents

PAGE

CHAPTER I. FOUNDATIONS OF THE SCIENCE OF ECONOMICS 3

1. Human wants: (1) Number and variety. (2) Food, clothing, and shelter. (3) Other wants. (4) Intermediate goods. (5) No limit to the wants of man
2. Nature, the provider
3. Economics: (1) Definition. (2) Economics as a science
4. Utility
5. Wealth: (1) Definition. (2) Are human beings wealth? (3) How to identify wealth
6. Income: (1) General. (2) Costs. (3) Net income
7. Wealth and income. Fund and flow
8. Property: (1) Definition. (2) Division of property. (3) Property in free persons. (4) Wealth and property. (5) Documents in evidence of property
9. Value: (1) A common denominator. (2) Transfer and exchange. (3) Barter and money exchange. (4) Definition of value. (5) Definition of price. (6) Value and price. (7) The unit of value
10. Exercises

CHAPTER II. THE FACTORS OF PRODUCTION . . 22

1. Production
2. The factors of production
3. Land
4. Labor
5. Capital: (1) Description. (2) Capital in production. (3) The accumulation of capital. Saving. (4) The form of capital instruments. Invention
6. The entrepreneur
7. Exercises

CHAPTER III. THE FORM OF THE BUSINESS UNIT. . 34

1. The individual proprietorship
2. The partnership

PAGE

3. The corporation
   (1) Modern growth
   (2) Nature of the corporation
   (3) Capital stock
   (4) Bonds and notes
   (5) Stocks and bonds compared
   (6) The property account of a corporation: *a.* Assets and liabilities. *b.* Losses and insolvency. *c.* The value of stock. *d.* Interest payments and dividends. *e.* Stock dividends. *f.* Watered stock
   (7) The corporation's income account
4. Exercises

CHAPTER IV. LARGE-SCALE PRODUCTION AND COMBINATION 57

1. The growth of large-scale production
2. Conditions necessary for large-scale production
3. Economies of large-scale production: (1) Power and machinery. (2) By-products. (3) Technical experiment and research. (4) Economy in administration
4. The field of large-scale production and its limits
5. Combination
   (1) Introduction
   (2) The large corporation
   (3) The formation of combinations
   (4) Motives leading to the formation of combinations
   (5) Horizontal combinations: *a.* Definition. *b.* Economies in buying and selling. *c.* Economies in operation. *d.* Saving in freight costs
   (6) Vertical combinations: *a.* Definition. *b.* Technical advantages. *c.* The United States Steel Corporation
   (7) Conclusions

CHAPTER V. DIVISION OF LABOR AND THE CONTROL OF PRODUCTION 72

1. Coöperation
2. Division of labor
3. Economic effects of division of labor
4. Division of labor and the extent of the market
5. Territorial division of labor
6. Division of labor a form of coöperation

PAGE

7. The control of production
   (1) Conscious and unconscious control
   (2) Mercantilism
   (3) *Laissez faire: a.* Reaction from government control. *b.* Economic freedom. *c.* The motive to productive effort. *d.* The rule of price
8. Taking stock of our economic organization
9. Shall the system be scrapped? (1) What the present system accomplishes. (2) The present conclusion

CHAPTER VI. PRICE AND THE CONSUMER . . . 92

1. The rôle of price
2. A pecuniary system
3. The behavior of prices: (1) A bird's-eye view. (2) Prices fixed by statute or custom. (3) Are sellers or buyers free to fix prices? (4) Is price a mystery?
4. Consumer's choices
5. Marginal utility: (1) The subjective side of utility. (2) Relative marginal utility. (3) Definition of marginal utility. (4) Graphical representation, a digression. (5) Meaning of marginal utility. (6) The law of marginal utility. (7) Some apparent price anomalies resolved. (8) Abundance and scarcity
6. Exercises

CHAPTER VII. DEMAND 112

1. Individual demand: (1) Marginal utility and demand. (2) Significance of this analysis. (3) How both parties to a trade can gain. (4) Real people. (5) Is all buying irrational?
2. Total demand: (1) Individual demand and total demand. (2) Definition of demand. (3) Time and place of the demand schedule. (4) Demand and desire
3. The law of demand: (1) General statement. (2) Special cases
4. Elasticity of demand: (1) General statement. (2) The limits of elasticity
5. Exercises

CHAPTER VIII. COST OF PRODUCTION 127

1. The different viewpoints of buyers and sellers
2. Production to order or for a general market
3. Immediate sales problem: (1) General statement. (2) Reservation price

PAGE

4. Remote sales problem
   (1) Price governed by cost
   (2) Future costs control
   (3) Analysis of cost
      a. Fixed and variable costs
      b. Economy of fully utilized plant
      c. Further expansion of output and unit cost
      d. The limit to large-scale enterprise
      e. Marginal cost: (*a*) Definition. (*b*) Marginal cost and average cost
      f. Relation of costs to output
5. Exercises

CHAPTER IX. COMPETITION AND THE MARKET . . 143

1. The market: (1) Definition. (2) Free market. (3) Setting the price
2. Competition: (1) Definition. (2) Competition and monopoly. Degrees of competition
3. Pure competition: (1) Market conditions. Definition. (2) The individual seller meets perfect elasticity of demand. (3) The seller has no "price problem"
4. Imperfect competition: (1) Inertia and ignorance. (2) Individuality of products. (3) Trade marks and special services. (4) Small number of sellers. (5) Demand not perfectly elastic. (6) Definition of imperfect competition. (7) The liberty of imperfect competition. (8) Limitations upon the seller's liberty. (9) Liberty and discretion in big business. (10) Quality competition in place of price competition
5. Advertising
6. Exercises

CHAPTER X. COST OF PRODUCTION AND SUPPLY UNDER PURE COMPETITION 159

1. Pure and imperfect competition
2. Cost of production and production policy: (1) Price above average cost. (2) Price below average cost. (3) Conclusion
3. Individual supply
4. Total supply: (1) Definition. (2) Supply and stock
5. Relation of quantity to price
   (1) Long and short run
   (2) Increasing costs: *a.* Condition. *b.* Diminishing returns

PAGE

(3) Constant costs
(4) No supply curve slopes downward
(5) Elasticity of supply
(6) The law of supply
(7) Long run and short run again
6. Decreasing cost industries
7. Exercises

CHAPTER XI. DETERMINATION OF PRICE UNDER PURE COMPETITION 173

1. Determination of price and quantity exchanged. Equilibrium: (1) Price determined by demand and supply. (2) Determination of the quantity exchanged. Clearing the market. (3) The process of adjustment. (4) The laws of price equilibrium
2. The causes of price change: (1) In general. (2) Demand and supply the sole determinants. (3) The retail egg market. (4) The basic law of price change. (5) Effect of perfect inelasticity or elasticity of demand or supply
3. Decreasing costs
4. Exercises

CHAPTER XII. COST OF PRODUCTION AND THE SELLER'S POLICY UNDER IMPERFECT COMPETITION . . 189

1. The seller's problem
2. The maximum profit
3. Marginal revenue
4. Marginal revenue and marginal cost: (1) In general. Price above average cost. (2) Price below average cost. (3) Conclusion
5. Influence of elasticity of demand
6. Ascertaining the facts of demand
7. Exercises

CHAPTER XIII. DETERMINATION OF PRICE UNDER IMPERFECT COMPETITION 202

1. The mechanism of price determination
2. Interaction of sellers' policies
3. The rôle of substitutes: (1) Their significance. (2) Influence of substitutes on size of demand. (3) Influence of substitutes on elasticity of demand
4. Price determination in the short run: (1) Price cutting. (2) Limitations. (3) Conclusions

PAGE

5. Price determination in the long run
6. Ascertaining production facts
7. Monopoly: (1) Imperfect competition and monopoly. (2) Monopoly and the public interest
8. Exercises

CHAPTER XIV. INTERRELATIONSHIP OF PRICES. PRICE THEORY IN PRACTICE 222

1. Interrelationship of prices
2. The present value of future income: (1) Capital and income. (2) Property and capital. (3) Present worth. (4) Present goods more desirable than future goods. (5) Interest. (6) The link between future and present values. (7) The discount process. (8) Capital and income. (9) Discount and the laws of price. (10) Discounted income and cost of production
3. Prices of the same good in different markets
4. Prices of competing goods
5. Prices of complementary goods
6. Prices of competing cost goods
7. Prices of joint cost goods
8. Prices of "tandem" goods
9. Price theory in practice: (1) Practical aspects of the theory of value. (2) The overworked term "value"
10. Exercises

CHAPTER XV. ECONOMIC RENT: THE INCOME OF LAND-OWNERS 246

1. The meaning of distribution
2. Definition of economic rent
3. The specific income from land
4. Diminishing returns
5. Market measurement of economic rent
6. Producer's surplus
7. Explicit and implicit economic rent
8. Lands of different quality
9. Margin of cultivation
10. Alternative uses of land: (1) Agrarian lands. (2) Urban lands
11. The value of land
12. Relation of rent to prices
13. Quasi rent

PAGE

14. Economic rent distinguished from other income
15. "Unearned increment"
16. Exercises

CHAPTER XVI. INTEREST: THE INCOME OF OWNERS OF CAPITAL 267

1. Time preference and capital formation
2. Nature of the loan contract
3. The loan market and the rate of interest
   (1) Definition
   (2) The supply of loan funds: *a.* Reaction of lenders. *b.* Corporate saving
   (3) The demand for loans
      a. Borrowing for consumption: (*a*) General description. (*b*) Consumption borrowing determined by time preference. (*c*) Borrowing by governments
      b. Borrowing for production: (*a*) Purpose of the borrower. (*b*) The productivity of capital
      c. Total demand for loans
   (4) Determination of the market rate of interest
4. Adjustment of the individual to the loan market
5. Lack of uniformity in interest rates
6. True and nominal interest rates
7. Changes in the rate of interest
8. The capitalist class

CHAPTER XVII. WAGES: THE INCOME OF LABOR . . 285

1. Economic stratification of labor and of wages
2. The demand for labor: (1) Why employers hire laborers. (2) Marginal productivity of labor. (3) Productivity discounted
3. The supply of labor
4. The general law of wages
5. Practical qualifications
   (1) Theoretical precision lacking
   (2) Interdependence of different types of labor
   (3) The sphere of bargaining: *a.* In general. *b.* Wages above marginal product. *c.* Wages below marginal product
   (4) Differences in wage rates within occupational groups
   (5) Local variations in wage rates
6. Discounted marginal productivity is fundamental

PAGE

7. Other types of labor income
8. Wages and the contrast of social classes
9. Labor's share of the social income. The law of proportionality
10. Exercises

CHAPTER XVIII. PROFITS: THE INCOME OF THE RISK BEARERS 306

1. The accountant's profit statement
2. Implicit rent, interest, and wages of management
3. Definition of profits
4. Actual profits and losses
5. The risk-bearing function: (1) Business risks. (2) Risk bearing and its reward. (3) Risk bearing not always a productive service
6. Profits tend to be short-lived
7. Profits not a price determinant
8. Criticism of the profit system
9. Exercises

CHAPTER XIX. PERSONAL DISTRIBUTION . . . 317

1. Functional shares in distribution. Summary
2. Total and average national incomes
3. Income distribution in America
4. Distribution of wealth in America
5. Interrelation of wealth and income distribution
6. Effect of inheritance
7. The problem of poverty: (1) Conditions. (2) The family unit. (3) The "standard of living." (4) Family incomes. (5) Causes of poverty and the remedies
8. The problem of inequality
9. Exercises

CHAPTER XX. PRINCIPLES OF MONEY 337

1. Nature of money: (1) Some historical examples. (2) The common characteristic. (3) Definition of money
2. Functions of money
3. Government control of money: (1) In general. (2) Legal tender
4. Money and currency
5. Varieties of money and monetary systems (definitions and general description)
6. Standard money systems

PAGE

(1) Principal kinds of money
(2) The ideal money commodity
(3) Types of standard money systems
(4) The gold standard: *a.* The monetary unit. *b.* Free coinage. *c.* The value relation between gold and money
(5) The bimetallic standard
(6) The gold bullion standard
(7) The gold exchange standard

7. Irredeemable money systems
   (1) Origin and essential features
   (2) General abandonment of the gold standard
   (3) Recent American monetary experience: *a.* Abandonment of the gold standard. *b.* Devaluation of the dollar. *c.* Gold reserve act and proclamation of 1934. *d.* The new monetary unit. Legal tender. *e.* "Gold profit" and stabilization fund. *f.* Silver policy. *g.* Summary. The present American monetary system
8. Exercises

CHAPTER XXI. PRINCIPLES OF BANKING . . . 365

1. The nature of credit: (1) Financing production. (2) Credit
2. The commercial bank
3. Lending: (1) Example of borrowing. (2) Interest and discount. (3) Proceeds. (4) Principal types of trade paper. (5) A bank's balance sheet. (6) Discounting a note
4. A typical bank statement
5. Bank deposits
   (1) Definition
   (2) Discount and deposit
   (3) The check
      a. Definition
      b. Payment by check
         (*a*) Within the bank
         (*b*) Collection between banks: In general. Clearing house. Clearing for outside banks. The volume of clearings. Out-of-town checks
6. Paying a loan
7. Note issue
   (1) Notes and deposits: *a.* Fundamental similarity. *b.* Differences between notes and deposits
   (2) Legal restrictions

PAGE

8. The normal business of banking
9. The reserve
   (1) Its function
   (2) How the reserve ratio is controlled: *a.* Buying and selling securities. *b.* Rediscounting and borrowing. *c.* Calling loans. *d.* Control by the rate of discount
10. The problem of elasticity: (1) Need of elastic currency. (2) Elasticity of bank credit
11. The function of credit
12. Peculiar importance of bank credit
13. Exercises

CHAPTER XXII. THE AMERICAN BANKING SYSTEM . 397

1. Historical bankground
2. Banking in the United States before the Civil War
3. The national banking system: (1) Origin. (2) National bank notes. (3) The problem of reserves
4. The Federal Reserve
   (1) Organization: *a.* Board of Governors. *b.* Federal reserve districts and banks. *c.* Member banks
   (2) Functions of the Federal Reserve
      a. Rediscounting and discounting
      b. Federal reserve notes: (*a*) General description. (*b*) Federal reserve bank notes. End of the national bank notes
      c. Reserves against deposits
      d. Open-market operations
      e. Centralized reserves
      f. Mobile reserves
      g. A national system of clearing
      h. The government's financial business
5. The American banking system in 1929: (1) Introduction. (2) Composition of the banking system. (3) Branch banking and chain banking. (4) Weaknesses of decentralized banking. (5) Affiliates. (6) Expansion of credit
6. The financial crises of 1929 and 1931
7. Bank failures

CHAPTER XXIII. AMERICAN BANKING SINCE 1929 . . 418

1. Banking legislation since 1929
2. Pre-"New Deal" remedies: (1) National Credit Corporation

PAGE

and Reconstruction Finance Corporation. (2) Glass-Steagall act
3. Banking collapse and government action
4. Provisions for meeting future panics
5. Structural reforms of 1933 and 1935
   (1) Deposit insurance: *a.* Description. *b.* Judgment of deposit insurance
   (2) Capital requirements
   (3) Branch banking
   (4) Holding company affiliates
   (5) Divorce of commercial and investment banking
6. Functional remedies of 1933 and 1935
   (1) Interest on deposits
   (2) Speculative uses of credit
   (3) Centralized control of the banking system
   (4) Control of credit: *a.* Concept of credit control. *b.* Open-market operations. *c.* Regulation of reserve ratios of member banks. *d.* Loans on real estate
7. The present banking situation
8. Other forms of government banking: (1) Agricultural credit agencies. (2) Loans to home owners
9. Criticism of government banking

CHAPTER XXIV. MONEY AND PRICES 437

1. The value of money
2. Value of money not constant
3. Measuring the value of money
4. Index numbers: (1) Price relatives. (2) The average, simple and weighted. (3) An index number of the aggregative type. (4) Practical index numbers
5. The problem of the price level
   (1) Introduction
   (2) Volume of trade
   (3) Quantity of currency
   (4) Velocity of circulation
   (5) The quantity theory of money: *a.* Algebraic relations. *b.* Significance of the quantity theory. *c.* Practical limitations of the quantity theory
6. Far-reaching effects of monetary fluctuations
7. Monetary reform
8. Managed money
9. Exercises

PAGE

CHAPTER XXV. PRINCIPLES OF GOVERNMENT REVENUE 460

1. Government and the economic organization
2. The functions of government
3. Government income
4. Industrial earnings
5. Financing the primary functions of government
   (1) General character of revenue
   (2) Characteristics of the principal forms of revenue: *a.* Taxes. *b.* Fees. *c.* Special assessments. *d.* Names may be misleading
6. Distribution of the tax burden
   (1) Equal distribution
   (2) Benefit or cost
   (3) Ability to pay
      a. Meaning
      b. Measurement of ability to pay: (*a*) General statement. (*b*) Proportional taxation. (*c*) Progressive taxation. (*d*) Regressive taxation
      c. The problem of taxing according to ability
      d. Another concept of progression
   (4) Benefit, cost, and ability
   (5) Escape from taxation: *a.* Lawful adjustments. *b.* Tax evasion
   (6) The shifting of taxes: *a.* Meaning. *b.* Direct and indirect taxes
7. The purpose of taxation

CHAPTER XXVI. TAXATION 480

1. The property tax
   (1) Evolution
   (2) The American property tax
      a. General description
      b. Position of the property tax
      c. Assessment: (*a*) The problem. (*b*) Expert assessment. (*c*) The ordinary local assessment. (*d*) Practical results
      d. Confusing wealth and property
      e. Taxation of corporations
      f. Reform of the property tax
2. The income tax
   (1) Historical background
   (2) The United States income tax: *a.* General description.

PAGE

b. Individual income tax base. *c.* Corporation dividends. *d.* Undistributed income tax. *e.* Progressive taxation of corporations. *f.* Income tax rates. *g.* Elasticity of the income tax

3. Shifting and incidence of taxes on property and income: (1) Taxes on economic rent or land value. (2) Capitalization of taxes. (3) Taxes on dwelling houses or their rent. (4) Taxes on business property or income
4. The poll tax
5. Consumption taxes
   (1) Introduction and classification
   (2) Excise taxes: *a.* General account. *b.* Shifting and incidence of excise taxes
   (3) The customs tariff: *a.* Economic effects of the tariff. *b.* The revenue tariff. *c.* Fiscal weakness of the protective tariff
6. Death taxes: (1) General description. (2) Some fundamental questions

CHAPTER XXVII. PUBLIC FINANCE 507

1. The cost of government
   (1) Cost of government and national income
   (2) Growth of public expenditure: *a.* General summary. *b.* United States federal government. *c.* All grades of government in the United States
   (3) Functional analysis of government costs in the United States: *a.* Federal government. *b.* State and local governments
2. The American revenue system: (1) Federal government. (2) Revenue system of the American states. (3) Revenues of American local governments
3. Public borrowing: (1) Purposes. (2) The security back of public debts. (3) High standing of government obligations. (4) The national debt of the United States
4. The budget
5. American financial policy during the past decade (1929–1939): (1) General character. (2) The double budget. (3) Expenditures, revenues, and deficits. (4) Progressive increase in cost of government

CHAPTER XXVIII. THE BUSINESS CYCLE . . . 531

1. Business fluctuations
   (1) In general

PAGE

(2) Types of economic fluctuations
(3) Independent irregularities
(4) Cyclical fluctuations: *a.* Description and classification. *b.* Seasonal cycles. *c.* Long-wave movements. *d.* Major and minor cycles
(5) The secular trend
2. The business cycle: (1) General description. (2) Phases of the business cycle
3. Revival
4. Revival breeds prosperity
5. Limiting factors in prosperity
6. Recession and depression
7. Forces initiating the business cycle
8. Consequences of cyclical fluctuation
9. Control of cyclical changes: (1) Conditions of the problem. (2) Flexible prices. (3) Control of wage rates. (4) The rate of interest. (5) Public works

CHAPTER XXIX. MONOPOLISTIC INDUSTRY AND GOVERNMENT INDUSTRY 557

1. The nature of monopoly
2. Capitalistic monopoly
3. Monopolistic industry: (1) The problem. (2) Historical. The trust movement
4. Anti-trust legislation in the United States
(1) Introductory statement
(2) The Sherman anti-trust act of 1890: *a.* Abstract of principal features. *b.* Operation of the Sherman act
(3) The trade commission act of 1914: *a.* Investigation of corporations. *b.* Advisory functions. *c.* Unfair competition. *d.* The amendment of 1938
(4) The Clayton act (1914) and the Robinson-Patman act (1936)
(5) Remedies provided by the anti-trust acts
5. Achievements of the United States anti-trust program
6. Is monopoly inevitable? (1) Statement of the question. (2) Forces favoring monopoly. (3) Limits to monopoly
7. Reasons for moderate results of the anti-trust program: (1) Standards set by legislation. (2) The task of administration. (3) Influence of patent laws. (4) Contradictory laws
8. Conclusions

PAGE

9. Government in industry
   (1) The attitude toward government industry
   (2) Financial results of government industry
   (3) The problem of government industry: *a.* Interest and profits. *b.* Profits and local public utilities. *c.* Corporation abuses. *d.* Politics. *e.* The motive in business. *f.* Conclusion

CHAPTER XXX. RAILROAD TRANSPORTATION . . 582

1. The railroads and the public
2. The railway problem
3. Railway expenses: (1) Fixed costs. Interest. (2) Other fixed costs and variable costs. (3) Decreasing costs. (4) Joint costs
4. Railway rates
   (1) Peculiarity
   (2) Charging what the traffic will bear
   (3) Discrimination between goods
   (4) Local discrimination: *a.* Description. *b.* The reason for local discrimination
   (5) Personal discrimination
5. The breakdown of competition
6. Monopoly inevitable and economical
7. Regulated monopoly
8. State regulation of rates: (1) Early history. (2) The "Granger laws." (3) The conflict of jurisdiction
9. Federal regulation
10. The railroad crisis
11. The emergency railroad transportation act of 1933
12. The motor carriers act
13. Labor legislation
14. Conclusions
15. Exercises

CHAPTER XXXI. INTERREGIONAL TRADE . . . 609

1. The bases of interregional trade: (1) Fundamental motive. (2) Local advantages and their causes. (3) The gain from specialization and trade
2. The law of comparative advantage: (1) Statement. (2) Diminishing returns and the law of comparative advantage. (3) Transportation and the law of comparative advantage
3. The control of trade
4. The foreign trade of the United States

PAGE

CHAPTER XXXII. FOREIGN EXCHANGE AND THE INTERNATIONAL BALANCE OF PAYMENTS . . . 621

1. The problem
2. Financing exports
3. Financing imports
4. The rôle of the banks in exchange
5. The exchange market
6. The rate of exchange: (1) General principles. (2) The gold points
7. Indirect exchange
8. The balance of payments
9. The international balance sheet of the United States
10. Limits to the flow of gold
11. Reaction of exchange rates upon imports and exports
12. Mercantilist fallacies
13. Exercises

CHAPTER XXXIII. CONTEMPORARY PROBLEMS OF FOREIGN EXCHANGE 643

1. Pre-war equilibrium
2. Pre-war exchange conditions
3. Irredeemable money and the par of exchange: (1) The problem. (2) The purchasing power par of exchange. (3) Purchasing power par and market rate of exchange
4. Stabilization of exchange rates
5. Dollar exchange
6. Conclusions
7. Exercises

CHAPTER XXXIV. INDUSTRIAL NATIONALISM AND THE PROTECTIVE TARIFF 654

1. International division of labor *vs.* national self-sufficiency: (1) International division of labor. (2) National self-sufficiency. (3) Effects of the protective tariff
2. Protection and national self-sufficiency
   (1) The case for protection
      a. The favorable balance of trade argument
      b. The gold argument
      c. The added industry argument
      d. The wages argument: (*a*) Increasing the demand for labor. (*b*) Making wages high. (*c*) Wages and labor cost
      e. Effect of withdrawing protection

PAGE

f. Alleged saving in transportation costs
g. Meeting the tariffs of other nations
h. Mutual dependence of nations
(2) Considerations against protection: *a.* What is the territorial unit for protection? *b.* Benefits of foreign competition
(3) Conclusion
3. Temporary protection: (1) Infant industry protection. (2) Protection against "dumping"
4. Protection on account of military necessity
5. Protection by bounties
6. Ship subsidies
7. International tariff agreements: (1) Preferential trade and tariff arrangements. (2) Reciprocal trade agreements of the United States

CHAPTER XXXV. LABOR PROBLEMS 675

1. Problems of labor
2. Unemployment
(1) Nature and extent
(2) Types of unemployment: *a.* Basis of classification. *b.* Technological unemployment. *c.* Seasonal unemployment. *d.* Cyclical unemployment
(3) Remedies for unemployment
a. Introduction
b. Greater mobility of labor
c. Public employment offices
d. Greater flexibility of wage rates
e. Stabilizing the demand for labor
(*a*) Voluntary action by industry
(*b*) Public works: Theoretical analysis. Public works program of the United States
3. Child and woman labor: (1) Introduction. (2) Employment statistics. (3) Wages of women and children. (4) Solution of the problem of child and woman labor
4. Industrial accidents
5. Hours of labor: (1) The shorter work day. (2) Hours and productivity

CHAPTER XXXVI. LABOR LEGISLATION 700

1. Unemployment insurance
(1) General nature and origin
(2) The Ghent system

PAGE

(3) Compulsory insurance in Great Britain: *a.* Original plan. *b.* Breakdown and reform. *c.* Lessons of the British experience
(4) Unemployment insurance in the United States
2. Legal regulation of the labor of children: (1) General summary. (2) In the United States
3. Regulation of the labor of adult women
4. Minimum wage laws: (1) General summary. (2) The United States fair labor standards act of 1938. (3) Appraisal of minimum wage laws
5. Workmen's compensation laws: (1) The problem. (2) Common law remedy. (3) Principle of compensation. (4) American laws
6. Security for the aged
(1) Pensions
(2) Old age insurance: *a.* General summary. *b.* United States social security act

CHAPTER XXXVII. LABOR ORGANIZATION IN AMERICA 721

1. Types of American unions: (1) Introduction. (2) The trade union. (3) The industrial union. (4) Conflict of rival types. The C.I.O.
2. Strength of unionism in America
3. Company unions
4. National labor relations act
5. The union program: (1) Incidental benefits. (2) Collective bargaining. (3) The struggle over union recognition. (4) The closed shop
6. Political activity of American unions

CHAPTER XXXVIII. INDUSTRIAL STRIFE 736

1. The strike: (1) General description. (2) Statistics of strikes in the United States. (3) Types of strikes
2. Violence in strikes
3. Legality of the strike
4. Public intervention in strikes
(1) Conciliation: *a.* In general. *b.* Conciliation promoted by government
(2) Mediation
(3) Voluntary arbitration: *a.* In general. *b.* Effectiveness of voluntary arbitration
(4) Transition stages
(5) Compulsory investigation

PAGE

(6) Compulsory arbitration: *a.* The affirmative argument. *b.* Experience in New Zealand. *c.* Compulsory arbitration in Kansas. *d.* Judgment of compulsory arbitration

CHAPTER XXXIX. SOCIALISM AND COMMUNISM . . 753

1. Radical movements
2. Definition of socialism
3. Socialist indictment of capitalism
4. Socialist analysis of capitalism
   (1) Karl Marx
   (2) The class struggle
   (3) Laws of capitalist development: *a.* Tendency toward concentration. *b.* Increasing number and misery of the laboring class. *c.* The rôle of crises
   (4) General appraisal of Marxian doctrine
5. The socialist program
6. Socialist tactics
7. Political action
8. Communism: (1) Doctrine and program. (2) Communist organization in the Soviet Union
9. Socialism in the Soviet Union: (1) Recent history. (2) Planned economy. (3) Economic organization. (4) Growth of soviet industry. (5) Labor and wages. (6) Soviet finance

CHAPTER XL. FASCISM. JUDGMENT OF RADICAL REFORM 781

1. Fascism and communism
2. Fascism in Italy: (1) Party organization. (2) Fascist social organization. The "corporative state." (3) Fascist achievements. (4) Economic welfare of the people under Fascism
3. Fascism in Germany: (1) The Nazi Party. (2) The party and the state. (3) Nazi social and economic program. (4) Government industry. (5) The "estates." (6) Labor organization. (7) Economic results of the Nazi program
4. General appraisal of radical movements: (1) Introduction. (2) Authority versus liberty. (3) Planned economy and efficiency. (4) Planned economy and capital formation
5. Judgment of socialism

SUGGESTIONS FOR FURTHER READING 809

INDEX 819

Economics

CHAPTER I

Foundations of the Science of Economics

Those activities of mankind to which we ascribe the adjective "economic" and whose investigation calls for a *science of economics* are the consequence of two fundamental laws of nature; the one relates to the character of man himself; the other has to do with the natural environment in which man lives.

Human wants: Number and variety. Organically man is an animal, one of the innumerable species of living things that populate the earth. Countless ages of evolution have brought man and all these other forms of life to their present state as separate species and their present equilibrium among themselves as sharers of the earth. It is especially to be noted that the process of evolution, while throwing man into the common struggle for existence, has developed in him certain characteristics which set him in a class apart from all other living things. Such peculiarity appears as soon as we give attention to the things that are desired by men.

At the first stages, it is true, we seem to find nothing to distinguish man from the lower animals. In common with all forms of animal life, man has certain organic needs which must be met or he will die. He must have food and water, and air to breathe. His bodily temperature must be constantly maintained within very narrow limits, and he must, therefore, be protected from extremes of cold and heat by appropriate clothing and shelter. To protect himself against the attack of natural foes, man must have the means of defense or escape. These are fundamental organic needs, and the simplest means appropriate to meet them are called "necessities," employing the narrowest sense of this somewhat overworked term.

But now we come to the difference between man and the lower animals. The latter are satisfied when the necessities of life are met. With man the necessities are only the beginning. No sooner are they met than he reaches out for the satisfaction of other wants, for the "comforts and luxuries." Even the means whereby modern men satisfy the fundamental organic needs have passed far beyond the realm of bare necessities. This attribute of man is so fundamental to every phase of human life, and in particular to man's economic activities, that we must take pains to grasp its full import even at the risk of what may at first sight appear as an elaboration of the obvious.

Food, clothing, and shelter. Food is necessary to prevent starvation, of course; but that is not what the average person is thinking about as he seats himself at the dinner table. He is thinking of the joy of eating. He studies the menu. He rejects oysters and decides on clams. He scans the soup list. He ponders on the choice of roasts: chicken, turkey, beef, lamb; and so on down through the vegetables, salad, dessert, fruit, and coffee. This ceremony is something more than the warding off of starvation. And the same is true, only in lesser degree, of the simplest home table. To the original hunger instinct has been added a host of tastes and desires, so that we demand of food a good deal more than that it keep us alive. The strong desires that have been imposed upon the foundation of the drink-seeking instinct are too well known to need elaboration. Drinking is certainly not confined to the supplying of moisture to keep the body alive.

The necessity of protection from heat and cold gives rise to the desire for clothing. But to the modern civilized man or woman clothing means a good deal more than this. In buying a suit far more attention is given to the pattern, the cut, the style, the fit of the garments than to their cold-resisting qualities. Hats, coats, dresses, shoes, stockings, collars, ties, and so on down the list — how much of all this could be eliminated if all we wanted were protection against the elements!

In most climates some sort of shelter is a necessity, but the dwellings of civilized men, from ancient days down to the present, have been much more than this. From the humble cottage to the royal palace, men have sought to make of their dwellings, not mere shelters from the weather, but things of beauty, of comfort, of ostentatious display, of luxurious enjoyment. One will generally have to search long in the modern village or city to find any habitation that does not represent a craving for something beyond mere shelter. Our houses, like our clothes, are responsive to vanity. They also represent (often grotesquely enough, to be sure) our aesthetic desire for that which is beautiful.

Other wants. Thus we have commenced to list the wants of mankind. We might continue, mentioning for example some of the things that the need of defense requires: strongly built houses, locks on doors and windows, protective weapons, policemen, armies and navies and their equipment, drugs and medicines, hospitals, jails and insane asylums, insurance companies — this is simply the beginning of the list. Consideration of the gregarious tendencies of mankind reminds us of the concentration of people in cities, towns, and villages, thus giving rise to special wants in the field of housing, transportation, etc. Even when people seek relaxation from their ordinary activities and surroundings and take vacations "in the country," they do not generally seek solitude. They require seashore and mountain hotels, resorts, and camps, where they can relax in company with others. Man has a thirst for knowledge, in its elements perhaps instinctive, but carrying him far beyond the promptings of mere instinctive curiosity. He wants books, telescopes, laboratories, schools, and colleges. He wants to travel and to visit new and strange scenes and so demands railroads, motor cars, and ships, hotels, and guides. We have a desire for the beautiful and so want music and works of art. We enjoy sport and recreation and therefore require golf clubs, baseball outfits, swimming suits, and all the paraphernalia of modern sport.

Intermediate goods. We must also note here the fact that modern civilized man wants many things which are not capable of satisfying directly any desire. The housewife wants a sewing machine, though there is obviously no satisfaction directly obtainable from the possession or use of this article. But the sewing machine will serve as an instrument for making and repairing the family clothing, thus helping to satisfy a fundamental need. It is not wanted for itself, but as a means to the gratification of another need. This is a characteristic of the whole group of things known as instruments, tools, and machines, which make up a surprisingly large part of the commodities which men want and for which they strive. They are called "intermediate goods" and will demand further study in the next chapter and elsewhere in our investigation.

No limit to the wants of man. This sort of study might be carried to any length. Enough has been said however to bring out the important conclusion with which we are at present concerned; namely, the infinite number and the bewildering variety of the things that man wants. Satisfying the instinctive cravings, securing the things necessary to sustain life and perpetuate the race — this is only the barest beginning. We soon forget that sustaining life is the basis of any of our wants. The more our wants are satisfied, the wider is the horizon opening up new and tempting vistas of wants still to be catered to. We are quite justified in our conclusion that man's wants, as regards their number and variety, are without limit. This is the first of the two great principles which mark out the field of economics.

Nature, the provider. The second principle is that the wherewithal to satisfy man's unlimited wants does not come without effort on his part. Nature may be a bountiful provider; she is in fact the source of everything man has or can hope to have. Some of the things he wants, such as air, sunlight, water, she provides gratuitously. But after noting these few exceptions we conclude that nature does not generally bestow her gifts with a free hand. For the infinite number of

things that man wants, man must work. Even the lower animals have to exert themselves to provide for their needs, and if man had been content with the sort of living which the lower animals thus obtain he would still be a lower animal himself. Man is always engaged in efforts to secure the things he needs for the satisfaction of his wants. His wants are unlimited; the things to satisfy them are strictly limited. There is never enough to go around. Here we have the explanation of the greater part of the activities of human beings through all history. Mankind has always been chiefly engaged in "working for his living." Here we have also the roots of the great human institutions — law, property, the family, etc., which have developed out of man's struggle for the things with which to satisfy his wants.

Economics: Definition. The insatiability of man and the niggardliness of nature are thus the foundation stones upon which rests the structure of economics. Starting with these two great premises, economics investigates the activities of man in his efforts to satisfy, so far as he may, his limitless wants out of the limited resources granted him by nature. *Economics is the science of man's activities devoted to obtaining the material means for the satisfaction of his wants.* Lest the reader miss the true inclusiveness of this definition, he is reminded of what will be made clearer a little later; *i.e.*, that the "material means" which serve to satisfy human wants include human beings as well as external objects.

When we stop to consider that an important part, not to say the greater part, of all the waking hours of most adult men and women is devoted in some form or other to the business of acquiring the means for satisfying wants, and that even the children are much engaged either in this same endeavor or in acquiring the education and training necessary to equip them to begin it a few years hence, we can realize that, although strictly limited by the definition, the field of economics is a broad one. We might proceed to catalogue the many topics included and the various problems to be in-

vestigated. But a mere glance at the table of contents of this book will perhaps suffice to give the reader a general notion of the subject matter of economics.

Economics as a science. We have been speaking of the science of economics, and it may help the reader to a still clearer notion of what economics is and also enable him to avoid some seductive pitfalls as he pursues his further path of investigation if we pause here to inquire briefly into the nature of science. A science may be defined as a body of knowledge in some particular field arranged in an orderly way to show relations of cause and effect. Science deals only with facts and causal relations. It is not the function of science to tell how things ought to be, to lay down moral precepts, to advocate that which is good, or to condemn that which is evil. Nor is it the business of science to lay down rules for the accomplishing of desired ends. That is the function of an art. Thus the sciences of astronomy and mathematics provide certain facts and causal relations; these are employed by the art of navigation to accomplish its desired ends. The sciences of physics and biology are similarly availed of by the art of medicine.

In thus sharply marking out the field of science we do not imply that moral or ethical considerations are to be ignored by the student of economics. Economics is a social science, and as such it grasps as its subject matter all attributes of man which affect his pursuit of the material means for the satisfaction of his wants. Among these attributes the moral and ethical notions of mankind are of importance and require full consideration. For example, the moral code of one community may dictate polygamy, while that of another insists upon monogamy, and these moral edicts are of great economic significance. The science of economics will tell us a great deal about the respective effects of these two systems of family organization, but we must not forget that economics cannot and does not attempt to tell whether monogamy or polygamy is right or wrong.[1]

[1] Cf. Francis A. Walker, *Political Economy*, Third Edition, pp. 19–21.

Utility. We need now to acquire clear notions of certain fundamental concepts which are inherent in the subject matter of economics and are an essential part of the equipment of the student.

Man desires those things that are able to satisfy his wants. Anything that satisfies a human want is useful; it has *utility. Utility is that quality of a thing which enables it to satisfy a human want.* This definition scarcely needs elaboration or explanation. We determine readily enough whether we want any particular thing; *i.e.*, whether it is "useful" or "useless."

In consequence of its character as a science, economics gives somewhat broader meanings to "useful" and "utility" than are employed in common speech. When the economist calls anything useful, he implies nothing as to its merits in an ethical sense. A thing is useful if it can satisfy a human want; that is all. He does not inquire whether the satisfaction of that particular want may be good or bad, wise or foolish. Opium, except when used as a medicine, may be physically injurious and morally bad, but this does not alter its classification as an object of utility so long as it does actually enable a human want to be satisfied. The use of chewing gum is probably foolish, but we cannot thereby deny its utility. The pictures which adorn the walls of many homes may be wholly bad from an artistic sense, but they are nevertheless economically useful. Of course this does not imply any lack of appreciation of the importance of aesthetic and ethical judgments. Such questions are simply outside the field of economics, as of any other science.

Wealth: Definition. *Wealth consists of all useful material things owned by human beings.* This definition needs some further examination. We have already settled the meaning of "useful." Useless things are not included in wealth, which thus embraces only those things which are capable of satisfying human wants. The reader should perhaps again be warned that the ethical character of human wants does not enter into the classification.

We limit wealth secondly to material things. This excludes such desirable things as honesty, good health, the speed of a race horse, the coldness of ice, the skill of an artist or a musician. This may possibly seem arbitrary and unfortunate at first. But we shall not have to look far to see the great advantage of this limitation and particularly the confusion which would follow if the concept of wealth were not thus limited. If the speed of a great race horse is wealth, what of the horse himself? Certainly he is wealth. But have we then two articles of wealth, the horse and his speed? And if so, should we not add a third, his strength; and a fourth, his beauty? Most of the forms of immaterial wealth with which we should have to deal if our definition admitted them are simply qualities of material things which are themselves wealth. To count both the material thing and its qualities as wealth would be productive of nothing but confusion. Such double counting is really to confuse wealth and utility. Similar confusion would arise from including human qualities, such as health, strength, skill, honesty, in the category of wealth. We shall consider a little later the question whether human beings themselves are wealth.

Having excluded from our concept of wealth all useless and all immaterial things, we next limit it to those useful, material things which are owned by man. The first effect of this limitation is to exclude those useful things which are furnished freely and abundantly by nature. Thus air is material, and it is about the most useful thing in the world. But on account of its abundance it is not owned. Why should one take the trouble to own it? We also exclude certain things which, from their nature or from the nature of the human institution of ownership, cannot be owned. We cannot own the sun or the moon; no person or group of persons can own the Atlantic Ocean or the Gulf Stream. Most material things that satisfy human wants however are owned. Ownership may be vested in the individual person, as in the case of one's suit of clothes, the farmer's land, the merchant's store, etc. Or various groups of persons, such as partnerships, corporations, etc., may own

land and buildings and virtually any kind of wealth which may be owned by individuals. Governments — national, state, and local — are large owners of wealth; they are simply another form of association of human beings.

Are human beings wealth? The requirement of ownership by human beings restricts the scientific concept of wealth in another important respect. Are human beings themselves wealth? As regards slaves, the question is clearly answered in the affirmative by reference to our definition. The slave is useful; he is a material thing; he is owned by his master. But how about free men or women or children? They are not owned under the legal systems of most modern nations. We might set up the fiction that free persons are owned by themselves or (in the case of minor children) by their parents. Free persons would then fall into our category of wealth, and there are some advantages to this classification. Free persons are material things, of course, and they have qualities which enable them to satisfy wants of other persons. Moreover it is not always easy to know just where to draw the line between free men and those who are owned. Logic and simplicity in the analysis of wealth would perhaps be served by including all human beings. On the other hand, it is not customary to think of free persons as owned, even by themselves, nor do we naturally think of free persons as wealth. The weight of advantage is on the side of our definition as stated, according to which, while slaves are undoubtedly wealth, free persons are not. Throughout the whole study of economics however it will be found necessary to take account of the utility of free persons and their services in satisfying human wants on the same terms as the utility and services of wealth.

How to identify wealth. We thus arrive at our definition of wealth. To determine whether any given thing is wealth or not, we have only to ask three questions: is it useful? is it material? is it owned? If the answer is yes to all these questions, the thing is wealth. A negative answer to any one of the questions excludes it.

Income. When an article of wealth satisfies a want of man we say that it furnishes *benefits* or renders *services*. By the benefits or services of wealth we mean *desirable events which it causes for human beings*. Keeping out the cold is a desirable event caused by a dwelling house; it is one element of the service of shelter, the benefit yielded by the house. The service of a piece of ice in the refrigerator is keeping things cold. A diamond ring renders service by satisfying the love of beauty or perhaps only the vanity of its wearer. The service of a slave is working in the field or as a domestic servant. Thus every article of wealth renders its appropriate service to man, else it would not be useful and so would not be wealth. In exactly the same way services are rendered by free persons, as when an actor performs before an audience or a lawyer tries a case for his client. The word *income* is used to include such services and is defined as follows: *Income consists of the benefits or services rendered by wealth or by free persons.*

Costs. Not all the events caused by wealth are desirable. Besides furnishing shelter, a dwelling house requires painting and repairs, it compels its owner to pay taxes and insurance premiums, and so on; finally it wears out or becomes obsolete and if not discarded must be torn down and rebuilt. The slave works in the field, but he requires food, clothing, and shelter, and the free farm laborer requires wages. *The undesirable events caused by wealth are called the disservices or costs of wealth.*

Net income. All wealth yields both income and costs. We cannot escape the undesirable events if we would have the desirable events; we put up with the costs in order to get the income of wealth. It follows that, in the long run, the income from any article of wealth must exceed its costs (how these are measured will appear later), else no one would care to own that article, and it would cease to be wealth. For example a storekeeper has a motor truck to deliver goods to his customers. The truck is an article of wealth; the income from it is the service of delivering goods or the receipt of money for delivering goods; its costs are the expenditures of money for the gaso-

line and oil it consumes, garage room, occasional washing, and repairs. These costs would not be borne if the service of delivering goods were not worth more than the costs. Eventually the truck begins to wear out; its speed diminishes and it cannot make some of the grades. It consumes more gasoline, and its loose piston rings and bearings are wasteful of oil. Repair bills become heavier and more frequent. Thus its income declines while its costs rise, and sooner or later the margin between them disappears. The machine is no longer useful, and it finds a resting place on the junk heap; it is no longer wealth. *The difference between the income and the costs of any article of wealth is its net income.* Some net income must be present or the article will not be wanted. In general then utility means the power to yield a net income or an income in excess of the costs, and this idea of a net income is inherent in the term "useful" as employed in the definition of wealth.[1]

Wealth and income. Fund and flow. There is a fundamental difference between wealth and income, which manifests itself in the matter of measurement. When wealth is measured, it is the quantity in existence at a certain time that is determined. In measuring income, we ascertain the amount of service that is rendered during a certain period of time. We do not speak of the yield of an orchard or the service of a carpenter at 1.25 P.M. of a certain date. It is ordinarily impossible to think of income except in reference to an appreciable length of time. An orchard yields so many bushels of apples per year, a mason lays so many bricks per day or per hour, etc. Wealth is a *fund;* it is measured as of a certain *instant* in time. Income is a *flow;* it is measured as of a certain *period* of time.[2]

[1] The reader who has any familiarity with accounting will note that these definitions of income, cost, and net income differ somewhat from those used by the accountants. They are not really in conflict. The accountant's terminology represents a special technical application of the more general concepts with which we are here dealing.

[2] By the methods of the infinitesimal calculus, the period taken for the measurement of income may be reduced to zero, thus making possible the measurement of income as a *rate* of flow at an instant of time. This concept is of great importance in certain advanced problems of economics, as it is in some of the natural sciences. It need not further concern us in this elementary book.

Property: Definition. Every article of wealth is owned by somebody. This relation between wealth and its owner is one of the most fundamental concepts of economics. It may be called ownership, property right, or property. *Property* is the technical term generally used in economics. It is defined as *the right to income; that is, the right to the benefits or services of wealth or free persons.* This evidently is what ownership means. When a person owns an article, he has a right in it; namely, the right to have it and use it, all of which is included in the right to the services or income from that article. The owner is protected in his property right by the laws and customs of the community in which he lives; other persons are prevented from interfering with his enjoyment of the services of his wealth. Since the term income is so broad as to include any use to which the owner could put his wealth, including even its destruction if he so desires, property might be defined simply as the right to wealth or free persons, without any change in meaning. The definition here chosen has the advantages, (1) of emphasizing the fundamental importance of income, (2) of somewhat readier adaptation to the innumerable cases of divided ownership, and (3) of avoiding any apparent conflict with the common usage of not regarding free persons as owned.

Division of property. A sole owner has the exclusive right to all the income of the article of wealth in question, as in the case of the owner of a house and lot, an automobile, a suit of clothes, etc. Ownership may however be divided, in which case each of the two or more owners has certain rights in the wealth; that is, the right to certain of its benefits or income. The tenant of a rented house has the right to its use only during a certain time, as limited by his lease contract with the landlord, who has retained the right to the services of the house at all other times; that is, before the beginning and after the termination of the lease. In a partnership the wealth held belongs jointly to the two or more partners. Each has the right to a certain part of the services or income of this

wealth, the exact division being specified in the partnership agreement. So in the corporate form of business organization, the wealth of the corporation belongs ultimately to the stockholders, of whom there may be hundreds or thousands, each with certain rights in the income from the corporation's property. The bondholders of the corporation represent another group of part owners, with certain other rights in the corporation's wealth, all specified in the bonds. Thus each bondholder has the right to receive certain interest payments at certain specified dates and finally at a specified date to receive a certain larger sum equal usually to the principal amount stated in the bond. That part of the income from the corporation's property which remains after the rights of bondholders and possibly other classes of part owners, such as note holders, etc., have been satisfied belongs to the stockholders and may be distributed to them in the form of dividends. Property may thus be divided in a great variety of ways. Every person who has a property right in an article of wealth is an owner (sole or part) of that wealth, though in common speech the term "owner" is sometimes applied to only one of the parties. Thus the landlord and the tenant are really both owners of the rented house. Neither is the exclusive owner, for neither has the right to all the services of the house. The property is divided between them, though we are accustomed to apply the term "owner" to the landlord alone. In like manner the bondholders are not usually spoken of as owners of the corporation, though fundamentally they are part owners as well as the stockholders.

All the services of wealth must belong to somebody. When property is divided, the sum of all the property rights of the several part owners must equal the total property right to all the services of that piece of wealth. For every article of wealth there must be a property right or property rights, and back of every property right there must always be an article of wealth or a free person or free persons.

Property in free persons. The last words of the previous paragraph will recall that property may be a right to the services of free persons as well as of wealth. If we had chosen to include free persons in the definition of wealth, all that has been said of property in wealth would have applied equally to them. As it is we must recognize that free persons render services to others and that these others have rights to such services. A baseball player signs a contract with a manager agreeing to play exclusively for him during the coming season. The manager has thereby a property right, the right to receive certain services from this player. To that extent the player's right to his own time and services — actually his right to himself — is diminished. It is a case of divided property again. Theatrical managers, moving picture producers, musical promoters, and others have property rights in actors and actresses, musicians, and other artists. Every employer has a property right to certain services of his employees. A promissory note gives the creditor a property right in the debtor; that is, the right to receive a certain sum of money. Lawyers have rights against their clients, and clients against their lawyers; parents have rights in their children, and children in their parents. Every business contract gives each party a property right against the other. In fact the property in persons is almost always divided. Practically no one is completely his own master, and there are all degrees of property rights of others, from the slave, a true article of wealth because completely owned by his master, through serfs, peons, indentured servants, to the freest so-called "free man."

Wealth and property. In popular speech, the terms property and wealth are not carefully distinguished. It is quite usual for example to speak of a house and lot as "property," to call household furniture "personal property," and so on. For the purposes of scientific analysis, economics draws a sharp distinction between property and wealth. The latter is a physical, tangible thing; it is material. Property, on the other hand, is a relation between wealth (or free persons) and

persons; it is immaterial. This distinction is fundamental and has many important bearings. Thus wealth is the source of income; whereas property is the distributor of income. The wealth of a community, together with its productive free persons, determines how much and what kind of income the people may enjoy. The property rights of the several members of the community determine how these incomes shall be divided among them, how much of the total community income each one shall enjoy. Property rights may be changed without changing the total amount of wealth. The material welfare of the human race is a matter of the satisfaction of wants and is affected both by the total quantity of wealth and by the way the property rights to wealth are distributed among individuals, families, and nations.

Documents in evidence of property. Property rights are sometimes expressed in legal documents. The owner of a farm or a building lot has a *deed*, which is a piece of paper containing a legal statement of his property right to all the benefits of a specified parcel of land. When one purchases the right to sit in a certain seat of a theatre on a certain date, his property is evidenced by a ticket. On the other hand, no legal document testifies to one's property right in the clothes he wears or the food on his table. Whether there shall be a document or not is a matter of custom or law. The property is the original and fundamental thing; the document, when there is one, is secondary and incidental; in many cases no document is considered essential. The reader must be on his guard against confusion between property rights and the documents in evidence thereof, particularly in those cases where common, and even scientific, usage has not found it necessary to employ different terms for these two things. For example a bank note is, strictly speaking, a piece of paper testifying to a property right of the bearer against a bank. Yet it is common to speak of the bank note as though it were the right itself. When once he has been warned, the reader is not likely to have any difficulty with such usage, loose though it confessedly is.

Value: A common denominator. Each kind of wealth, income, or property has its appropriate unit of measure. But there is one unit in which all kinds of wealth, income, and property may be measured and which is by far the most important of all units. This is the unit of value or price. The universality of exchange is one of the characteristics of the modern economic organization, to which much attention will be given in the succeeding pages. Value and price are, next to wealth itself, perhaps the most important of the fundamental concepts with which the student of economics must equip himself at the outset of his study.

Transfer and exchange. When Mr. Jones gives his dwelling house to his wife, there is a change of ownership. The house, which formerly belonged to Mr. Jones, now belongs to Mrs. Jones. A *change of ownership of wealth* is called in economic terminology a *transfer*. Transfer takes place whenever wealth is bought or sold, bequeathed, or given away. Now suppose Mrs. Jones sells the house to the Brown Realty Co. for $20,000. There has been another transfer of the house. But this time there has been also a transfer of money, from the Brown Realty Co. to Mrs. Jones. These two transfers together make an *exchange*, which is defined in technical language as *a pair of voluntary transfers between two owners, when each transfer is made in consideration of the other.* A gift is not an exchange, since there is only one transfer. As is said, there is "no consideration."

Barter and money exchange. *Exchanges made without the use of money are called barter.* For example, the farmer brings ten dozen eggs to the village storekeeper and takes in return ten gallons of gasoline and ten quarts of motor oil. This transaction fits the definition of an exchange. "Swapping" horses or anything else is barter. At some time in the history of any people barter was probably the usual form of exchange, such as there was, and a certain amount of barter goes on among all peoples today. But the civilized world long ago abandoned barter as the normal form of exchange. In practically all ex-

changes today one of the transfers is a transfer of money, and money exchange has thus taken the place of barter. The subject of money will come up for special study in later chapters. For the present it is sufficient to know that *money is anything which is generally accepted in exchange for other things.*

Definition of value. From exchange come two of the most important concepts of economics and business, not to say of modern human life in general; namely, value and price. As the term is employed in economics, *the value of anything is the quantity of any other thing that would be given in exchange for the first thing.* If ten dozen eggs could be exchanged for two bushels of potatoes, the value of the eggs is two bushels of potatoes, and the value of the potatoes is ten dozen eggs. If a ton of stove coal is sold for twelve dollars, the value of the coal is twelve dollars, and the value of twelve dollars is one ton of stove coal. It is not customary to speak of the value of money in this way. On the other hand, values of practically all other things are regularly expressed in terms of money.

Definition of price. *The price of anything is the amount of money that would be given in exchange for one unit of it.* If a pound of cotton would sell for twenty cents, the price of cotton is twenty cents per pound. In stating the price of anything, the unit of measure used must always be stated or understood.

Value and price. The terms value and price are of course used more or less vaguely and inconsistently in popular speech, and there is some difference in usage among economists. The definitions laid down here agree as closely as is possible with popular usage, and they will prove useful for economic analysis. As here defined, value and price are closely related to each other. They are distinguished from each other in that (1) value may be expressed in terms of any kind of wealth, property, or service, while the price of anything is expressed only in money, and (2) the term value is used for any quantity of a good, whereas price relates specifically to one unit. In practice value as well as price is usually expressed in terms of money; in which case value is price multiplied by quantity,

and the value of one unit is identical with the price. Thus if a farm sells for $25,000, we may say that the value of the farm is $25,000, or that its price is $25,000. We should not say that the price of fifteen tons of coal is $150, but should say that the value of fifteen tons of coal is $150, because the price is $10 a ton.

The unit of value. Value thus gives us a common unit in which to measure all kinds of wealth, income, and property. This is the unit of money value; *i.e.*, the value of the monetary unit, in America the dollar. Since value is almost always expressed in terms of money, we may ordinarily call the dollar the unit of value, not bothering always to say the unit of "money value." Thanks to this common unit, it is possible to total a list of various articles of wealth and property rights. Bushels of potatoes, yards of cloth, pounds of sugar, acres of land cannot be added together. But their values can be. It is possible therefore for the business man to make an "inventory," for the citizen to make up his "tax list," for anybody to make a list of his possessions and get the total, and it is possible to keep accounts of wealth and property and income.

EXERCISES

1. Which of the following may properly be called wealth? Explain in each case.

(*a*) A fountain pen. (*b*) An abandoned farm. (*c*) A factory building. (*d*) A state highway. (*e*) The Gulf of Mexico. (*f*) A farm whose ownership is under dispute. (*g*) Fish in the ocean. (*h*) A moving picture actor. (*i*) A college dormitory. (*j*) An automobile on the junk heap.

2. Which of the following are wealth and which represent property rights? Explain in each case.

(*a*) A share of stock. (*b*) A theatre building. (*c*) A bank deposit. (*d*) A dollar bill. (*e*) A suit of clothes. (*f*) A life insurance policy. (*g*) A delivery truck. (*h*) A football ticket. (*i*) An apple orchard.

3. During the last depression the United States Government borrowed billions of dollars from the people and from banks through the sale of bonds.

(*a*) Does this debt affect the wealth of the United States? (*b*) During the years these bonds are outstanding, interest will be paid the bondholders out of funds raised by taxation. Does the annual payment of interest on a national debt of this kind affect the wealth of the country? (*c*) Will the wealth of the country be affected if the bonds are paid off?

NOTE. This exercise refers to immediate effects, not to remote and indirect consequences.

4. In 1932–33 a number of states adopted laws providing for a moratorium on farm mortgage indebtedness. The holders of mortgages were unable to collect interest or undertake foreclosure.

(*a*) Did such laws affect the wealth of these states? (*b*) Did such laws affect property rights?

5. Describe the income created by each of the following.

(*a*) A farm. (*b*) A musician. (*c*) A watch factory. (*d*) A doctor. (*e*) A baseball player. (*f*) A fishing vessel. (*g*) A housewife.

CHAPTER II

The Factors of Production

Production. The first step in the study of production must be to settle exactly what is meant by the term. Who are the producers of the world? Common usage would lead us to say at once that the farmer, who raises wheat, is a producer; he obviously produces wealth, in that he creates something which did not previously exist. But a little reflection will show that he is a creator or producer only in the sense that he has been instrumental in effecting a change in the form or relationship of various physical elements already in existence so as to make them suitable for human consumption. The cotton manufacturer likewise adds nothing to the world's store of materials, but when he twists cotton fibre into thread and weaves thread into cloth he has added to the utility of the cotton in the eyes of the consumer. If we examine other forms of agriculture or manufacturing, we shall find that the essence of the particular operation is to bring about a change in the form of the good which will bring it one stage nearer the condition where it may satisfy a human want. Agriculture and manufacture, carpentry and blacksmithing, and innumerable other lines of activity, by thus changing the form of material goods, are active in the *creation of utility.*

But this is not the only way that utility is created. There may seem to be little similarity between farming and the operation of a railroad if one views only the technique of each business, but the purpose is essentially the same. Wheat on the farms of the West is quite as useless to the consumer of the East as if it had never been grown, and the service of the railroad in bringing it to the consumer is just as essential as that of the farmer in growing it. The grocer who keeps flour in stock to meet his customers' demands is performing a similar

service, for the consumer not only wants a particular good at a particular place, but he wants it at a definite time. To-day's flour is patently more useful in satisfying today's hunger than flour which will not be available until next week. Likewise a productive service is rendered by those who facilitate the sale of goods. When A desires to sell wheat, he gives his broker a warehouse receipt for wheat stored in some elevator; the broker sells the wheat to B by delivering to him the warehouse receipt. A service has been rendered, for which the broker will collect commissions from A and B. The wheat is presumably one step nearer to the direct satisfaction of a human want. Utility has been created, by a change in ownership, even though the actual wheat has not been touched.

Thus are illustrated the four forms which the production of wealth may take — the creation of *form utility*, of *place utility*, of *time utility*, and of *ownership utility*. All effort directed toward the production of wealth may be classified under one or another of these four headings; agriculture and manufacturing result primarily in the creation of form utility, all the varied transportation services are engaged in creating place utility, merchants of all classes and bankers, brokers, and other middlemen are creating time utility and ownership utility. To the first one alone the name production is sometimes given, doubtless because of the obvious character of the change, but in view of the essential nature of the other three forms it is difficult to justify such usage, and we shall use the term production of wealth to include all four.

Although our primary interest in the study of production centres around the production of wealth, it should not be forgotten that man has wants which cannot be satisfied by wealth alone. A mechanical piano may satisfy the longing of one for music, but to another it is a sorry substitute for the performance of the pianist. If the labor of the pianist in making the roll for the mechanical piano is productive, certainly his performance on the concert platform must also be productive. We must therefore include in our definition of production the

rendering of direct personal services by free persons. Completely stated, *production consists of (a) the production of wealth, which is defined as the creation of utility in wealth, and (b) the rendering by free persons of direct personal services.*

This definition is intentionally broad, including any effort which results in the satisfaction of a human want. In defining utility, it was pointed out that scientific definitions have nothing to do with questions of ethics or morals. Like utility, the word production carries with it no moral significance, although economists are not therefore unaware that effort spent in some directions may be less commendable than effort spent in others, and that the results of human exertion may be fatal to individuals or disastrous to society. The study of production is the study of human activity in satisfying wants, whether or not that activity can be approved on moral or ethical grounds.

The factors of production. In spite of the fact that man, physically and intellectually, is today probably little different from what he was at the time when historical records begin, and in spite of the fact that the natural forces and laws have remained the same, the effectiveness of man's efforts devoted to the satisfaction of his wants has during the period covered by history immeasurably increased. The explanation lies in the fact that man has gradually accumulated knowledge of the natural environment and learned to make it serve him in production.

In his struggle to guide and control natural forces to his own ends man has enlisted in his service two powerful partners, and these three — man, nature, and capital — work together in production, with results which, when compared with man's first efforts, are almost unbelievable. These three partners are commonly called the *factors of production* under the titles, *land*, *labor*, and *capital*, the first two being figuratively used to stand for nature (not merely land in the narrow sense) and man. The rest of this chapter will be devoted to the study of their place in the economic organization.

Land. Land is essential in providing "standing room" for the population of the earth; it is the source of the food supply for the people and the raw materials for production and is the original source of all that man has. It is more than mere land, for it embraces the whole natural environment, including the oceans, lakes, and rivers, the topography of the land, the character of the soil, the amount and seasons of the rainfall, the range of temperature at different seasons, the mineral deposits, the water power, and a host of other features, all of which are important and may have a determining effect on the character of occupation and the type of industry in a particular region.

Labor. Land without labor is useless for the purposes of man. Even in the spots most favored by nature some effort is required to secure the goods desired, and on most of the earth's surface much exertion is necessary to make a living. For the majority of men labor is painful and is avoided if possible. Only when the desirability of things which can be secured by labor is considered to outweigh the unpleasantness of mental or physical exertion will the effort be undertaken. Mental effort is as truly labor as physical exertion. Probably the physician derives more pleasure from his work than a factory helper does, but our only standard of whether a given effort is labor or not is whether it contributes to production. The efforts of the physician, the lawyer, the judge, the business executive, and the government official all add to the total amount of enjoyable goods and services which constitute the income of society. Some of it directly increases the flow of goods and services, some indirectly by making the work of others more productive.

Capital: Description. The third factor in production is *capital.* Ordinarily the term capital is synonymous with wealth; that is, useful material articles owned by man, but in discussing the factors of production it is convenient to limit the term capital to wealth which man has produced, thus excluding land. Land is distinguished from capital in this

classification because of its peculiar relationship to production, but it should be remembered that it constitutes a particular and important class of capital in the ordinary use of that term.

At this point we are particularly interested in those capital goods which are useful in production. There is an important group of capital articles which is desired primarily for its own sake rather than as instruments of production, including the mass of goods ready for consumption — food, clothing, furniture, etc. To this we give the name *consumers' capital*, while we call the tools, implements, machines, and materials which are to be used to produce consumable goods *intermediate goods* or *producers' capital*. It is however not always possible to draw a sharp line between producers' and consumers' capital, and the terms are not always mutually exclusive. Some goods are capable either of being used for further production or of yielding their satisfactions at once. Coal may be mined for use in factories, or it may be used in the kitchen stove with the double purpose of warming the room and cooking food, or it may be burned in the open grate. A ship may be capable of use as a yacht or as a carrier of freight. Moreover virtually every sort of consumers' capital may, when used to satisfy wants, be at the same time serving the ends of production. This is obviously true of the form of consumers' capital *par excellence*, food, and it will be recognized as true of many of the less essential forms and even of many mere luxuries, such as athletic goods. But in spite of these qualifications we shall find the concepts useful in analysis, and with this warning the reader can easily avoid confusion.

Capital in production. We are at the present time so accustomed to tools and machinery of every description that it is difficult to imagine conditions in which they were entirely lacking; yet at some stage, long before the beginnings of history, man must have been dependent on the unaided efforts of his hands. Such a situation, where brute strength or cunning is relied upon to secure the means of subsistence, may be called *direct production*. What the first implement was is a

matter of conjecture; it may have been nothing more highly developed than a club. Then possibly someone conceived the idea of attaching a sharp stone to the club, and an axe or hatchet was the result. Certain it is that tools came into use very slowly, and for centuries only the rudest and most primitive were known. Even during most of the centuries of which we have historical records, the machines and tools were of the simplest sort. But with the first tool came a definite advance, for man had started on his yet unfinished process of changing the conditions of his environment to meet his rapidly growing needs.

Production with tools and implements is called *indirect* or *roundabout* production. One may for example acquire the art of catching fish with his hands — this is direct production; but if he take time to make a fishing net he can greatly increase the effectiveness of his labor, including, as of course we must, the labor spent in making the net. Instead of proceeding directly to secure the desired products, time and energy are first spent in making tools, with the result that the production is made vastly more effective and, passing beyond the small range of activities open to one without tools, production brings forth things which could not conceivably have been made with the unaided hands.

We are not however particularly interested in the contrast between direct and indirect methods of production; the economies of the latter method are so patent as to require no further demonstration. The contrast of significance is between a less and a more intensive resort to the roundabout methods of production. The labor necessary for the erection of a modern flour mill may be spread over many years, the materials for its construction may be gathered from the four corners of the earth, and hundreds of men may be engaged on one or another of the preliminary processes, whereas the old-fashioned gristmill could be made in a few weeks' time by a relatively small number of laborers; but the output of the modern mill per unit of labor necessary to erect and operate it is consider-

ably greater than came from its predecessor. Wherever we find an example of intense recourse to the roundabout or capitalistic method of production, we may safely conclude that it is because the effectiveness of labor, as measured by the flow of goods, is thereby enhanced. Capital instruments are labor-saving devices, not only from an individual's point of view, but from the point of view of society as well.

The accumulation of capital. Saving. The degree to which indirect methods of production may be carried depends, in part at least, on the quantity of capital which is available for productive purposes. Other elements enter, such as the progress of invention, the size of the plants in the industry, the demand for handmade articles, and so on, but the quantity of capital is so fundamental a consideration that a study of the way in which it is accumulated will repay our consideration.

The formation of capital depends upon the readiness of men to postpone the gratification of wants in the present in the hope of increased satisfactions in the future; in other words, it rests on *saving*. In relatively simple cases this is quite clear. The savage who has gathered a supply of corn may be tempted to consume it before the spring, but if he can save enough for seed the following year he can hope for a more abundant supply in the autumn. A farmer today may use the proceeds from the sale of his crops to buy either a new automobile or an agricultural machine. The former will yield him pleasure at once, but the latter will enable him to expand his farming operations and to have a larger surplus of grain in the future. Nearly every manufacturer is compelled at some time or other in his career to choose whether to use the profits of the business to expand his scale of living or to "put them back in the business" in the form of enlarged facilities.

In these cases we can easily see that self-denial — saving — has added to the stock of capital. For the majority of people however the connection between saving and the productive processes is not so clear or so direct. Today the great mass of

people gets its living by exchanging services for money and with this money purchasing the various articles which are required or desired. To the individual the saving of a part of his income does not seem directly to affect production; it appeals to him solely because of the necessity of providing against illness or old age or because money can be left in a savings bank or loaned or used to purchase stocks and bonds and can thereby earn interest. Yet it has precisely the same influence on production as saving by the manufacturer.

If we consider saving from the standpoint of society as a whole, it will be apparent that it entails the devotion to the production of intermediate goods of effort which might have been spent in producing goods capable of yielding their satisfactions at once. It further implies that the consumption needs of those working on the intermediate goods must be met from the flow of consumable goods produced by others. Were it not so they could not live.

We can imagine a simple kind of social organization where the tasks are arbitrarily divided among the different men and women, where all share in the common product, where the farmers are required to set aside a portion of their produce to support the miners and others working on intermediate goods, receiving their recompense eventually in the form of agricultural implements which will greatly increase the effectiveness of their efforts.

But although we may be able to find isolated cases where the intermediate worker is supported by direct payment in commodities, such as "grubstaking" a prospector, it is far from the normal case. The ordinary individual saves money, not commodities, and the laborer receives his remuneration in the form of money. This feature serves to obscure the nature of the problem and make it seem more difficult than it really is. A clerk who saves $500 in the course of a year does so by denying himself the satisfactions which might have been obtained from the consumption of goods. When he places the sum in the savings bank, it is not kept in the bank's vaults

until such time as he may require it but is loaned to some borrower. Now it is possible that the borrower may want the money to purchase goods for his own immediate satisfaction, but most loans are made to men actively engaged in business and for business purposes. Let us suppose that the savings bank uses the money to buy a bond, a part of an issue of bonds put out by a railroad in order to be able to finance the construction of a branch line. The railroad uses the funds thus secured to engage men to work on the construction of the line and to purchase the results of past labor — steel rails, ties, etc. The clerk has actually passed on his control over consumable goods to the laborers engaged in the intermediate processes; by reducing his own consumption he has made it possible for them to work on non-consumable goods and yet to live. During the years which may be required to construct the railroad line the men engaged in all the preparatory stages might have been producing food, clothing, etc., and for the time being have greatly increased the flow of consumable goods, but the eventual result of abstinence on the part of some in society is the creation of a great machine of production, the railroad, which makes the efforts of workers in all lines many times more fruitful. The saver gets his reward in the form of interest, but he also shares in the social gain, which consists of a more abundant flow of consumable goods.

It is difficult to overestimate the importance of the capitalistic or indirect methods of production and the necessity of saving, and it is easy to overestimate the adequacy of the volume of savings at present available in the world. Were the people of the nations of the earth suddenly to stop saving, spending all earnings on consumable goods, it would be but a short time before the entire structure of our civilization would be crumbling in the dust. On a smaller scale this happened in Russia. In the days before the World War Russia's industrial fabric depended in large measure on loans from foreign nations. Under the Revolutionary and early soviet régimes Russia to a large extent was cut off from foreign

sources of supply and in the disorder which prevailed was living in part on her capital. For a time there was practically no replacement of capital goods which had deteriorated or worn out. The result was an enormous decrease in the efficiency of Russian labor and the threatened extinction of many lines of industry.

The form of capital instruments. Invention. The effectiveness of labor in production depends not only upon the amount of capital available for use in production, but also upon the form which this capital takes. At the present stage in industrial development we could doubtless use to advantage a greater number of the tools and machines with which we are already familiar, but there must be some limit to the gain which comes from duplicating our present machinery. A more profitable line of development lies in improving the type of machines. Not two printing presses of the same capacity, but one improved press with twice the ordinary capacity.

The inventive genius of the world is constantly studying the possibilities of improving old machines and devising new machines to take care of processes now performed by hand. Society profits by the inventions which make it possible to secure a greater flow of commodities with a smaller expenditure of effort. The saving which we have expressed in general terms appears to individual members of society in the guise of a lower price and therefore means that a given income represents command over a larger number of commodities than formerly. Measures which stimulate inventive genius are generally in the public interest. The inventor or the one to whom he transfers his invention may make a fortune in a few years through the monopoly which a patent right confers on him, but the people as a whole benefit still more. In the long run the hope of making a fortune is the spur which leads the inventor to persevere through the period of fruitless experimentation. The wealth which Edison accumulated appears a small price to pay for the services he performed for

society in placing so many comforts within the reach of people of modest means.

The entrepreneur. Land, labor, and capital cannot work together without organization, direction, and control. Moreover in every productive undertaking there is *risk; i.e.*, the chance that the enterprise may prove not profitable and that the capital devoted to it may be lost and the human effort wasted. The owners of the several factors of production are, as such, not prepared to undertake this managerial function or to assume the risks of business. The landowner will provide land for a business if somebody will guarantee him a fixed rent. The laborer will work for an employer if the latter will by contract make himself legally liable to pay a definite wage. Finally the capitalist, in so far as he is performing the capitalistic function exclusively, lends money at interest to someone who can persuade him that the amount of the loan will be paid back in full. Someone must borrow the capital, decide what to produce, and employ labor and land to produce it, in the hope that the venture will be successful enough to pay him profits over and above the fixed interest which he must pay to the owner of the capital. If the business is a failure, he must use his own resources to pay back the capitalist and to discharge his legal obligations to the landowner and the workers. *Anyone who performs the function of organization, management, and risk bearing is termed an entrepreneur.*

The entrepreneur is thus the man who makes the fundamental decisions as to what commodities to produce and how to combine labor, land, and capital in their production. He determines the degree to which it is profitable to supplant manual labor by labor-saving machinery, the amount of a given good which can profitably be produced. Whether a new invention is to be given a trial depends on his judgment. He studies the nature of the industrial operations, tries to simplify them as much as possible, and assigns laborers to one task or another according to his opinion of their abilities. The responsibility for the entire internal and external organ-

ization of a business falls on his shoulders — the coördination of production, finance, buying, and selling. The larger the business, the more incentive there is to delegate to others all tasks of a routine nature, and the more closely does the work of the chief executive partake of the nature of decision-making.

So important is this function in the modern economic organization that the entrepreneur is by some writers elevated to the rank of a fourth "factor of production." Since whatever the entrepreneur does falls clearly within our definition of labor, there seems no need of thus putting him in a class by himself. He is economically a laborer, although in the popular separation of social classes he generally falls in the capitalist rather than the laboring class. In fact the entrepreneur is quite generally a considerable owner of capital and so a capitalist as well as a laborer. We are able nevertheless to distinguish between the functions of laborer, capitalist, and business leader, even though they may be united in one person.

EXERCISES

1. The creation of a physical object may be unproductive; and there may be production when no physical object is created. Explain what is meant, and give illustrations.

2. Which of the following are producers? Explain in each case.

   (*a*) A stenographer. (*b*) A locomotive engineer. (*c*) A life insurance salesman. (*d*) A racketeer. (*e*) A dentist. (*f*) A mechanic. (*g*) A beggar. (*h*) A burglar. (*i*) A professional baseball player. (*j*) A fortune teller.

3. A, B, C, D, and E each saves $1,000. A buys a piece of land. B deposits his savings in a bank. C invests his in the stock of a newly formed paper manufacturing enterprise. D spends his on a trip to Hawaii. E sets up an automobile repair shop. Explain how each of these activities affects the formation of capital.

4. Show how the factors of production are combined in each of the following enterprises:

   (*a*) A dairy farm. (*b*) A coal mine. (*c*) A retail store. (*d*) A doctor's office. (*e*) A fishing enterprise.

CHAPTER III

The Form of the Business Unit

Control of production in the modern world is largely in the hands of the entrepreneurs, who organize and direct industry, induced to do so by the expectation of private gain in the form of business profits. In the course of the extraordinary era of economic progress ushered in by the Industrial Revolution, old ways of conducting business have been modified, and new forms of business organization have been introduced in order that the several branches of industry might adapt themselves to changing conditions and might function more easily and efficiently and that increased profits might be secured. The form of the business unit and its manner of functioning are matters of considerable significance to society.

The individual proprietorship. In this form of organization the proprietor is in sole charge of the business, responsible for its success or its failure. He may (in most states) start up in business when he chooses and cease operations when he so desires, without the bother and expense of legal formalities. Unless an activity is specifically prohibited by law, such as the coining of money or the carrying of mail, no line of business is closed to him. Responsible to himself alone, he is in position to make decisions at once without seeking either the advice or the assent of an associate, and he is thus able to take advantage of the opportunity of the moment. His business secrets cannot become common property through the indiscretion of associates.

While there are advantages for the small business in this form of organization, there are certain drawbacks which make it undesirable for a concern of any great size. In the first place the individual is seldom able to invest as much capital

in his own business as can be secured by a partnership or a corporation. And if he is able to invest a large amount of capital, the great risk involved is a deterrent, for he is personally liable for all debts of his business. Hence at the present time we find relatively few large enterprises organized and operated by a single individual. While the number of manufacturing plants owned and operated by individuals is over one-half of the total number, the output of these plants forms an insignificant part of the total output. In agriculture alone the individual is still the predominant figure.

The partnership. The partnership is a more complex type of organization than the individual proprietorship, for it involves relations between the individual partners as well as the relationship of the partnership to the public. Essentially it is a voluntary association of two or more individuals for the performance of a specific or a general object. This object may be accomplished in a day or a week, or the association may be formed for more general purposes, such as the establishment and operation of a factory, and may last a longer time.

In any case the purposes for which the partnership is formed, its duration, the ways in which it may be dissolved, and the mutual rights and obligations of each partner to the other will be the subject of an agreement usually set forth in written form. As was true in the case of the individual proprietor, the partnership can be created with little legal formality and without delay.

In the event of inability to meet all debts from the funds of the partnership, each partner is liable to the full extent of his personal resources for the debts of the business, even though the agreement may make one partner responsible for only a certain proportion. If Jones and Brown suffer large losses and the debts exceed the resources of the partnership, and Jones finds himself unable to meet his share of the losses, Brown must make up the difference if he is able to do so. This is known as *unlimited liability* and is one of the chief

points of difference between the partnership and the corporation.[1]

The fact that there are several men united together makes it possible for the partnership to control a larger amount of capital than can the individual. Few concerns are able to finance themselves with their own capital exclusively; they must rely to some extent on bank credit, secured through loans from banks, and on commercial credit, obtained by purchasing goods to be paid for in the future. The ability of the partnership to secure capital in these ways is likely to be greater than that of the individual, in so far as the unlimited liability of several men offers a greater margin of safety to the creditor than the unlimited liability of one.

Another advantage of the partnership over the individual proprietorship may come from the association of abilities. An important question will be viewed from more than one point of view, and the decision which is reached may therefore be sounder. Offsetting this there is the possibility of serious disagreement and consequent delay in making decisions, but at least snap judgments are likely to be avoided. A certain degree of specialization is possible, as in a wholesale house where one partner is entrusted with the buying, another with the selling, and the third with financial questions.

One of the chief weaknesses of the partnership is the prospect of its enforced dissolution on the death of a partner. Under the common law doctrine,[2] unless modified by statute law, the death of one of the partners automatically dissolves the partnership and makes necessary an accounting and distribution to his heirs of the deceased partner's share of the net resources of the partnership. This may be a hardship, both to the partnership and to the heirs. But the heirs cannot

[1] It should be noted that there are a few limited liability partnerships, which are in this respect similar to the corporation (see p. 39 below). These are exceptional.

[2] Common law is based on universal custom which has received recognition and sanction in decisions rendered by the courts. Statute law consists of laws enacted by a legislative body, such as the Congress of the United States or the legislature of one of the states, and enrolled on the statute books.

join the partnership without the formation of a new partnership.

The theory on which this doctrine is based is that the partnership is a peculiarly personal relationship and that anything which occurs to interfere with this must necessarily bring the original arrangement to an end. A logical corollary is that a partner may not transfer his interest in the partnership without causing a dissolution of the partnership, because no associate can be forced on the other partners against their will.

But the partnership has such obvious advantages over the individual proprietorship that, before the corporate form was devised and made common, it was the prevailing form of organization in enterprises of moderate and large size, and it is still not uncommon to find large concerns conducted for many years under a partnership agreement, although the corporation has steadily increased in favor and now overshadows the partnership in importance.

The corporation: Modern growth. Though the most significant type of business unit in the United States at the present time, the corporation is of comparatively recent development, the great increase in the number of corporations having come since the Civil War. In the fields of railroading, banking, and insurance the corporation is practically supreme today. Even in manufacturing, where its growth has been slower, it has a position of dominance. In 1929, although only 48 per cent of all manufacturing establishments were owned by corporations, they turned out almost 92 per cent of the total value of the products, and employed 90 per cent of all wage earners in manufacturing. In some lines of manufacturing the importance of the corporation is even more striking. For example, in the manufacture of rubber boots and shoes corporations have entire control. They turned out 97 per cent of all the agricultural implements made in the United States and 99 per cent of the clocks and watches. In occupations formerly thought the province of the small retailer, the corporation has made its appearance, as is evi-

denced by the rapid growth of chain stores in the grocery, drug, tobacco, restaurant, and other fields. Taking our economic activity as a whole, 57 per cent of the total national income in 1929 was produced under the corporate form of business enterprise and the remaining 43 per cent under other forms of business organization.[1]

Nature of the corporation. A corporation has been defined in part as "an artificial being, invisible, intangible, and existing only in contemplation of law." This definition indicates one of the most distinctive characteristics of the corporation; in the eyes of the law it is a legal or artificial being, a fictitious person, conceived to have an existence apart from that of its owners. By contrast, the partnership exists in the persons of the partners. For example, suits by or against the partnership are brought in the names of the partners, but the corporation sues and is sued in the corporate name.

Furthermore the corporation depends for its existence upon the sanction of the law. The individual business man and the partnership carry on business by virtue of rights long recognized by common law, but the corporation exists only through the official act of the state. If the organizers fulfill the requirements as laid down in the corporation laws and make application in due form to the proper state official, a *charter*, or *certificate of incorporation*, will be issued to them defining the rights and duties of the corporation.

Called into being by the act of the state, the corporation exercises only such powers as have been conferred upon it. A partnership may engage in any activity which is not prohibited, but the corporation is restricted to the range of activities sanctioned by the terms of its charter or the general laws, either explicitly or implicitly, as essential to its existence. Acts beyond the scope of the corporation's powers are termed *ultra vires* and may result in the revocation of the charter and the dissolution of the corporation.

[1] *Big Business: Its Growth and Its Place.* Twentieth Century Fund, Inc., 1937, p. 17.

The doctrine of personal relationship, which is stressed in partnerships, is not so important in the case of corporations; the association of capital is of greater significance than the association of individuals, and this finds its exemplification in the right of each of the owners in the corporation — *i.e.*, each shareholder — to transfer his property rights in the corporation at will, without seeking the assent of the other owners, either individually or collectively. It follows that the death or disability of one of the owners (shareholders) in the corporation does not bring about its dissolution. This makes the corporation a more permanent form of organization.

Another distinctive feature of the corporation, as compared with the partnership, is the *limited liability* of the owners in the corporation for corporate debts. This is not a necessary mark of the corporation; there have been corporations the shareholders of which have had full personal liability for corporate debts, but it is usual at the present time to limit the liability of the owner to his investment in the business. A man who has purchased 500 shares of stock in a corporation is liable to lose what he paid for them, but the law does not lay hold of his personal resources to satisfy the creditors of the corporation, except in certain special cases where the law provides for *double liability*, as it formerly did in the case of stock issued by national banks. In this instance the owner of a $100 share of stock in a national bank was not only subject to the risk of losing the money he had invested but could also be required to contribute $100 more to meet the debts of the bank.

Finally we should note that the conduct of the affairs of the corporation is delegated by the owners to hired managers. The shareholders elect representatives from among themselves, called *directors*, to care for their interests and to supervise in a general way the operation of the business, but the active management of the corporation falls upon the officers, who are elected by the directors. The board of directors is the policy-forming body. The stockholders do not participate directly in policy forming or in the routine conduct of the

business, though there are certain actions of the directors which law and custom require the stockholders to approve.

Capital stock. The capital stock represents ordinarily the original and permanent investment of the owners in the corporation. Suppose that a group of men incorporate the Steel Manufacturing Corporation and agree to contribute $1,000,000, divided into 10,000 shares of $100 each. A takes, let us say, 5,000 shares and pays $500,000 into the treasury for them; B takes 1,000 shares, and so on until all the shares are sold. The corporation now owns $1,000,000 in cash, and the organizers of the corporation have certain property rights in the corporation. The collective name for these property rights is *capital stock*, and if we say that the capital stock of the company is a million dollars, it is an elliptical way of saying that the value of the property rights which the owners of the corporation hold in the corporation (by virtue of their original investment) amounts to a million dollars.

Each owner of a share of stock, or *shareholder*, becomes a part owner in the corporation, and as such is possessed of certain rights in the corporation, such as the right to receive dividends when legally declared, the right to share in the resources of the company if it is liquidated, and usually the right to vote at the meetings of the shareholders. A *certificate of stock*, signed by the proper officials of the corporation, serves both as evidence of the shareholder's rights of ownership and, when properly endorsed, as a convenient method of transferring his rights to another. It must be emphasized that a stock certificate is not synonymous with a share of stock; a share is one of the units into which the total property rights of the owners have been divided, and the certificate is merely presumptive evidence that a part of the total rights is vested in the holder. This is an example of the distinction between a property right and the document in evidence thereof, which was pointed out in Chapter I.

Many corporations find that their financial needs are best met by the issue of two types of capital stock, which differ

from one another in some important respects. These are *common stock* and *preferred stock.*

Common stock usually carries with it the exclusive right of voting at shareholders' meetings. Preferred stock also may confer the right of voting, but usually the preferred stockholder has no voice in the management so long as his dividends are regularly paid. The ownership therefore of more than one half of the outstanding common stock will give control of the corporation, since the owner has one vote for each share of stock he holds. In fact in a large corporation a much smaller percentage will usually suffice because of the lack of interest on the part of a number of the small stockholders and of division among the larger ones.

Dividends are declared on common stock by action of the board of directors when profits have been earned and when it is judged wise to distribute them, and the amount or the rate of the dividend is entirely within the discretion of the board. Hence the stockholders in a new corporation are uncertain of a return on their investment for some time, conservative policy dictating that the earnings be reinvested in the business, but as the corporation becomes more and more successful the dividend rate may rise until it is much higher than the rate of yield on any other type of security.

Preferred stock differs from common stock in that a fixed rate of return is promised, and in view of this the rate of return is somewhat lower than that which can be expected eventually from common stock. Furthermore dividends must be paid on preferred stock before any can be declared on the common stock. Ordinarily the preferred stockholder has no redress if the company is unable to pay dividends in any given year, and if during the succeeding year very large profits are made he does not then receive dividends for the previous year. But there is frequently *cumulative* preferred stock, upon which back dividends must be paid up before any dividends can be declared on the common stock. This protects the preferred stockholder against manipulation of the accounts to

conceal profits and avoid thereby the paying of dividends. Preferred stock may also be *participating*. This means that when a dividend has been paid to the common stockholders at the same rate as to the preferred stockholders, both the common and the preferred stockholders share in proportions already determined in any further dividend payments made for that dividend period.

The owner of this type of stock has preference over the common stockholder in another direction also; *i.e.*, in the distribution of assets in case of liquidation. The common stockholder receives his share of the assets only after the preferred stockholder has been paid in full. If the remaining assets are insufficient to redeem the common stock in full, the amount is divided pro rata among the shareholders; if there is more than enough to redeem the common stock, then the remainder is usually divided pro rata among the common stockholders.

Bonds and notes. A part of the funds for the operation of the corporation are often secured through the sale of its bonds and notes. A bond is a certificate of indebtedness issued by a corporation, bearing interest at a stipulated rate and payable at stated times, and having a certain number of years to run before the corporation is obliged to redeem it. Essentially it differs from a promissory note only in the length of time before maturity, the type of security on which it rests, and the greater formality with which it is issued. The notes of a corporation, apart from the promissory notes of a few months' tenor, partake of most of the characteristics of the bond. The majority of them however run for a few years only, whereas bonds may not mature for as many as forty years after their issue. The notes frequently pay a higher rate of interest than do the long-time bonds. This arises from the fact that note issues are resorted to at a time when the ruling rate of interest is too high to make it desirable to issue bonds; by the time the notes mature it is expected that the interest rate will have fallen so that, if necessary, long-time bonds may be sold to replace them.

The bondholder has no voice in the management of the business so long as his interest payments are met with regularity, except in so far as his contract with the corporation may limit its freedom of action in respect of matters which have a bearing on its ability to pay the principal or interest of the bond. If the security behind the bonds is a mortgage on the physical equipment of the concern, as is frequently true, default in the payment of either interest or principal may force the bondholders to foreclose the mortgage and take over the operation of the business to protect their own interests. In case of the liquidation of a business the bonds are redeemed before either preferred or common stockholders receive anything.

Stocks and bonds compared. The essential distinction between bonds and stocks is that the former represent a definite contract on the part of the corporation to pay money to the bondholders, whereas stocks only entitle their holders to receive their shares of such business profits as may be earned and their shares of the corporation's net assets upon liquidation. As regards the risk of failure to receive payment of income and return of the amount invested, the bonds of any given corporation are obviously safer than its stocks. On the other hand, they carry no chance of very large returns (beyond the fixed interest rate). Finally they give no control over the conduct of the corporation's business. Preferred stocks carry more risk than bonds but give the chance of a somewhat larger return; they generally give no control. The common stockholders take the greatest risk but enjoy the chance of the highest return; and they have usually complete control. Of course this does not mean that bonds are always safer than stocks. The stocks of a sound, well-managed corporation would carry less risk of loss than the bonds of a corporation whose assets were of questionable value or whose management was inefficient or dishonest.

There is however another risk to which all investors are subject; *i.e.*, the risk of a decline in the purchasing power of

the monetary unit. Here the bondholder bears the greater risk. Both his income and his principal are defined in the contract with the corporation in terms of dollars. If the general level of prices should rise, the bondholder's command of commodities would decline. The shareholder, either by an increase in the dividend rate or through an appreciation in the market value of his securities, may be left in a financial position substantially unaltered or even improved. By the same token, the bondholder stands to gain from a fall in the general level of prices which is not violent enough to threaten the soundness of the business. The stockholder gambles more than the bondholder on the success of the enterprise. The bondholder gambles more than the stockholder on the price level.

The property account of a corporation: Assets and liabilities. Many of the features of fundamental importance in connection with the corporation can be illustrated to best advantage by the use of a simple property account, or balance sheet. A balance sheet is a record, as of a particular time, of a real or fictitious person's *assets* and *liabilities* and of the relationship between them. Assets consist of wealth and property rights against other free persons and their wealth; a person's liability is a property right of another against him or his wealth. Thus if A owns a house, it appears on the asset side of his account; his promissory note to B for $500 appears on the liability side of his account and on the asset side of B's account.

Referring to our former example we will assume that the Steel Manufacturing Corporation has been incorporated with a capital stock of $1,000,000, all of which has been paid for in cash. The balance sheet would then appear as follows:

BALANCE SHEET OF STEEL MANUFACTURING CORPORATION

July 1, 1930

Assets		*Liabilities*	
Cash	$1,000,000	Capital Stock	$1,000,000

Before there can be a balance sheet, a fund of assets must have been segregated under a common ownership and control. These assets are owned by the corporation. Thus cash appears on the assets side of the balance sheet above. The total of claims against the assets must be equal to the total fund of assets. These claims are called *equities* or *liabilities.* The corporation's capital stock is such an equity, since it represents the stockholders' claims against the assets; it therefore appears on the liability or equity side of the balance sheet.

Now let us assume that the corporation markets $500,000 worth of its bonds, and spends $1,200,000 for plant, including machinery, etc. Its balance sheet would then read:

BALANCE SHEET OF STEEL MANUFACTURING CORPORATION

July 31, 1930

Assets		*Liabilities*	
Cash	$ 300,000	Bonds	$ 500,000
Plant	1,200,000	Capital Stock	1,000,000
	$1,500,000		$1,500,000

We have now two types of liabilities — the capital stock, which is a property right of the owners, and the bonds, which are property rights of outsiders.

During the next ten years the plant is in operation making and selling the goods for which it was built. It has sold some of its goods and has purchased some of its raw materials on credit; it has on hand a supply of materials of various kinds, raw and finished, and it has increased its investment in the plant. Furthermore the years have been profitable, and we find this reflected in the balance sheet of July 1, 1940.

Some of the items on this balance sheet need explanation. The items "accounts receivable" and "notes receivable" among the assets are property rights which the corporation has against other persons and which are evidenced respectively by entries on the books of the corporation and by promissory notes. Presumably they arose from the sale of goods. The

BALANCE SHEET OF STEEL MANUFACTURING CORPORATION

July 1, 1940

Assets		*Liabilities*	
Cash	$ 102,560	Accounts Payable	$ 60,000
Accounts Receivable	75,000	Notes Payable	50,000
Notes Receivable	25,000	Interest due Bondholders	15,000
Raw Materials	100,000	Bonds	500,000
Finished Goods	150,000	Capital Stock	1,000,000
Plant	1,800,000	Surplus	627,560
	$2,252,560		$2,252,560

corresponding items among the liabilities refer to similar obligations to others on the part of the corporation.

The other item which needs explanation is the surplus. It will be noted that since the date of the last statement our assets have grown from $1,500,000 to $2,252,560, an increase of $752,560. At the same time claims of outsiders against the corporation have grown from $500,000 to a total of $625,000 (bonds, interest due, notes payable, and accounts payable), an increase of $125,000. The increase in total assets exceeds the increase in liabilities to outsiders by $627,560. This represents a gain from all the corporation's business, and this gain belongs to the stockholders. The corporation can never have assets greater than its liabilities; they must always balance. To indicate the growth in the value of the stockowners' holdings we might possibly add the $627,560 to the capital stock, but it is desirable to leave this item intact, in part to show the amount of the original investment, and partly for technical reasons. This new item is therefore introduced to represent the growth in the owners' share in the assets of the corporation.

It should be emphasized that surplus is not cash or money or any form of wealth. It is a property right. The sum of capital stock and surplus gives the accountants' estimate of the total worth of the property rights of the owners against the corporation and is known as the *owners' equity* in the cor-

poration, or the *net worth* of the corporation to the owners.[1] The net worth may always be found by subtracting the liabilities to all others than the owners from the total of the assets. The reader will observe that, in any property account or balance sheet, the asset side is a list of property rights against wealth and persons, whereas the liability or equity side shows to whom these property rights as a whole belong. The liability side is thus a record of *proprietorship*.

Losses and insolvency. The relationship between these items representing ownership may perhaps be made clearer by taking a case of declining assets. Referring again to the last statement, on page 46, let us suppose that a note for $10,000 proves to be uncollectible, making it necessary to reduce the notes receivable and the total assets by $10,000. Losses fall first upon the owners of the business, and we will therefore reduce the surplus by $10,000 to make the two sides balance. A fire occurs and wipes out the plant entirely, causing a loss of $500,000 in excess of the insurance. The total assets will be decreased by $500,000, and another $500,000 will be subtracted from the surplus and from the total liabilities or equities.

Losses can continue to occur without affecting the outside creditors' position until the point is reached where capital stock and surplus have been wiped out entirely. When that point is passed, the corporation is said to be insolvent. In precise terms a corporation is insolvent when its assets are less than its liabilities to others than the owners. When the assets are equal to or greater than the liabilities to outsiders, the corporation is solvent.

[1] The word "capital" is a much abused term, used in various senses which have become so entrenched in popular and even scientific language that we unfortunately cannot restrict it to one meaning without too violent a break with common usage. We have already defined capital as one of the factors of production (see Chapter II). We must now note that the term capital may be used also to designate what is in the text called the "net worth" of a corporation or of a partnership or individual. This should always be distinguished from the "capital stock" when it is not the same as the net worth. And finally we may use the same overworked term to designate all the resources at the disposal of a business enterprise, incorporated or not; that is, the total assets.

The value of stock. It has been customary to assign a definite value to each share of stock and to engrave this value on the stock certificate. This value is known as the *par value*, and the par value, multiplied by the number of shares outstanding, gives the figure at which the capital stock is carried on the books of the corporation. While there are certain conveniences which attach to this custom, it has tended to confuse certain investors and others as to the true worth of a share of stock as reflected in either its actual or potential earning capacity. To engrave on a certificate of stock the words "Shares, $100 each" may be but the expression of a pious hope; the shares may be worth no more than the paper on which they are printed, or they may be worth much more than $100 per share.

A more reliable index of real worth than par value is found in the *book value*, which gives the accountant's view of the value of each share of stock. This is easily obtained by taking the value of the owners' equity, as shown by the balance sheet, and dividing it by the number of shares. According to the last balance sheet above, the owners' equity was $1,627,560, and there were 10,000 shares outstanding; each share must then have a book value of $162.756. The book value fluctuates with the increase or decrease in the owners' equity; large earnings swell the surplus and send the book value up, while losses, which of course must be borne by the owners, decrease the surplus and cause book value to decrease.

The book value will represent the real value of stock to the extent that the accountant's estimate of the value of the assets is correct. In some items there can be little possibility of error; if there is cash on hand or deposits in the banks, the accountant has generally little latitude. But in determining the figures at which the plant shall be carried in the statement and in estimating the proportion of bills and notes receivable which are uncollectible there is room for considerable error. The conservative accountant has a tendency not to overvalue the assets, in order to be on the safe side. Some accountants

do not attempt to make current valuations of the more permanent assets, such as land and buildings, but adhere for long periods to cost. On the other hand, the fly-by-night concern may greatly overvalue the assets in the attempt to bolster up the selling price of its stock, or to improve its credit standing, or possibly to conceal transactions which have "milked" the corporation for the benefit of insiders. The book value, though in general much more significant than the par value, may be greater or less than the real value.

A third estimate of the value of a share of stock is its *market value*, the selling price of the share in the open market. This is based on the estimate of the dealers in the market of the earning power of the corporation. If there is an active market for the stock of a particular corporation, enough interest will be manifested by investors and professional traders in the affairs of the corporation so that the market value, aside from temporary speculative fluctuation in either direction, may be taken as a fairly trustworthy indication of its true value. It represents the appraisal of those who may be presumed most competent to judge.

Interest payments and dividends. We have already stated that the payment of interest on bonds is a part of the contractual relation between the corporation and the bondholder. If we assume the interest rate in our example above to be 6 per cent per annum, payable semi-annually, the corporation evidently has a charge of $15,000 to meet every six months. The payment of $15,000 for this purpose would change our balance sheet as follows: (1) cash would be decreased by $15,000, and (2) the item "interest due bondholders" would be wiped out entirely.[1] A simultaneous and equal change in

[1] From the fact that the item "interest due bondholders" in the statement of July 1, 1940, is $15,000, the exact amount of the semi-annual interest charge, we infer that this is the date on which the semi-annual payment of interest is due. On any other date this item would be something less. Immediately after the payment it is zero, as stated in the text; thereafter it increases gradually, as the liability "accrues," until it again reaches $15,000 on the date when payment is due six months later.

both assets and equities leaves the totals equal, and no further change in the statement is required.

Let us assume next that the directors decide to pay a 5 per cent dividend to the stockholders. This means 5 per cent on the capital stock, which is $50,000, or $5 per share of stock. Cash would be decreased by $50,000, and surplus would be similarly decreased. The owners' equity is reduced to $1,577,560, and if we calculate the book value of each share of stock we will see that it is now $157.756 (*i.e.*, $1,577,560 divided by 10,000); in other words it has decreased by precisely the amount paid in dividends on each share of stock.

Stock dividends. Suppose the Steel Manufacturing Corporation decides to declare a *stock dividend* of 50 per cent. Each shareholder will receive half as many shares as he now possesses. A, who owns 5,000 shares, will receive 2,500, and the other shareholders in proportion. Clearly this does not disturb the assets in the slightest; so far as our statement is concerned all the changes will be among the equities. Capital stock will be increased to $1,500,000 and the surplus of $627,560 will be reduced to $127,560. The net worth of the business is not altered, but the book value of each share of stock is decreased, for now we divide the net worth by 15,000 instead of by 10,000.

It is quite evident that the position of the stockholders has not changed at all so far as their claims upon the assets of the corporation are concerned. The 5,000 shares which A owned had a book value before the stock dividend of $162.756 per share, or a total value of $813,780; now he holds 7,500 shares, each having a book value of $108.504, or a total value of $813,780. The book value per share of stock has declined, but the total book value is not affected. In fact the stock dividend, except in a figurative sense, is not a dividend at all, but merely a transfer from surplus to capital stock.

Since the stock dividend makes no change in the position of the stockholder with respect to the corporation, it may be asked what purpose it can serve. Without attempting a com-

plete answer, the two most common purposes may be cited. In the first place, it may happen that the market value of the stock is so far in advance of the par value that a share of stock is less easily traded in. Five shares which sell for $100 each are much more convenient than one share which sells for $500. The stock dividend may then be resorted to in order to increase the number of the shares and lower the value of each. But the chief motive is probably to avoid the appearance of unusually high profits and high dividend rates. A dividend of 20 per cent on a capital stock of $1,000,000 is no greater than a dividend of 10 per cent on a capital stock of $2,000,000, but it seems greater; the former may attract considerable notoriety, lead to the suspicion of monopoly, and invite competition, while the other passes by without exciting comment.

Watered stock. Stock is said to be "watered" when the corporation issuing it has not received full value for it. There are several different ways of watering the stock. One of the most common is to issue stock in payment for property or services which are overvalued. This may be done innocently and in good faith, or it may be done with intent to defraud. Suppose that $100,000 worth of stock, par value $100 per share, is sold to investors for cash, and that another block of $100,000 is given in payment for a tract of land which, at the most liberal estimate, is worth not more than $50,000. The statement would show assets of $200,000 and liabilities of $200,000, and the book value of each share would be $100. A true record on the books of the corporation of the value of the assets would however record them as worth $150,000 and would thus show a book value of only $75 per share. A stock dividend without a sufficient surplus to justify it is a form of stock watering. In our example on page 50 a stock dividend of a million dollars could have been declared only by raising the recorded value of the assets by $372,440.

The essence of stock watering is overvaluation of the assets of the corporation, and it may be resorted to for any of the purposes stated in a previous paragraph dealing with book

value. Like the legitimate stock dividend, its purpose may be to conceal large profits, though in this case it really conceals them instead of merely avoiding an appearance of high profits. Whatever its purpose, stock watering involves falsification of accounts and carries at least the presumption of dishonest intent.

The corporation's income account. The balance sheet or property account of a corporation shows the condition of business at a given instant of time, but it does not give all the information that may be required regarding the condition of the business. Even a comparison of two such accounts for successive years, while more informing than a single balance sheet, does not tell the whole story. It does indeed show the changes in the various items in the balance sheet and it may record the profitableness of the business, but it does not tell us how the changes have come about, nor what has occasioned the profits or the losses. To complete the picture we must have another account, called the *Income Account* or the *Profit and Loss Statement*, which gives in condensed form the history of the operations of the business over the accounting period in question.

In order to illustrate the content, form, and meaning of a typical profit and loss statement, let us assume that the City Grocery Company is dealing in eggs solely. On January 1st it had on hand eggs which cost it $40,000. During the year it purchased eggs costing $171,200, and it paid out in freight and cartage $3,100 to bring these eggs to its warehouses. Its sales of eggs amounted to $215,000, and at the close of the year it had on hand eggs which had cost $42,300. Tabulating these figures (see page 53) we get a first approximation of the operating profit of the business for the year, or the *gross profit*.

The gross profit obtained by deducting from net sales the cost of the goods sold is not the final net profit actually earned; other expenditures have been made and must be deducted. Expenditures for advertising, salesmen's salaries, commissions, and the maintenance of a delivery service are necessary to

Net Sales		$215,000
Deduct: Cost of Goods Sold		
Inventory Jan. 1, 1939	$ 40,000	
Purchases	171,200	
Freight & Cartage Inward	3,100	
	$214,300	
Less: Inventory Dec. 31, 1939	42,300	
Cost of Goods Sold		172,000
Gross Profit		$ 43,000

keep the merchandise moving from the warehouse to the consumer. Expenditures of this type are classified as *Selling Expenses* and must be deducted from gross profit. They cannot be included in the cost of goods sold because, for the most part, they do not relate to any particular sale. Commissions could conceivably be deducted from selling prices, but convenience and uniformity demand that they be grouped with the other selling expenses.

In seeking the final net profit from operation it is also necessary to take into account another major group of expenses known as *General and Administrative Expense.* Such are the salaries to executives and to the office employees who keep the books and accounts, send out monthly statements, make collections, adjust complaints, supervise the purchasing of merchandise, etc. Office supplies of all kinds are consumed. Moreover provision must be made for warehouses, offices, show rooms, etc., with all the necessary equipment, and depreciation on all these assets must be treated as an expense. Characteristic of these expenses is the fact that they relate almost entirely to periods of time rather than to the volume of sales. Depreciation occurs, and rent and taxes accrue at a uniform rate whether sales be large or small. Deducting selling expense and general administrative expense from gross profit gives *net profit from operation* or *net profit on sales* (see page 54).

Logically the profit and loss statement should stop at this point, but in every business there are incidental receipts and payments usually financial in character and not a part of the principal business of the corporation. Space may be leased

to an outside tenant; interest may be received on investments; rent or interest may be paid; discounts may be earned or given, and the like. For want of a better title such income is called *Other Income* and such expenses *Other Expense*. Adjusting the net operating profit for these additional items gives the *Net Profit* for the year.

The profit and loss statement of the City Grocery Company can now be presented in the usual form, and its meaning should be clear. It will be noted that, instead of giving the

PROFIT AND LOSS STATEMENT

City Grocery Company

(For the year ending December 31, 1939)

Net Sales			$215,000
Cost of Goods Sold		$172,000	
Selling Expenses:			
Advertising	$ 5,000		
Salesmen's Salaries	10,000		
Delivery Expense	3,000	$ 18,000	
General and Administrative Expense:			
Salaries	$11,200		
Office Expense	3,240		
General Expense	2,000		
Depreciation	1,500		
Taxes	800	$ 18,740	$208,740
Net Operating Profit			$ 6,260
Add: Other Income:			
Interest Earned		$ 300	
Rental Income		240	540
			$ 6,800
Deduct: Other Expense:			
Interest Paid		$ 1,100	
Loss on Sale of Real Estate		700	1,800
Net Profit			$ 5,000
Appropriation of Net Profit:			
Dividends Declared		$ 3,000	
Carried to Surplus		2,000	
Total Net Profit			$ 5,000

details of calculation of gross profit, the final figure for cost of goods sold is given and combined with other groups of costs in determining the total operating expenses.

A comparative balance sheet for the beginning and ending of the accounting period is given here.

COMPARATIVE BALANCE SHEET

City Grocery Company

(As of the close of business December 31, 1938, and December 31, 1939)

	Dec. 31, 1938	*Dec. 31, 1939*
Assets		
Cash	$ 5,000	$ 4,200
Accounts Receivable	30,000	32,500
Inventory	40,000	42,300
Fixed Assets	25,000	31,000
	$100,000	$110,000
Liabilities		
Accounts Payable	$ 26,500	$ 34,500
Mortgage Payable	10,000	10,000
Capital Stock	50,000	50,000
Surplus	13,500	15,500
	$100,000	$110,000

From this we see that the net worth of the business has increased by $2,000, but the balance sheet does not show us what brought this about. Nor does it tell us that $3,000 was earned, declared, and paid out in the form of dividends. For this and other facts we must have recourse to the income account. Neither the balance sheet nor the income account by itself is sufficient; together they show us in what respects the business has changed and what has caused these changes.

EXERCISES

1. Arrange the following items in the form of a statement of the assets and liabilities of a corporation: land and buildings, $90,000; capital stock, $100,000; bonds outstanding, $50,000; machinery and equipment, $75,000; government bonds, $5,000; cash, $7,000; surplus, $66,740; taxes due, $3,260; accounts receivable, $12,000; raw materials on hand, $26,000; finished goods on hand, $20,000; accounts payable, $15,000.

2. There are 1,000 shares of capital stock outstanding. What is the par value of each share; the book value?

NOTE. In this and each of the following five exercises, refer to the original statement as in Exercise 1.

3. A cash dividend of 6 per cent is declared and paid to the stockholders. Make the necessary changes in the statement. What is the book value per share of stock now?

4. The company declares a stock dividend of 50 per cent. Make the necessary changes in the statement. What is the new book value per share? Stockholder A held ten shares of capital stock before the stock dividend was declared. Calculate the book value of his holdings of capital stock before and after the stock dividend.

5. A fire breaks out and destroys the stock of raw materials and the warehouse, valued at $5,000, in which they were stored. These losses are not covered by insurance. Make the necessary changes in the statement. Is the company still solvent?

6. Make such changes in the balance sheet as would be necessary to illustrate stock watering.

7. The general ledger of the Bolter and Hammer Hardware Co. contains the following income and expense accounts:

Cost of Goods Sold	$32,650.00
Advertising	250.00
Salaries of office force	1,240.00
Workshop income	110.00
Interest paid	125.00
Sales (net)	47,375.00
Wages of sales force	4,650.00
Office supplies and postage	70.00
Delivery expense	950.00
Interest received	40.00
Rent paid	1,200.00
Miscellaneous expenses (general)	500.00
Depreciation of store building	130.00
Dividends declared	2,500.00
Taxes	559.50

From the above facts prepare a profit and loss statement for the company for the year.

CHAPTER IV

Large-Scale Production and Combination

The growth of large-scale production. An outstanding feature of industry during the past hundred years or more has been the continued growth in the size of the producing unit. Taking the cotton manufacturing industry of the United States as an example, we find that between 1859 and 1925 there was an increase of 55 per cent in the number of cotton manufacturing plants, the average number of workers per plant had more than doubled, the average primary horsepower per establishment (an index of the use of power-driven machinery) had increased enormously, and the average value added to the product per establishment had increased by 334 per cent. If we were to examine the statistics showing the growth of other industries we would find differences in detail but the same general trend — a great increase in the average number of workers, the average amount of capital, and the average output per establishment.

Attention should be directed also to the importance which the very large establishment has in our present-day industrial life. According to the figures of the 1930 census industrial establishments with an annual product valued at $1,000,000 or over formed only 5.6 per cent of the total number of such establishments, but they were responsible for 69 per cent of the total product in value, and employed 58 per cent of the total wage earners in industry. On the other hand, establishments with an annual output of less than $100,000, while actually forming 68 per cent of the total number of producing units in the year 1929, accounted for only 6.2 per cent of total industrial production and employed but 10 per cent of industrial labor. In 1929 the establishments employing from one to five laborers comprised 45.4 per cent of the total number,

whereas the establishments employing over 500 workers formed 1.3 per cent of the total; these large establishments however employed 37.8 per cent of the total wage earners in industrial plants, as contrasted with the 3.2 per cent employed by the establishments in the former group. Establishments employing more than 1,000 laborers formed only 0.47 per cent of the total number of establishments, but they employed 24.4 per cent of the total number of workers.

Two conclusions of considerable importance can be drawn from these figures. The first is that small-scale industry is still of significance in our national life, particularly to supply local markets. "Big business" has not entirely supplanted the small producer, even in the lines of industry where the advantages of large-scale production seem most marked. On the other hand, it is true that the large unit has been growing, both absolutely and relatively, at the expense of the smaller-sized units. Production for national markets has been supplanting production for local or regional markets, and with this change has come the decline of the small unit and the substitution of the larger.

Conditions necessary for large-scale production. One of the first prerequisites for large-scale production is obviously a product which is in great demand, for unless a large output can be marketed profitably there is no scope for increased refinements in the productive apparatus. Another requirement is that the article be not one which is demanded because of the finish which hand labor may give it. It must be amenable to the machine process, capable of standardization, and one which does not entirely lose attractiveness in the eyes of the purchaser thereby. The finer grades of clothing and many articles of personal adornment and luxury can never be produced by large-scale methods.

Economies of large-scale production. Large-scale production, where it is found, exists for the simple reason that it is more efficient than small-scale production, and by this we mean that the larger unit effects economies in production

which make it possible to sell the goods at a lower price. This not only extends the range of commodities over which the consumer has command and therefore enlarges his living, but it is advantageous in a more profound sense, for it implies the more effective utilization of the limited supplies of labor and capital.

Power and machinery. Large-scale production gives rise to economies in the use of power-driven machinery, both as to power and as to machinery. The saving in power comes from the lower unit cost of installing and operating a large power plant. Though the large plant costs more than the small one to install, it costs proportionately less, and greater efficiency in the production of steam results. Hence the cost of power to be attributed to each finished article is less.

Machinery can be used for more of the processes in a large factory than in a small one, for if a machine is to be used to greatest advantage it must be used continuously. If the machine is used only a part of the time, the cost charged against each unit of the good undergoing the particular process may make its use more expensive than hand labor. A large factory is able to employ machines which, though they may save only a fraction of a cent per unit produced over the cost of hand labor, are still worthwhile because of the large volume of production. This is one way of stating the fact that large-scale production permits of the greater application of the principle of division of labor. There can be greater specialization by workers, and processes can be more finely divided and therefore more profitably subjected to machine methods.

By-products. Another economy lies in making by-products available for use. The meat packing industry offers one of the most illuminating examples of this. According to figures given by Swift and Company, by-products such as hides, fats, etc., form 45.7 per cent of the total weight of a live steer. A list of the finished goods which are made in whole or in part by Swift and Company from these by-products includes the following: glue, fertilizer, oleomargarine, lard compounds, soaps, glycerine, leather dressing, animal feed, buttons, combs, gela-

tine, ice cream, candy, pharmaceutical products, snuff containers, gold beaters' skins, and sausage containers. In the early days the materials from which these goods are made had to be carted from the city and buried; so far from being a gain, they were the occasion of considerable expense.

The economy which results is plain; any net returns obtained from the by-products diminishes by so much the price which must be obtained for the principal product in order to cover the expenses of production. The meat packers for example often pay more for live cattle and hogs than they receive for the dressed meat; the difference is made up in the return from by-products.

Technical experiment and research. A large company is better able to conduct experiments than a small company. Experimentation is a risky and sometimes an expensive affair. Many trials are necessary before a device or a process is developed to the point where it is of practical use. For a small concern the financial burden may be too heavy to carry, whereas the company with greater resources can carry the load through the lean years, trusting that in the long run the improvement will more than pay for itself.

Economy in administration. Within limits the cost of administration and supervision is not much greater for a large factory than for a small one, certainly not proportionally greater. The head of a department can generally take care of a large department about as effectively as of a small department. The chief additional expense consists in a larger clerical force. Similarly the cost of buying raw materials and selling finished goods may be decreased per unit. Selling in large quantities means a real saving to the seller in packing, shipping, and sales costs, and a lower price can therefore be quoted to one who buys on a large scale. In selling the finished good there is better opportunity to coördinate the activities of salesmen, thus covering a given territory without overlapping and with greater effectiveness, while national advertising can be used with greater profit.

The field of large-scale production and its limits. Manufacturing is the field in which large-scale production obviously finds its greatest opportunities and from which may be drawn the most illuminating illustrations of its principles. Many other lines of industry have however proved themselves amenable to large-scale production. In transportation, banking, and marketing the size of the organization has greatly increased in the last century. Doubtless this increase in size has come partly in response to growth in the size of the manufacturing establishment and is partly a cause of that growth. In any event concentration of manufacturing in large plants requires more effective service than could be afforded by the transportation, banking, or marketing system of a century ago. The one great industry which has not proved a profitable field for the application of large-scale methods is agriculture. There are of course farms where hundreds of workers are employed, but they are the exception and not the rule.

Large-scale production has its limits, even in the fields most favorable to it. Probably for each industry there is a plant of optimum size, beyond which it is unwise to attempt to go. A part of this difficulty lies in the physical awkwardness of handling a large bulk of goods under one roof, but a still greater difficulty is the weakness in management which may appear. Since this weakness is common also to combinations of several plants under one management, its further discussion may better be postponed until after our investigation of the problem of industrial combination.

Combination. Rarely does large-scale production as such give rise to problems of vital concern to society or raise question of regulation in the public interest. Sentimentalists deplore the elimination of many small shops and the passing of the independent craftsman, but while we have undoubtedly lost thereby a distinctive and virile element in our economic life, few would seriously suggest that we sacrifice the greater gains which have come from mass production in order to try to recapture the spirit of an earlier day.

This however is not always true of the *combination*, which we may define as *a union of several industrial units under a common management*. There are indeed serious problems of contemporary interest and importance raised by the great corporations, the combinations, and the combinations of combinations.

The large corporation. We have become so habituated to the rôle played by the great corporations that paradoxically perhaps it may be a surprise to discover the extent to which they control our economic life. According to one estimate the two hundred largest non-banking corporations in the United States had at the beginning of 1930 total assets of $81,074,000,000; the assets of all the non-banking corporations, something over 300,000 in number, amounted to $165,000,000,000. Almost half of the assets of all non-banking corporations were in the control of less than one per cent of the total number of such corporations. The same authorities also estimate that these corporations owned 38 per cent of all business wealth, other than banking, and 22 per cent of all national wealth.[1]

A somewhat different picture of the importance of large corporations is given in the table on page 63, which shows the number of corporations in selected lines having in 1929 an income of $1,000,000 or more and the percentage of total net income in the given line which accrued to the larger corporations.[2]

In recent years the large corporations have shown a growth which is much more rapid than that of the small or medium-sized corporations. In the years 1907 to 1927 the assets of the two hundred largest non-financial corporations increased more than twice as rapidly as the assets of other such corporations; they reinvested a larger portion of their earnings; they

[1] For a further discussion of this topic and for an illuminating analysis of the social implications of the position of the modern corporation the student is referred to Berle and Means, *The Modern Corporation and Private Property*, New York, 1933, from which these figures are taken.

[2] Lewis Corey, *The Decline of American Capitalism*, New York, 1934, p. 387.

CONCENTRATION OF CORPORATE CONTROL

Industry	*Number of corporations with net income of $1,000,000 up*	*Net income (millions)*	*Per cent of all net income*
Manufactures	627	$3,338	64.0
Mining	65	278	84.6
Public utilities	230	1,805	86.0
Trade	93	316	27.5
Service	31	108	34.4
Finance	283	1,048	47.7
Total	1,329	$6,893	60.5

secured a larger share of new capital by floating securities; and mergers were more common and more significant among them.[1]

The formation of combinations. Many a corporation has become a combination quite simply and naturally when it was found profitable to build a new plant from earnings or from funds secured by the floating of additional securities. Some combinations came into being when one company purchased the property of another company and then completely merged the two into a larger single company. The American Tobacco Company, formed in 1890, exchanged its stock for the plants, business, brands, and good-will of five companies, and the American Sugar Refining Company, organized in 1891, is another example among the many that might be cited.

The most common, the most effective, and the least expensive method of building up a large combination is however through the holding company. A corporation which buys the controlling shares of stock in another company, either with its own stock or with cash, becomes in effect a holding company. Frequently a new corporation is chartered, as in the case of the United States Steel Corporation, to buy up a controlling number of shares in the companies which are to be combined. Little capital is required to start a holding company, since the shares of stock of the member companies may usually be obtained in exchange for the stock or bonds

[1] Gardner C. Means, "The Growth in the Relative Importance of the Large Corporation in American Economic Life," *American Economic Review*, October, 1931, pp. 10–42.

of the holding company. Unity of policy may be obtained by the election of the proper men as directors in the constituent companies.

Motives leading to the formation of combinations. In large measure combination has been resorted to as a means of securing substantial control of a particular field of production in order to obtain monopoly profits. Two other motives however have been prominent. Of these the first has been to make possible speculative profits through the issuance of large volumes of stock; the second has been to obtain even greater economies in production than were possible simply through large-scale production.

We have therefore three distinct problems. The first, and in some aspects the most important, concerns itself with the growth of monopolies — the so-called "trusts" — including their effect on society and the methods by which they may be regulated in the public interest. The second involves methods used in marketing securities and the means by which the operations of investment bankers, underwriters, and the security exchanges may be controlled in order to prevent abuses. The third problem, to which the remainder of this chapter will be devoted, centres around the possible social advantages of combinations as producing units. Are there economies in production possible to a number of large plants under single management which cannot be obtained by a single large producing unit? If there are such economies and if they are of considerable significance, this fact will have an important bearing on the first problem.

Horizontal combinations: Definition. We may distinguish two types of combination, the *horizontal* and the *vertical*. As the name suggests, a horizontal combination is one which unites under a single management several plants producing the same product or products. It is a type which is quite common in the United States, being exemplified by many of the "trusts" which exist or have existed in the past, such as the American Sugar Refining Company, the United Shoe

Machinery Company, and the American Tobacco Company.

Economies in buying and selling. A horizontal combination may possibly secure a saving in the purchase of its raw materials. Buying in much larger quantities than the single plant, it may approach the producer directly and eliminate the middleman's profits, or it may secure its goods from a wider productive area and thus obtain its raw materials more cheaply. But it is to be doubted if the savings in this line are great enough to be of real consequence.

In the sale of its products there is a possibility of saving. A reduction in number of salesmen and amount of advertising together with greater effectiveness in the use of both sales media may lower the sales cost. The tobacco trust, which had been formed by the combination of several large producing units, was dissolved by a court decree in 1911, whereupon the business was turned over to a number of separate companies. "The subdivision of the business provided for in the dissolution decree led to a duplication of selling organization and an increased overhead expense; and the result was a general increase in selling costs. . . . In every branch, except flat plug and Turkish cigarettes, the selling costs showed an increase. The advertising expenditures of the successor companies also greatly increased as compared with the expenses of the trust. The advertising expenses of the trust in 1910 in all branches except cigars were $10,895,132, while those of the successor companies in 1913 amounted to $23,623,564, or more than double." [1]

Economies in operation. The greatest field for economies would appear to lie in the operation of the plants. Each plant can presumably secure maximum efficiency by producing a certain number of articles. If forced above that point or below it, the unit cost of goods is increased. With several plants operating under unified control it is possible so to arrange things that all the slack in production will fall on one or more plants,

[1] Eliot Jones, *The Trust Problem in the United States*, 1921, pp. 149–150.

leaving the others running as usual. The American Sugar Refining Company is said to have used its Brooklyn refinery for this purpose, slowing down operations there as demand slackened and speeding up in response to increased demand. In some industries the risks and uncertainties involved in production may be reduced if plants are located in different parts of the country. A labor strike may not affect all of the plants at the same time, and a tie-up of the transportation facilities of all the plants at one time is unlikely, so that if the closing of some is forced the others can be operated more intensively.

A combination which makes a variety of products is able to specialize in factories, devoting one factory to the making of a particular article, rather than allowing all factories to make all articles, and in this way the utmost advantages of specialization are attained. The devices or patented articles that have proved most successful can be made the common property of all the operating units, heightening the efficiency of each. With several plants it is possible to start competition between them, thus enlisting the enthusiasm of one manager to excel the others. Cost statistics may be available for all the plants, so that if one is unsuccessful in some respect the trouble can quickly be diagnosed and remedied. It is claimed that more efficient executives can be secured because of the wider field from which to choose. The inter-plant competition, it is asserted, will disclose men who are most capable of handling the more difficult tasks of coördination and centralization.

Saving in freight costs. The savings on cross freights are sometimes cited as proving the superiority of the combination. If under separate management there are two plants, one in New York and the other in Chicago, each may sell goods in the other's district and incur freight charges unnecessarily great. Under combination the good will be sent from the nearest point. But the importance of this saving, while real in many cases, can be greatly exaggerated, for if the good be bulky and of low value the distance it can be shipped is limited by the cost of freight, whereas if it be a small and expen-

sive object such as a watch, freight charges form such a small portion of its value as to be of little moment.

Vertical combinations: Definition. Vertical combination involves the gathering together under one management of the various processes necessary to turn out a finished good. It is sometimes called *integration* of industry and is the opposite of that process of differentiation and specialization which has been characteristic of the factory system. In some industries there has always been a certain amount of integration. Thus in the United States spinning and weaving have commonly been undertaken by the same factory, although in England they are separate industries. In the steel industry of the United States can be found the most impressive illustrations of vertical combinations, although horizontal combinations have been by no means uncommon.

Technical advantages. There are real technical advantages in combining the different stages of manufacture of steel products, advantages which result in the saving of fuel for reheating the metal, of time and labor in moving or handling material, and in the more complete utilization of by-products, such as blast-furnace gas. Further advantages may be obtained through the control of sources of iron and coal. The steel industry differs from some others in that each unit not only may use the products of another as raw material but may pass its own finished or semi-finished products on to serve as raw materials for the others. Coal and iron are essential for the blast furnace, while the ultimate products of the blast furnace are requisite in the mining and transportation of coal and iron.

The United States Steel Corporation. The extent of the process of integration in the United States can best be pictured by a brief reference to the United States Steel Corporation, one of the largest corporations in the United States, incorporated in 1901 with an authorized capital stock of $1,100,000,000. The corporation secured control of a number of other corporations by an exchange of stocks and bonds,

and at the outset it controlled eleven corporations. It is a fact of interest that almost all the companies which were united to form the United States Steel Corporation were themselves combinations. Some of them represented merely the horizontal type of combination, such as the American Steel Hoop Company, which was a consolidation of nine concerns formed with the desire to limit competition and secure larger profits. Some already showed signs of a fairly complete integration, such as the Federal Steel Company and the Carnegie Company. The United States Steel Corporation was not exclusively a vertical combination, for several of the constituent companies produced the same articles. The result of the consolidation was the creation of a giant corporation, which was practically self-sufficient so far as materials were concerned and which embraced the manufacture of practically every known article in iron and steel.

The question which immediately presents itself in this connection is whether the Steel Corporation has justified itself from the point of view of social saving. There is no doubt that the Steel Corporation has been a profitable venture; there may be doubt however as to whether the profits had their origin in economies of operation. In many instances of course the company avoided the payment of profits to others by the operation of a coal mine or a railroad itself, but so far as it merely took profits which would otherwise have gone to others this does not represent economy from the standpoint of society. So far as the economies to be obtained from the integration of manufacturing processes were concerned, they were probably already obtained to the full degree by some of the larger and better equipped among the constituent companies. The chief economies possible were from the more economical subdivision of the business among the plants by utilizing only the most efficient plants and dividing the work among the plants according to the nearness of the market. One writer in discussing this question states that "The conclusion would seem to be that the Corporation by virtue of

its ownership of the cream of the ore and coking coal lands and of the iron ore railroads (not to mention its financial connections) had an advantage over its competitors but that this advantage did not demonstrate the superior economy of the trust form of organization." [1]

Conclusions. It cannot be doubted that both vertical and horizontal combinations offer the possibility of economies over and above those which can be obtained by the single large manufacturing plant. But it is exceedingly difficult to determine the extent to which these economies have been realized in actual fact. Combination and monopoly power are so closely associated that it is not possible always to attribute the success of any given combination to the appropriate factor. It is however a fair statement of the case to say that where a combination has been conspicuously successful the profits have arisen in large part from a monopolistic position enabling the combination to control output and raise prices to the public. When monopoly is absent economies in the field of marketing are probably a more important factor than economies in production in the narrower sense.

As a matter of fact the success of the large combinations of the United States has not been striking, if success be measured in terms of profits. Not only have profits been smaller than anticipated by the promoters of the consolidations, but in many instances they have been smaller than the aggregate profits of the constituent companies in the period before consolidation. Despite the economies which combination might have been expected to bring, the United States Steel Corporation during its first year of operation earned but 78 per cent of the amount earned by the member companies in the period immediately preceding the formation of the combination, and for the first ten years of its life its average earnings were somewhat less than its earnings during the first year. Professor Dewing has made a comparison of the earnings of a number of representative combinations with the earnings of the com-

[1] Eliot Jones, *The Trust Problem in the United States*, 1921, p. 219, note.

panies which were united and has found for the group as a whole that the earnings of the first year, as well as the average earnings for the first ten years of the existence of the combinations, were materially below the level of the previous earnings of the member companies.[1]

The failure of many combinations to reflect economies of operation and management in their earnings seems to indicate that fundamental defects or weaknesses are consequences of size of organization. A combination may realize some of the economies of large-scale management only to have these offset or more than offset by losses traceable to other factors. There are many losses to which a huge combination is peculiarly susceptible, but undoubtedly the chief weakness which organizations of tremendous size (including both large unit plants and combinations) have to face is the difficulty of finding men of the calibre required to direct the work. It is possible to find many an organization which seems to work with effortless, machine-like precision, but examination will usually disclose a dynamic personality behind the organization, supplying the initiative, making the important decisions, and inspiring his colleagues with enthusiasm and confidence. The founder of almost any of our large industries has been such a man. His personal fortunes have been closely tied up with the fortunes of the business, and not the least of his incentives has been the desire to leave his business as a monument to himself. When the reins of management pass from his hands into those of another, the same combination of ability, energy, and audacity is rarely found. Business ability of the highest order is required by all large-scale production; it is essential for securing the economies of combination, but it is not sufficiently plentiful to give assurance that large combinations will be able by more efficient production to undersell the large single units.

This conclusion does not however give warrant for the in-

[1] A. S. Dewing, *The Financial Policy of Corporations*, 1934, vol. iv, Appendix. For other recent studies see *Recent Economic Changes*, 1929, vol. i, pp. 179–218.

ference that the day of the large combination is over. On the contrary many indications point to a continued growth of the largest combinations. As pointed out earlier in this chapter, the acquisition of monopoly power or the elimination of intense competition or the opportunity to secure profits from security issues may be sufficient motives for management to try to expand the scope of corporate activities, even though the result may be in opposition to the interests of society as a whole.

CHAPTER V

Division of Labor and the Control of Production

Coöperation. The key to an understanding of the modern industrial system is to be found in the idea of *coöperation.* By this is meant that the divers elements in society are consciously or unconsciously working together toward a common end.

There are historical examples of communities working together and consciously regulating their various relationships to each other. Such societies as the Oneida Community, the Amana Society, Brook Farm, and others were organized for this purpose. These examples of small isolated socialistic communities are however of only incidental interest to our present inquiry. Our task is to obtain a bird's-eye view of our present-day economic organization in order that we may see in what sense it is a coöperative system and may discover the motive forces which impel the various members of society to work together and the regulating forces which direct their efforts.

Probably the earliest and certainly the simplest type of coöperation is found where men get together to perform a task which is too great for the strength of a single man. In former years the farmers of New England worked together at a barn-raising or in lifting a millstone into place. There are and always will be in everyday life countless examples of this form of coöperation, which we may call *simple coöperation,* and which is distinguished by the fact that the several cooperating workers are all performing the same kind of work.

Division of labor. Coöperation becomes of vastly greater importance when it passes beyond this simple stage and leads to specialization, in which each worker confines himself to one kind of work, made possible for him because other kinds of work are cared for by other workers. Such coöperation is

called *division of labor.* It is of varying complexity according to the degree of specialization. The first steps occur when industry is separated into distinct trades, when one man devotes his entire time to making shoes, another to raising wheat, another to making cloth, and so on through the various trades with which we are familiar. But the mere separation of trades, in which each worker performs all the several operations of his own trade, taking the raw materials and finally turning out the finished product, has only begun to explore the possibilities of division of labor. Every trade is itself a complex of many distinct operations, and division of labor has not reached its final goal until each of these operations is made a separate task and assigned to a special group of workers. Although it did not originate there, it is in the modern factory that the more complex division of labor has been developed to its highest point. Here we see the extent to which it has been found possible to split a trade up into distinct minor operations, each performed by a different group of workers. In the making of a pair of shoes the soles are cut by one man, the various parts of the uppers by others, one part is stitched by still another, and so on. Complex division of labor may be carried on without power machinery, but its possibilities are usually greatly enhanced by the employment of machinery for many of the operations. In the making of the Goodyear welt shoes it is said that there are one hundred and eighty-five distinct operations, of which one hundred and fifty-seven are performed by machines.

Economic effects of division of labor. Simple coöperation is in fact so simple that it will require no further discussion. Our attention for the remainder of this chapter will be devoted primarily to division of labor as the really fundamental form of coöperation.

The predominant fact about division of labor is that it increases man's efficiency in production. In the first place, division of labor, by permitting each worker to specialize upon one trade or even upon a single operation in a trade, permits

him to acquire a degree of skill and dexterity which would otherwise be out of the question. Furthermore the simplification of the individual operations in a factory makes it possible for a worker to become amazingly skillful in his given operation.

Division of labor also makes possible the saving of time which the unspecialized worker wastes in putting away the tools required for one job and getting the tools for another. A certain amount of time is used in really getting started, in deciding where and how to begin on the new job.

When division of labor does not prevail, there is also the social waste which comes of leaving tools idle during a good part of the day. If analyzed, it can be seen that this is also a waste of time for the group as a whole. Assume that we have five individuals and five possible trades. If each individual follows all five trades, twenty-five sets of tools are necessary; if there is specialization, only five sets will be required, and they can certainly be made in a fraction of the time required to make the twenty-five, thus releasing labor for other employment.

Division of labor not only economizes the use of tools and machines, but it leads to the invention and utilization of complicated labor-saving machinery which would otherwise not be possible or economical. As the processes of any trade are subdivided each operation becomes more and more simple, and sooner or later it becomes possible to turn over the simpler ones to machines, thus setting free labor for other tasks and increasing the product of industry. The reduction in the physical labor of tending machines has made it possible to use women for a variety of tasks which men alone used to perform, and the decreased difficulty of the task itself has considerably shortened the time spent in acquiring the necessary skill to perform it.

Finally with division of labor it becomes possible for a man to find the work for which he is best fitted and in which he can show the greatest dexterity and speed. It by no means follows

that in our society this adaptation will always take place, and in many cases it is quite apparent that it does not take place. There is a tendency to follow customary lines of work. The son of a banker may follow banking as the line of least resistance, even though his natural bent may be in the direction of medicine. There is many an automobile mechanic who would be a more useful member of society as a farm hand, and doubtless many a farmer who is wasting his talents on the farm. But there is at least the possibility of this adaptation today, whereas formerly, as on the medieval manor, nearly every man had to be a farm laborer whether he would or not. As a special application of this principle, we have the inestimable advantage of giving full-time employment to capable business leaders, whose important place in our economic organization we have already noted. If one will think of some of the great industrial leaders — Carnegie, Harriman, and Ford, to mention only three names — and try to imagine them in the manorial period in any rank of life he chooses, he can better appreciate the extent of the opportunity which the present organization of society gives to native talent and the degree to which society profits from the exercise of such talent.

All of these results of division of labor promote production, by economizing labor, increasing its efficiency, and making more effective use of capital. With a given amount of human labor there emerges a richer flow of commodities and services for the satisfaction of human wants. Or we can look at it from another standpoint and say that division of labor makes it possible to turn out the same quantity of goods with less expenditure of effort, thus leaving more leisure for the enjoyment of the products of industry. It should be noted however that not everything that can be said of division of labor is in its favor. It is objected for example that modern specialized labor is monotonous and deadening in its effects on the workers, that modern machine tending is so largely automatic that skill is at a discount and labor is kept at a dead level, that the highly specialized laborer is unfitted for work

in another trade. There is doubtless some measure of truth in these charges, but they may easily be exaggerated. Labor in the medieval manor was hardly free from monotony, and the shorter working hours of modern industry probably turn the scale in favor of the present day as regards the deadening effect of labor. For machine labor it can be said that, automatic though it be, the demand for skilled labor is still active and that labor could scarcely be more nearly on a dead level than it was on the manor. The charge that specialized training today makes change of occupation difficult is the reverse of the truth. An operative in a woollen factory could fit into a bicycle factory more easily than a house painter could become a blacksmith. On the whole such drawbacks as are inherent in the division of labor are certainly not sufficient to prevent its net effect being overwhelmingly on the credit side of the account.

Division of labor and the extent of the market. Whether the farmer builds his own barn in addition to farming or hires a carpenter to do it for him depends in part on conditions over which he, as an individual, has little or no control. Assuming that his true bent is farming, he will prefer to hire the carpenter, but if he cannot sell the wheat that he does not need for his own family, he will have to build his barn. If the carpenter cannot sell his services he is forced to raise his own wheat, even though he be an indifferent farmer. The possibility of exchanging the surplus which one individual produces is a necessary preliminary to division of labor, and as exchange becomes easier there is a tendency to subdivide the tasks and to proceed from simple to complex division of labor. When the market consists only of the town and the countryside, the village cobbler has little use for assistance and can himself fabricate the whole shoe. If he does require one or two helpers, there will be little division of labor among them. Each will be able and will be required to make a whole shoe, although it may develop that one will be particularly skillful in a special part of the work and devote a large part

of his time to that operation. But if the shoemaker is called upon to manufacture several thousand pairs of shoes a year, it will be economical for him to hire men to perform minor operations, for then he can provide them with continuous work. It would further be profitable for him to use machinery to perform some of the operations.

Territorial division of labor. The extent to which division of labor will be carried depends then on the extent of the market, on the volume of goods which consumers are willing to purchase. The demand for a particular product may increase with the gradual growth of the population or as the result of a change in popular taste. But in the past the most striking cause of increased demand for the product of a particular manufacturer or locality has been a widening of the market through improvements in the means of transportation; in other words, through the addition of a demand which was previously satisfied by other producers or not satisfied at all.

Certain regions are peculiarly fitted by natural endowment for the pursuance of certain occupations. The soil and climate of the American southern states make them particularly suitable for the growing of cotton; the coal and iron of Pennsylvania make specialization in iron and steel products natural; the fertile plains of the mid-western states offer unusual advantages for growing grain. The specialization in production which is thus dictated is called *territorial* or *regional division of labor*. Under almost any circumstances there will be some degree of specialization in the products of the region, but unless adequate transportation facilities exist between different regions each must be self-reliant to a great degree. It is a familiar fact that a century ago New England was in large measure dependent on her own food supply, in spite of the fact that her soil has never been so productive as that of the West. Yet New England could not specialize in manufacturing until she could market a volume of goods far in excess of the needs of her own people and could satisfy her food re-

quirements by grain from the West. The great farming regions of the West, on the other hand, would never have been able to specialize in agriculture except for the possibility of selling their products in New England and other regions and purchasing from New England and other regions their stock of manufactured goods. It was the development of the transportation system that made possible this interchange of goods and this territorial division of labor, to the great advantage of both regions.

Division of labor a form of coöperation. Production with division of labor is clearly coöperative. Laborers, employers, capitalists, and landowners are working together in the sense that each is in part dependent on the efforts of others. In the long chain of productive operations each producer is dependent on the one below him for his raw materials. The cotton manufacturer relies on the grower of cotton, the cotton middleman, the coal miner, the manufacturer of cotton machinery, and literally a host of others, to provide him with the wherewithal to conduct his operations. The agriculturist grows more wheat than he can use, confident of his ability to sell the surplus and with the proceeds secure the cotton cloth that he needs. Coöperation in production consists in the creation of a surplus and the exchange of that surplus for the products of others. The laborer working for wages just as truly exchanges his surplus as does the manufacturer who hires him, for the manufacturer pays him the present worth of his product in money instead of giving him a portion of the product to sell, as might be done and as is sometimes done with farm labor.

This fact of coöperation is by no means of academic interest only; it is of great practical importance. In the past material progress has followed in the steps of coöperation. As in the past so in the future, the path of material progress lies in the direction of increased coöperation, of greater specialization both within industry and between regions having natural aptitudes in the production of certain goods. Any artificial

restraint therefore which hinders this specialization decreases by so much the effectiveness of labor.

The control of production: Conscious and unconscious control. When we survey production as a whole we see the truly marvellous spectacle of millions of men engaged in thousands of different tasks, turning out quantities of goods varying in kind from the most essential to the most frivolous, transporting these goods where they are most urgently wanted, financing their manufacture, shipment, and sale, and performing personal services of the most varied character. Each producer is apparently quite independent in his choice, yet each attempts to turn out exactly the amount of his particular product which can be sold, and with few exceptions the goods all seem to reach the consumers where and when they are wanted. This coördination of human activity cannot be the consequence of accident, nor is it merely a fortunate coincidence; there must be some control. Now such control may be of two types. The first type we find exemplified today in Soviet Russia, where legal authority restricts the freedom of the individual producer and prescribes in minute detail the character and scope of individual activity. The second type of control is exercised by economic forces operating under a régime of personal liberty. We may call it *unconscious regulation* in contrast to the first, which is consciously exercised.

Mercantilism. This second form of control represents a reaction against the restrictions of an earlier system of conscious control. In this earlier period the economic and political philosophy, to which the name *mercantilism* is given, held that regulation of the activity of the individual was a necessary function of the state, partly in order to protect the subjects of the state against the dishonesty or oppressive power of other nationals or of foreigners, and in part to increase the prosperity and prestige of the nation. So we find, from an early time well into the eighteenth or even the nineteenth century, laws and regulations restricting the freedom of action of the individual. In England there were, for example, laws fixing

the price of bread and the rates of wages, laws providing for the inspection of cloth, laws regulating the conditions of apprenticeship, laws prohibiting the export of bullion and of certain raw materials, laws restricting the movement of labor from one part of the country to another, while in France during the period of most minute regulation the individual could not determine what he should produce, how much he should produce, where he might produce it, or to whom he might sell it.

***Laissez faire:* Reaction from government control.** During the eighteenth century the conviction spread that there had been too much government, too much interference with private business, and there arose a strong reaction against the restrictions of mercantilism. *Laissez faire* — let things alone — became the watch cry of philosophers, economists, and statesmen. *Laissez faire* in the first place emphasized the importance of the individual and of individual welfare. Secondly it assumed that individual welfare and national prosperity are not incompatible. It went further and asserted that individual prosperity is at the very foundation of national prosperity; that the individual, if allowed to follow his own self-interest without external restriction, would exert himself to the utmost in the line of endeavor for which he was best fitted and would therefore be a more productive and hence a more valuable member of society than if his occupation and his conduct in that occupation were prescribed for him by either law or custom. As each man followed his own interest, the clash of conflicting interests would lead to modifications and compromises which would result in approximate justice to all.

Translated into terms of concrete policy, *laissez faire* supported (1) free trade, as opposed to protective tariff, (2) no attempt to fix or regulate prices or to regulate either the quantity or quality of goods produced, (3) freedom on the part of the individual to enter any occupation he might choose, in any place he might think advisable, (4) the determination of wages by unfettered bargaining between em-

ployer and employee, (5) the limitation of the functions of government to (*a*) defense and the maintenance of peace and order, (*b*) the dispensing of justice, (*c*) the establishment and maintenance of certain public works and institutions, such as the paving of streets, the improvement of harbors and means of communication, the care of education, etc., which could not be made commercially profitable as private enterprises.

Economic freedom. The reaction against mercantilism spread rapidly in the eighteenth and ninteenth centuries, and the general acceptance, particularly in England, of the doctrines of *laissez faire* ushered in a period in which the dominant note was the endeavor so far as possible to give free play to the forces of individual initiative. Despite the fact that the scope of government regulation has at many points been broadened beyond the limits of extreme *laissez faire*, the conditions under which production has operated have on the whole been those of economic freedom in contrast to legal authority. Taking a broad view of the present system, we note that it is founded on three major principles: (1) personal liberty, (2) private property, and (3) individual initiative and control of enterprise, with assumption of the risks of industry.

So generally is personal liberty today taken for granted that we are in some danger of failure to appreciate its full significance. In truth personal liberty is a comparatively new thing, only recently gained as the climax of a struggle of centuries. Throughout most of the world's history freedom was a rare privilege enjoyed by the favored few. The masses of the people were slaves or serfs, more or less completely subject to the will of their owners or overlords. Even after slavery was abolished and the status of serfdom gradually liberalized and finally developed into comparative personal freedom, the legal restrictions upon economic activity which we have been studying continued. Real individual freedom is scarcely a century old.

The institution of private property also is a remarkable and fairly modern institution. There was probably very little private property in ancient days, common ownership by the tribe or nation or ownership by the chief or king being the rule. The feudal system was based upon ownership of at least the most important element of the wealth, the land, by the king or other ruler. Even today a certain part of the wealth of every community, more in some countries than in others, is owned in common as the property of government. But in most modern nations the great bulk of all the wealth is owned by individuals. Private property involves not only the right of the owner to enjoy all the benefits of his wealth but also the right to say who shall own it after his death; that is, the right of bequest or inheritance. This attribute of private property greatly increases its significance as one of the conditions which control production in the modern world.

The development of individual initiative and control of enterprise, succeeding the régime of control by legal authority, has already been sufficiently discussed. Today, in spite of reactionary tendencies in countries which have been affected by the recent communistic and Fascist movements, it is generally not the congress or the legislature that says what sorts of goods shall be produced and what quantities of each, where they shall be marketed, and at what prices sold. Such matters are generally left to the personal judgment and free initiative of the men who direct modern industry, the farmers and miners, manufacturers and merchants, bankers and brokers; in short, the entrepreneurs.

The assumption of business risk is a normal corollary of the development of individual control of industry. Risk we may define as the chance of loss or injury. It takes many different forms — the uncertainties of fire, accident, change of price, strikes, bad debts, change of fashion, declaration of war, and so forth without end. Basically risk exists because of lack of human knowledge. To an omniscient person, who could foresee the future with perfect accuracy, there would be no

risk, but to actual human beings risk is always present because no one is omniscient; and it is small or great according to the degree of our knowledge.

There are ways by which a business man may reduce risk, and some risks may be transferred, at a price, to professional risk bearers such as insurance companies or to professional speculators, but many remain on the shoulders of the individual entrepreneur. Against these risks his only defense is the initiative and judgment which he can display in the control of his business, and the only reason for his willingness to assume these risks is the chance thus given him to avoid loss and make a profit.

The motive to productive effort. The goal of all industry is production, in order that human wants may be satisfied. Each producer is, under the institution of personal and economic freedom, at liberty to produce whatever he chooses, and the institution of private property permits him to have and dispose of what he produces. Thus is the desire for the satisfaction of wants liberated to become the motive force behind production, as it could never be without freedom and private property. The slave may cherish wants without result in increased application and diligence. He does as little as he can without incurring the wrath of the overseer, because he knows that there is small connection between his efforts and the satisfaction of his wants. A freeman, on the contrary, has the right to enjoy the fruits of his toil, and he is spurred on to efforts more or less energetic and effective according to the urgency of his wants and his productive capacity.

But, as we have seen, modern society is organized upon a coöperative basis, which means, in the first place, that one does not produce the things that will satisfy his own wants but must first exchange his own product for goods produced by others and secondly that production must submit to an extremely complex organization controlled by certain individuals, the entrepreneurs. For the great mass of the workers,

the connection between effort and satisfaction of wants, though it still exists, becomes less direct and less obvious. Nevertheless the basic motive remains; the wage earner is aware that the amount of his wages depends, in some measure at least, upon the efficiency of his productive efforts, and the entrepreneur thinks of the profits of his business. The modern motive to productive effort is thus frankly self-regarding, and modern industry proceeds along the lines of competition, which means in general that each is acting independently in his own interest without primary regard to the interests of others.

The rule of price. Such being the motive to production, and its direction being in the hands of the entrepreneurs, what is it that guides the judgment of the latter and tells them what goods to produce, what quantities of each, where to send them, and to whom to sell them?

The answer is price. Prices determine profits or losses and so invite or repel the entrepreneur as he casts his eye over all the possibilities in the field of production. Wages, the price of labor, turn laborers from one industry to another. Price directs the flow of savings into new capital equipment for this industry or that. Price determines the amount of iron ore that will be mined, the amount of wheat that will be raised, whether wheat will be sent to one country or another. Price is so omnipresent and so subtle a force that few are fully aware of its significance or realize how completely it controls our economic life.

A simple illustration will make this clear. Ask any farmer what he would do if he knew or firmly believed that the price of wheat were going to double. Having every reason to anticipate greater profits, he would undoubtedly try to grow more wheat per acre and plant some acreage to wheat that had previously been devoted to another crop or lain fallow. In his efforts to produce more wheat he might bid for laborers, offering them higher wages to induce them to leave their present jobs, and he might purchase new machinery, so stim-

ulating the manufacture of farm machinery and directing capital into that line of industry. Conversely a fall in the price of wheat, threatening lower profits or even losses, will generally tend to reduce its production, may lower farm wages, and tends to check the flow of capital into farm machinery. The farmers of this country rejoiced in the high prices of the World War. Profits were large and production expanded enormously to make up for the interruption to the flow of wheat from Russia, Australia, and the combatant countries. The return of peace brought with it an approach to normal conditions. Australian wheat could get to the market, and France could devote capital and labor to raising wheat. The price of wheat dropped, and American farmers found themselves in a desperate plight. The explanation was simple; the world production of wheat was too great to be sold at a profit by all of the farmers. The automatic check was working, painfully no doubt, in the direction of decreased production, the only way to raise the price to a remunerative level.

Price is thus the medium through which the consumer makes his voice heard, telling those whose judgment directs industry how much of each good to produce and whether any particular good shall be produced at all. If someone is willing to pay a sufficient price, anything conceivable will be produced, from bread and shoes to diamond shoe buckles. If for something else no one is willing to pay a price from which the producer can make a profit, then that thing will not be made. There is no production of the old-fashioned spinning wheels or of clothing of a bygone style or of gold horseshoes, because no one will pay the price. High prices of the products in any particular line attract the attention of the entrepreneurs and they direct land, capital, and labor into that branch of production. Falling prices and smaller profits are the automatic check to overproduction. Prices of raw materials and wages govern costs and have the opposite effects, tending to stimulate production when low and putting on the brakes when high.

Competitive price is actually determined by bargaining between sellers and buyers, producers and consumers. Each consumer of wheat is striving to purchase his wheat at the lowest possible price, while each seller strives to sell for the highest possible price. The only restraint imposed is that of self-interest. A seller hesitates to be too grasping lest he lose his trade to another seller; a buyer fears to hold off for a low price too long lest he lose the opportunity to buy at any price. If the existing stocks appear deficient, this bargaining of sellers and buyers tends to a rise in price and economy in present use, with more held back for future needs. The opposite results follow the fall in price which is induced by abundant present stocks. And finally the high or low price encourages or discourages future production to replace the present stocks.

Taking stock of our economic organization. In the foregoing pages of this book our special attention has been directed to the infinite number and variety of man's wants and the extraordinary range and intensity of his activities in the quest of their satisfaction, and we have sought to present a sketch of the economic organization which has been built up for this purpose. The central theme, giving coherence and meaning to the infinitely complex picture, is that modern economic life is coöperative, based upon division of labor, individual freedom, private property, and the control of production and distribution by price.

It must not be presumed however that the productive system works perfectly. No human institution is perfect. The critical student of current affairs has frequent occasion to observe that the productive machine does not run with perfect smoothness. He may note an occasional breakdown in the transportation system, a disturbance in the mining industry which interrupts the normal flow of coal from the mines to the consumers, a strike or lock-out in a factory, or frenzied speculation in land which is whipping prices up to a level that cannot be maintained. It is evident that to many a per-

son the alleged freedom to choose an occupation, to produce what he wishes, and to have and enjoy the product of his industry seems more imaginary than real. It is certain that not all the men whose judgment guides the industrial machine are the ones best fitted for their posts. Capital does not instantly and without exception distribute itself among the several industries in perfect adjustment to their respective needs. There is obvious maladjustment of the labor force, and no one can claim that every parcel of land is always put to its most productive use. Competition appears not always to apportion rewards according to productive contribution but sometimes to favor the merely shrewd or the actually dishonest. Private property, in company with free competitive industry and reënforced by inheritance, occasions, or at any rate permits, great inequality in the distribution of the products of industry and of the capital instruments from which further products are to flow.

At frequent intervals have come business crises, which are marked, on the one hand, by widespread unemployment and acute want and, on the other hand, by the inability of producers to sell their products except at heavy loss. Drastic curtailment of production, rapidly falling price level, bankruptcies, and business distress are other characteristics of such periods.

It must also be evident to anyone who looks beneath the surface that, when government authority was displaced by free enterprise under the control of price as the directing force back of production, the control of production was actually given over to the consumers, or more specifically to those who have the wherewithal to pay for the things they want. As has been observed, nothing will be produced, no matter how beneficial it might presumably be to mankind, unless people are able and willing to pay a price equal to its cost of production, and any conceivable thing will be produced, no matter how frivolous or even positively harmful, so long as someone is prepared to pay for its cost. Production is thus directed,

not toward the maximum of things most beneficial to mankind on the basis of some ethical or moral standard, but toward the things which are wanted by those who can pay for them. It is this which gives special significance to the inequality in the distribution of wealth which appears to be an inevitable concomitant of the present economic system. Those who have abundance do not merely receive a greater share of the product of industry; they have a disproportionate voice in saying what things shall be produced. The machinery of production is directed to the turning out of steam yachts and race horses, fine silks and costly wines, even though there be those who are in need of the food and clothing which might otherwise be produced in greater abundance; land is set aside for private estates and hunting preserves which might otherwise be devoted to the production of a more abundant flow of food and other necessities. Moreover, the decisions of a comparatively small group as to whether to spend or invest their incomes may have important reactions upon the whole economic structure.

Why our economic organization thus falls short of perfect functioning is a question whose answer cannot be attempted at this point in our study. There are however two causes so basic and so generally evident as to warrant mention. In the first place, unregulated competition may at certain points degenerate into monopoly, and it is obvious that unregulated price as the regulator of production can function in the public interest only when it is competitive price.

In the second place, there is in the economic world an evident lag between cause and effect. To illustrate, a decline in the price of wheat, caused by a decrease in the demand for wheat, will eventually (other things being equal) cause a decline in the area of land devoted to raising wheat and in the quantity of wheat placed on the market. It may however have surprisingly little immediate effect because of the commitments which farmers have already made, and over a short period of years its effect may even be an increase in the volume

of wheat to be marketed. An individual farmer finds that to pay taxes, interest, insurance, and other costs he must sell more bushels at the lower price. The farmer who has just been making a living at the old price should, to avoid farming at a loss, turn now to a profitable occupation, but transfer of occupation is not effected easily or quickly. He may think it better for a time to raise enough wheat to maintain life, try to stave off his creditors, and wait for better prices.

To the natural slowness of adjustment we find added such artificial impediments to price adjustment as the maintenance of a fixed price policy by a monopoly, or the stubborn resistance to wage cuts which is characteristic of trade unions. These causes provide opportunity for even more serious maladjustments to make themselves manifest in the economic structure.

Shall the system be scrapped? Now there are those to whom these weaknesses of the present industrial régime loom so large that their patience is exhausted and they cry: away with the whole system! Industry is going to rack and ruin in the hands of irresponsible and incompetent entrepreneurs; let us substitute government ownership. Competition is nothing but legalized selfishness, in which the race is to the unscrupulous and grasping, not to the honest and efficient; let us therefore substitute the real coöperation which socialism offers. Worst of all is the intolerable inequality in the distribution of the good things of life; let us therefore prohibit inheritance, let us have government ownership of the capital instruments of production, or better still let us abolish private property entirely and entrust all wealth, along with the direction of industry, to the socialistic state.

What the present system accomplishes. Now the time to pass judgment upon proposals to alter radically the existing economic system is not at the beginning of our study of economics but at its close, when we shall better understand the details of the present system and the possibilities of proposed substitutes. What our present purpose requires is to attain a

clear notion of the general outline of the existing system and its operation. The essential foundation of the philosophy of economic freedom, with its corollaries of private property, individual control of industry, and competition, is that, by leaving the several individuals free to act in their own interests, the maximum total production will be obtained and to that extent at least the interests of the whole community will be served. As was pointed out in the first chapter, the welfare of mankind depends first upon the total amount of production and second upon its distribution. Any "reform" which aims to improve distribution must see to it that the effectiveness of production is not incidentally impaired, lest the remedy be worse than the disease.

There can be no doubt that the system of economic freedom does offer a powerful motive to production and does spur men on to energetic effort. Equally without doubt is it that economic freedom is an incentive to leadership. Executive ability of a high order is by no means overabundant, and it is indispensable that such leadership be found out and placed at the disposal of industry. He is most successful who not only perfects the internal organization of his own business but who makes the most perfect adjustment of his business with other types of business round about him. The present system does on the whole call forth efficient leaders and make them cooperate. Again there can be no doubt that economic freedom offers a powerful stimulus to saving. The motives for saving are varied, but they are particularly strong in a system of private property. Whatever these motives may be, we find that most of them are conditioned on the responsibility of the individual for his own present and future welfare, which is a corollary of freedom. The present system thus promotes the accumulation of capital needed for production.

Again it is certain that the automatic control of industry through competitive prices does on the whole compel producers to bring forth the goods the people want and to deliver them to the consumers at about what it costs to produce

them. This is an accomplishment not to be lightly cast aside, even though the direction of industry may not perfectly conform to some preconceived standard of ethics or morals. And finally the present system of distribution, in spite of all exceptions, does bear a very definite relation to the productive contributions of the several individuals.

The present conclusion. It will be noted that the statements of the foregoing section leave ample room for all the exceptions to which attention has previously been called and for the perfectly evident conclusion that the present system falls far short of perfection. It is therefore to be presumed that the present system is capable of improvement. In fact the system is already the result of first a movement away from government control and later a material reaction from the extreme of *laissez faire*. That further modifications of the present economic system are to come is a reasonable, indeed an inevitable, inference. That the whole system will throughout the world generally be scrapped and something quite different substituted is most unlikely, just as unlikely as that it will be crystallized and perpetuated in exactly its present form.

We have now reached the point in our study where we can at least appreciate the well-nigh infinite complexity of our economic organization and can recognize that he who would attain to such understanding of its nature as to be able to pass wise judgment upon it has set himself no mean task. At the present stage of our study it would therefore be most unscientific either to jump to the conclusion that the imperfections of the present system warrant its discard in favor of something radically different or to conclude from the undoubted services to mankind of the present system that it is perfect and should be defended against any change. In the chapters now before us we shall seek a more complete and intimate acquaintance with the economic system and shall thereby at least approach nearer to that knowledge and understanding which alone can qualify one to sit in judgment upon it.

CHAPTER VI

Price and the Consumer

The rôle of price. The modern economic system is one of universal, albeit unconscious, coöperation, made possible by the control over production and distribution which is exercised by price. With virtually universal division of labor, economic life centres about the exchange of goods and services, through the medium of money. With few and trifling exceptions, whatever anyone produces must be sold for money, and by the same token whatever anyone acquires must come by purchase with money. Price therefore controls each person's enjoyment of the material things of life. The rich man has many enjoyments because he can pay the prices demanded by those who furnish commodities and services. The poor man has less to enjoy, because he cannot pay the prices.

Most of us are producers as well as consumers. To the one who works for salary or wages, the price which his labor commands is the main factor determining his income and, jointly with the prices of goods in general, determining the total amount of his material enjoyments. To the manufacturer, the merchant, and the business man generally, everything centres about price. The prices of labor (wages), of materials, and of capital (interest) determine his costs; the prices of his product determine his income. Upon these prices depends the result of his enterprise — whether profit or loss.

A pecuniary system. This pecuniary character of modern life, while in a sense self-evident, is not always given its due place in our thinking. In order to arrive at fundamentals, we are frequently reminded that money is not the aim of economic exertion but is only a medium, serving the purpose of passing goods from producer to consumer with a facility not possible under a régime of barter. The ultimate social

purpose of production is not the obtaining of money, but the production of goods to satisfy human wants. And the acquisition of money is not for its own sake, but simply as a means to the possession of the things that satisfy wants.

These are indeed profound truths. But they are not the terms in which the business man thinks. To the shoe manufacturer, his factory is operated, not to the end that the community may be comfortably shod, but that profits may accrue to himself. His payment of wages and other costs are not thought of as the means whereby his laborers and others may obtain satisfaction of their wants; they are considered as payments of money, necessary to the conduct of the business. The disposal of his product is thought of in terms of the money brought in, which he regards as the true purpose of his enterprise. In a certain sense the shoe manufacturer's goal is not making shoes, but "making money." This is the business man's point of view. And the business man is the centre of the modern economic system. The modern system is in truth a pecuniary system — a money economy — centred about the exchange of goods and services for money prices.

The behavior of prices: A bird's-eye view. If now, convinced of the importance of the subject of price, one turns his attention to actual prices and their behavior, the first impression is likely to be one of utmost confusion. An infinite number of articles is on sale, each at its own particular price, from the one cent postage stamp and the five cent lead pencil to the million dollar steam yacht. To make confusion worse confounded, it appears that there is no stability in the realm of price. Prices are constantly changing. For example, the price of wheat in England in the fourteenth century was on the average about twenty-five cents a bushel in United States money; in the year 1527 the price suddenly more than doubled. In the fall of 1929 wheat was in the neighborhood of $1.35. Here are a few random examples from American price history. Pine boards cost twice as much in 1891 as in 1860. The price of salt, on the other hand, declined by 1891 to

about half its price in 1860. The price of milk was the same in 1891 as in 1860. During the next generation prices of most things increased. For example the wholesale price of salt rose from seventy-nine cents a barrel in 1891 to $2.70 in 1921, three and one-half times what it was in 1891. White pine boards cost nearly five times as much in 1921 as in 1891. Some few things declined in price. Coffee for example cost 16.7 cents per pound wholesale in 1891 and 7.2 cents in 1921, a decline of more than half. Here are the quotations of cotton on the New York market on June 21 of certain selected years from 1905 to 1939, showing a remarkable succession of changes up and down:

PRICE OF COTTON (MIDDLING UPLAND) IN NEW YORK

1905	$0.09	1917	$0.257	1924	$0.2965
1910	0.153	1920	0.3875	1932	0.0535
1915	0.096	1921	0.1105	1935	0.1185
				1939	0.0944

Individual prices sometimes appear to fluctuate without reference to each other. There is nevertheless a tendency of all prices to move up and down as a group. This is what is meant by references to changes in the general price level. Throughout the world's history there has been a strong tendency for prices to increase — in spite of temporary setbacks.

It is estimated that prices today are on the average from five to ten times as high as they were a thousand years ago. In the United States the price level fluctuated moderately during the generation before the Civil War. From 1860 to 1865, prices as a whole more than doubled. Then there followed a steady decline till about 1897. During the next generation there was an almost uninterrupted rise, reaching its climax in 1920, when prices on the average were more than three times as high as in 1897. This climax was followed by a drop in 1921; but from 1922 to 1929 prices remained on a fairly stable level, about twice as high as in 1897 and about 50 per cent higher than in 1913. The stock market crash of 1929 was

the noisy announcement of one of the most abrupt downward movements of prices in the world's history. By 1932 wholesale commodity prices reached a level below that of 1913, and they continued below the pre-war level until September, 1933. At the present writing commodity prices are about 10 per cent higher than in 1913 and about 25 per cent lower than in 1926.

Prices fixed by statute or custom. There was a time when prices were determined largely by custom and law. People in the Middle Ages talked of "just prices," "fair prices," etc. When a seller sought to exact a price higher than the customary price he at once laid himself open to suspicion. So also of the buyer who sought to compel a price lower than the customary one. In general a price was just if it was as it always had been; the just price was the customary price. Custom was powerful in those days, and the mere force of custom alone had great influence in preventing serious fluctuations in prices. To the aid of custom came law, civil and religious, which undertook, more or less completely, to prescribe what were the fair and just customary prices and to forbid departure therefrom. But neither law nor custom was able permanently to prevent the price changes which followed in the wake of changed economic conditions. There was some attempt to adapt the law to these changes, and the legal prices were altered from time to time, but the law was unable either to resist the changes or to adapt itself to them. Evidently there were forces at work more powerful than custom and statute law.

The legal determination of price gradually declined and finally ceased to exert an important influence, though the policy has never completely disappeared. In the United States today legislative authority regulates the rates charged by the railroads for transporting both passengers and freight. Trolley fares, telephone rates, telegraph rates, etc., are subject to legal control. Rates charged by gas companies, electric companies, water companies, and other local public utilities

are generally limited by official authority. But the mention of these cases is sufficient to remind us that they are the exceptions and not the rule and that prices in general are evidently not determined by legislation.

Custom likewise still plays a considerable part in the control of price. The fees charged by physicians, lawyers, and other professional men, barbers, bootblacks, etc., are quite generally restricted by custom. But it is evident that, in spite of custom, the prices of all these services do change and that there is at no time universal conformity to the customary schedule. What custom does in these cases is not to determine prices, but rather to act as a steadying force, keeping most prices in line with the customary standard and preventing many great and sudden changes. And after all these cases again are the exceptions, not the rule.

Are sellers or buyers free to fix prices? People sometimes talk as though they thought the retail sellers of goods had complete power over prices. Does not the merchant, we are asked, fix his prices as he sees fit? What does the customer on the other side of the counter have to say about it? He can take the goods or leave them; that is all. But obviously things are not so simple as this for the retail merchant. In the first place, he was a buyer before he became a seller, and the prices he charges are obviously related to the prices he had to pay for the goods. Finally the buyer does have an influence on prices, exercised through his power to refuse to buy.

Is price a mystery? How is it then that prices come to be what they are? We have seen the extraordinary variety of prices. Some at least appear to be entirely freakish. A diamond may sell for a thousand dollars, a loaf of bread for ten cents. How is it that the mere luxury is worth ten thousand times as much as a prime necessity of life? A dress suit, worn only on rare occasions, may cost a hundred dollars; a suit for everyday wear, only forty-five. Why does the price of the less necessary thing exceed that of the more necessary? We have seen that prices are constantly changing, some rising while

others are falling, and the average of all prices moving steadily downward for years and then as steadily rising. Is all this complexity a mystery incapable of explanation? Is price merely a matter of chance? Or is it the natural result of causes which may be discovered? Certainly price is of enough importance in human affairs to make the search for its explanation worth while if there is promise that the inquiry will be rewarded by success. The investigation of the forces that determine price is the task next before us.

Consumers' choices. From the broad social point of view, goods are produced in order that human wants may be satisfied. The consumer finds himself confronted with an apparently infinite variety and quantity of goods more or less capable of satisfying his personal wants. He is invited to make his choice, to take whatever things he wishes and as much of each as he may desire, subject only to the condition that he give money in exchange in accordance with a stated scale of prices. This is indeed an onerous condition, and to the ordinary person it immediately imposes a stringent limitation upon his purchases. He abandons at once any idea he might have entertained of possessing the magnificent yacht; he turns away from the tailor shop where a suit would be made to his order for seventy-five dollars. He finally makes his purchases — from a limited number of goods and in quantities of each which are strictly limited. His purchases are the result of a series of decisions, in the making of which he has had to weigh the relative merits of many competing wants, involving the exercise of more or less careful thought and judgment. It is the sum total of these individual judgments and choices of consumers that makes up the demand for goods.

Although it is true that the total demand of consumers is one of the chief price-determining forces, the fact is that the purchases of the ordinary individual are generally only a minute fraction of the total quantity of any good offered for sale. To the individual it appears that prices of the goods in which he is interested are independent quantities, over which

he has no control, and we are justified in making the same assumption for the purposes of our present inquiry.

Everyone is vitally interested, as a consumer, in the problem of so disposing his limited stock of money among the unlimited number and quantity of goods offered to his choice that his total enjoyment may be as great as possible. And since it is these choices of consumers which determine the kinds and the quantities of the things that may be sold, every producer is vitally interested in the mental processes by which consumers arrive at their choices. The investigation of this subject is the first step in the search for the laws of price.

Marginal utility: The subjective side of utility. Let us consider a family living in a village in which there is only one grocery. Let us suppose that on a certain day the housewife finds the grocer no longer offering eggs at the customary price of thirty cents a dozen but on the contrary disposed to charge each separate customer the very highest price he can get. What price would the housewife pay rather than go without eggs? In the first place, she would certainly take count of her present stock. If there were at home five or six dozen eggs, she would doubtless offer to pay little if any more than the customary price, trusting to supply herself later under more favorable conditions. If there were only one or two dozen on hand, she might be willing to pay a substantially higher price rather than go home empty handed. If the egg box were quite empty, she would endure a still higher price, this time measuring the sacrifice the family would suffer from going without eggs. On the other hand, we may assume that the stock of eggs at home is so great that our housewife is not interested in buying any, even at the customary price. In such case it would take a bargain price to tempt her to buy.

We can change our supposition by supposing that, with no possibility of purchasing eggs, the housewife were asked to sell a dozen eggs. Clearly the sacrifice suffered through giving up a dozen eggs would depend upon the number of eggs on hand. The store room might be so well supplied that the

housewife would part with a dozen for thirty cents. On the other hand, were she down to her last dozen, she might think a dollar or two was not too much to compensate for doing without eggs.

From this rather crude example we readily derive the principle that the utility of an additional dozen eggs depends upon the number of dozens possessed, varying inversely with the number possessed, and that the same principle holds of the utility of a dozen subtracted from the stock on hand. Generalizing still further, we find ourselves introduced to a human characteristic which relates itself to all goods (wealth and services). Heretofore we have studied utility objectively, as that quality of wealth and of free persons which enables them to satisfy human wants. Here we encounter utility subjectively, from the viewpoint of the person whose want may be satisfied. *Utility thus expresses the satisfaction which a person expects to realize from a commodity or service.*

When a person apportions a given quantity of any good to different uses, the wants to be satisfied by the several units are not generally of equal importance. Suppose there is a water shortage and a family finds itself in possession of ten buckets of water for the day's use. One bucket might be set aside for drinking, one for cooking, one or two for washing the dishes, two or three for bathing, and so on down the list of uses of water in a descending scale of importance. If the number of buckets on hand were less, less water would be devoted to each use. But it would be the less important uses that would suffer the greater curtailment; some one or more of the less important uses might have to be given up entirely; baths might be omitted. If the quantity were greater, other uses still less important might appear. In this example the wants to be satisfied by the several units are different in kind as well as in importance, but the principle holds equally when the wants are all of the same kind, as in the case of eggs.

Strictly speaking utility is not capable of measurement, as we measure the length of a road in miles or the area of a farm

in acres. Utility is subjective. The utility of any given thing is therefore not a fixed quantity; it is different for different persons and for any one person at different times and under different circumstances.

Relative marginal utility. However it is possible to observe the relative utilities of two objects to a given person by noting the terms upon which he would make an exchange. If a boy is willing to give five sticks of gum for ten marbles, we may infer that the utility to him of a stick of gum is twice that of a marble. If later, having now more marbles and less gum, he will exchange only on the basis of two sticks of gum for ten marbles, we infer that the utility of a stick of gum is now five times that of a marble.

In a money economy such as ours we are able to observe the way in which people choose between the purchase of any particular good and other possible ways of spending their money. Whenever one purchases anything he gives a certain quantity of money in exchange. To the ordinary individual money is not something desired for its own sake. It merely represents a certain quantity of purchasing power, which may be used to buy any of the things that normally enter into his consumption. Thus the purchase of a tennis racket for ten dollars is a transaction in which there is exchanged for the racket purchasing power which is valuable because of all the other things that could have been bought for ten dollars. So instead of considering the relative utility of tennis rackets and each of the innumerable other objects which might compete for this purchaser's attention, we may consider relative utility in terms of the amount of money that the individual would be willing to give up in exchange for a unit of the article in question.

Let us now refer back to the example of our housewife considering the purchase of eggs. Having none at all in the larder, she would be willing, if necessary, to pay a fairly high price for a dozen eggs. Let us assume that under these circumstances she would pay a dollar (if she had to). This

means that a dozen eggs would just compensate her for the sacrifice of a dollar; that is, for one dollar's worth of all the other things that she might purchase. If she already had a dozen eggs on hand, another would not be worth the sacrifice of a dollar's worth of other goods, but some smaller amount, seventy-five cents' worth, let us say. Using figures thus which, though arbitrary as to absolute amounts, are in proper relation to one another, we may set up a schedule showing the respective amounts of money which this person would be willing to sacrifice (if she had to) in order to change her stock of eggs by one unit (the dozen) under various numerical circumstances:

Quantity possessed	*Number of cents which would be sacrificed in order to gain one unit*	*Number of cents which would just compensate for the loss of one unit*
0	100	
1	75	100
2	55	75
3	41	55
4	30	41
5	22	30
6	16	22
7	11	16
8	7	11
9	4	7
10	2	4
11	1	2
12	0	1
13		0

This schedule illustrates the possibility that a person may have so much of a commodity that an added unit (a thirteenth dozen eggs) will afford him no additional satisfaction at all. Beyond a certain point additional units of the commodity may indeed lessen his total satisfaction by burdening or inconveniencing him. Thus, extending the schedule, we might show negative changes in utility.

Definition of marginal utility. Either the second or the third column in the preceding table may be logically headed "ratio of the marginal utility of eggs to that of money" in

accordance with the following general definition: *The ratio of the marginal utility of any good to that of another good for any person is equal to the number of units of the second good that would just compensate him for a change of one unit in the quantity of the first good.* For most purposes it is not convenient to speak of the ratio of the marginal utilities of two different commodities, and realistically in the modern world individual consumers do not think in terms of barter, because the typical transaction which any consumer carries through is one in which he exchanges a certain sum of money for a good or a service. It is usually convenient therefore to measure the tastes of a consumer in terms of *the ratio of the marginal utility of some particular good or service to that of money.* This ratio we shall define as the *relative marginal utility of the good* or, for the sake of brevity, as its *marginal utility.* Hereafter the term marginal utility may be understood to mean "relative marginal utility."

In the application of this concept however a certain convenient numerical simplification is introduced to avoid confusion. A "change of one unit" may occur either through addition or through subtraction. Reference is more frequently made to the "addition of a unit"; but the "loss of a unit," or the "subtraction of a unit," is an equally important possibility. In order not to have two ratios of marginal utilities for each and every quantity, the marginal utility is generally understood to be the marginal utility of a quantity *which includes the unit added.* Thus relative marginal utility through addition and through subtraction is generally considered the same for any particular quantity. Revising the preceding table in accordance with this convention, we have a marginal utility schedule as shown on the following page.

Graphical representation: A digression. Study of the facts presented in a schedule is generally made easier if the schedule is translated into a curve or graph. Such graphical aids to the presentation and interpretation of statistical material have proved very useful both in the study of economics and in the conduct of practical business affairs. For this reason and be-

Quantity	*Marginal utility*
1	100
2	75
3	55
4	41
5	30
6	22
7	16
8	11
9	7
10	4
11	2
12	1
13	0

cause we shall have frequent occasion to use the graphical method of presentation in this book, we may well take this opportunity to inquire into the essential features of the most commonly used type of graph.

Referring to the utility schedule above, we note that it presents two separate magnitudes. The problem is to construct a graph which will show the relations between these magnitudes. We first provide a chart, prepared as follows (see Figure 1, on page 104). Draw a horizontal line parallel to the bottom of the page. This line is technically known as "the axis of abscissas," but for present purposes it may be more simply designated as the *horizontal axis.* Toward the left-hand edge of the page draw a vertical line, intersecting the horizontal axis at right angles. This vertical line is technically known as "the axis of ordinates," but we will call it the *vertical axis.* The point where the two axes intersect, in the lower left-hand corner of the chart, is called the *origin.* This point, marked *O*, represents zero on two scales, a horizontal scale and a vertical scale. It is customary to mark the right-hand extremity of the horizontal axis x; and the upper extremity of the vertical axis y. The horizontal axis may thus be referred to as the x axis, or Ox; and the vertical axis, as the y axis, or Oy.

On the horizontal axis mark off a scale to represent quantities; *i.e.*, in this particular case numbers of dozens of eggs.

Begin this scale with zero at the origin, and mark it off evenly along the axis with appropriate numbers. The axis is thus divided into a number of equal parts called scale divisions. The greatest of the numbers on the scale should be at least as large as the largest quantity to be represented in the chart.

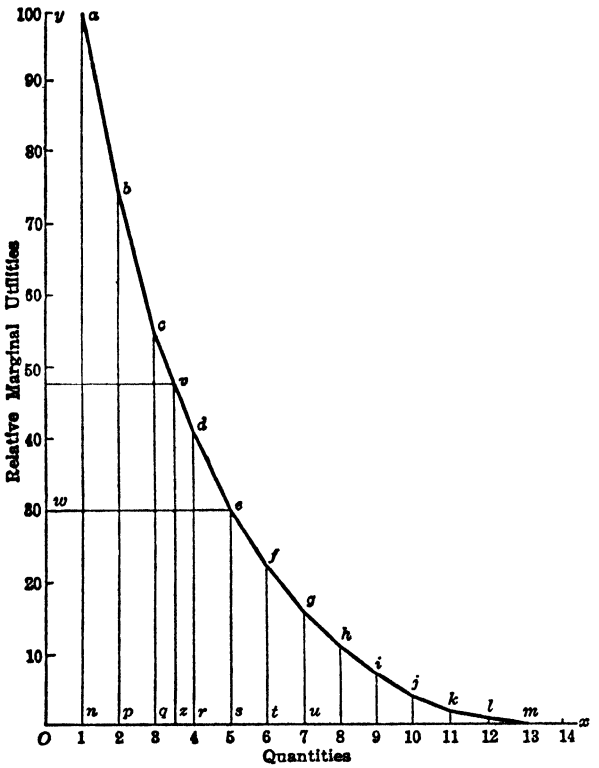


FIG. 1

In this example the largest quantity is 13. Spread the scale so that the largest given quantity appears well over to the right; this is in order that the magnitudes we deal with may not be crowded too closely into the left-hand corner of the chart. Skip no intervening numbers between zero and higher magnitudes; let this scale be continuous from left to right.

In almost all English and American economic charts,

quantities of goods are measured along the horizontal axis; and other magnitudes (utilities, prices, costs of production, etc.), along the vertical axis. Although this convention is purely arbitrary, its use serves to avoid confusion. To make the chart absolutely clear, however, it is advisable to print the word "Quantities" along the horizontal axis, and another appropriate word or phrase along the vertical axis.

In this case we print the words "Relative marginal utilities" along the vertical axis, and along this axis we mark off a scale. Begin this scale, as well as the other one, with zero at the origin. Mark it off evenly upward with appropriate numbers. Let the greatest of these numbers be at least as large as the largest measure of utility to be represented in the chart. In this case this largest measure is 100. Spread the scale so that the largest given measure of utility appears roughly as far above the origin as the largest given quantity does to the right of the origin. The scale divisions on either of the two axes need not, and usually will not, be equal to those of the other. In the vertical scale, as in the horizontal one, skip no intervening numbers between zero and higher magnitudes. If these initial directions are followed, the entire figure when completed will comfortably cover the designated space; the scales can be easily and correctly read; and the inclined lines in the figure will not be too steep or too nearly horizontal for quick, accurate interpretation.

Now let us proceed to locate on the chart points representing the respective relations between the possible number of dozens of eggs, from one to thirteen, and corresponding relative marginal utilities. The first line of the utility schedule indicates that the marginal utility of one dozen is 100. To represent this supposition graphically, locate 1 on the horizontal scale (as shown in Figure 1), and proceed in a straight line vertically upward until a point is reached exactly on a level with 100 on the vertical scale; or from 100 on the vertical scale proceed in a straight horizontal line to the right until a point is reached exactly above 1 on the horizontal

scale. In either case we arrive at the point marked *a* in Figure 1. This point represents the supposed fact that the marginal utility of one dozen eggs is 100.

In similar fashion twelve other points (*b*, *c*, . . . *m*) may be located, representing twelve other suppositions regarding the marginal utilities of various numbers of dozens of eggs as shown in the utility schedule. Thirteen points in all are located. They are connected by a series of lines. Thus a "curve" *am* is obtained, which is known as a *marginal utility curve.*

Having constructed the curve, we pause to consider its interpretation. It is a marginal utility schedule, a person's mental attitude, converted into graphical form. Every significant point on the curve bears a definite relationship both to the horizontal and to the vertical axis. The distance of such a point from the vertical axis, as measured by the horizontal scale, indicates a certain number of dozens of eggs. The height of this point above the horizontal axis, as measured by the vertical scale, indicates the amount of money our imaginary housewife would be willing to sacrifice to obtain a dozen eggs which increases her total stock to the number indicated.

Certain additional lines in Figure 1 would not ordinarily be necessary, but they have been drawn to assist us in detailed study of the chart. Consider, for example, the line *we*, drawn from *e* perpendicular to *Oy*. Its length, the same as *Os* on the horizontal axis, indicates that the point *e* refers to five dozen. In conformity with this information the height of the point *e*, equal to *se* or *Ow*, indicates that if a fifth dozen is acquired it will just compensate this person for the loss of 30 cents. The height of the point *e* likewise indicates that the loss of any one of a total of five dozen could be compensated by the sum of 30 cents. From points *a*, *b*, *c*, *d*, *e*, . . . *m*, other similar definite readings may be made.

When it is possible to use very small units, the points on a marginal utility curve may become so numerous as to give us practically a smooth curve. Thus when we represent the

marginal utility of sugar, measured in pounds, the diagram is somewhat as in Figure 1. Measured in quarter-pounds, we should have four times as many significant points. Measured in grains, the points would be so close to each other as to form practically a continuous curve. Actually we seldom are able to use units small enough for this result, but when dealing with units which are capable of subdivision, it is possible to carry the graph to a further stage of usefulness. Suppose it is desired to know the marginal utility of $3\frac{1}{2}$ dozen eggs. Obviously this will be somewhere between 55, corresponding to 3 dozen, and 41, corresponding to 4 dozen. To obtain the answer graphically, it is first necessary to change the curve from a series of straight lines connecting the ends of the several ordinates to a curved line passing through them. Assuming this "smoothing of the curve" to have been done, the marginal utility corresponding to $3\frac{1}{2}$ dozen is now readily shown by the point v, indicating that when $3\frac{1}{2}$ dozen (Oz) are considered the marginal utility, measured by zv, is 48. This process is called interpolation.

The simplicity of the illustration used in this elementary discussion of the graphical method does not adequately emphasize the services of the curve as an aid in analysis of statistical data. The student of intricate economic phenomena in the business world finds that so vast and complex is the mass of data with which he must deal that other methods of analysis are inadequate. Of course the curve adds nothing to the evidence of the data from which it is constructed; it does however reduce complex data to a simplified form, so that inherent truths may be brought to light, thus abbreviating the search for the laws which govern the relations of different factors.

Meaning of marginal utility. It must be emphasized that our marginal utility schedules and curves measure relative marginal utilities. Because utility cannot be measured and because all that we know either about ourselves or about others is the way in which we will choose between alternatives

that are presented to us, the concept of an absolute marginal utility would have no meaning.

Since the marginal utility of a good, as we define it, is relative to the marginal utility of money, its marginal utility will change with changes in the amount of money possessed even when the amount of the good itself remains the same. If for instance you were suddenly given a thousand dollars to spend within the next ten days, the relative marginal utility of practically everything that enters into your budget would be sharply increased. The income you would have available for expenditure during the next ten days would suddenly have become so abundant that the sacrifice, let us say, of five dollars for a theatre ticket would seem much less important than before. The additional satisfaction you would gain from witnessing the show would be much greater than before measured in terms of money; that is, in terms of the other things you might buy with any given number of dollars. Similarly the relative marginal utility of any good to a wealthy man will usually be much greater than to a poor man unless the former already possesses far more of the good in question than the latter.

We must also remember that a marginal utility curve does not show a historical sequence of uses or utilities; it depicts a series of simultaneous possibilities — not in the sense that all of them can be realized at the same time, but in the sense that any of them may be realized under the same general circumstances. Thus at a given moment our imaginary housewife may have thirteen dozen eggs, or she may have none at all, or she may have any intermediate number. The relative marginal utility curve shows the additional satisfaction in terms of money she derives from any dozen eggs which increases her total stock to any indicated number within a given range of possibilities.

It is somewhat misleading to say, as is sometimes done, that "marginal utility is the utility of the last unit." The units are all alike. Once in the possession of the user any

unit will render him equal service, afford him equal satisfaction, no matter in what order it may have been acquired. If a difference in utility is occasioned by the addition or subtraction of a particular unit, this change cannot be attributed to any peculiar characteristic of that unit. Any other unit would produce the same result.

The law of marginal utility. Once we have a clear idea of the meaning of marginal utility, the economic law of the relation between the quantity of an economic good and its marginal utility becomes obvious. This principle is known as the *law of marginal utility* and may be stated as follows: *At any given time the relative marginal utility of any good to any person varies inversely with the quantity under consideration.* A glance at Figure 1 will make the meaning of this perfectly clear.

This law is in part a result of the well-known psychological law that when an identical stimulus is repeated the response gradually becomes weaker. The appetite for food is the simplest illustration. Having eaten one apple, the second gives us less satisfaction, the third still less, and very soon we reach the point of satiety, when no further satisfaction, perhaps even repugnance, results. The same loss of appetite would have appeared more gradually if we had considered the separate bites.

It should be realized, however, that the sort of marginal utility which commonly determines economic action does not arise from immediate use or consumption, but from anticipation. When the housewife buys sugar, she does not generally use it and pay for it pound by pound. Ordinarily she buys enough to last at least several days. For purposes of immediate consumption the marginal utility of a relatively small amount of sugar would be zero; but for the anticipated uses of the next week or ten days the same amount of sugar may have a high marginal utility. Therefore she buys not only this amount but also several pounds more. To the psychological principle of satiety must therefore be added the principle that man takes account of the future and makes provision

for his needs beyond the present moment. But just as there is satiety in immediate consumption, there is a like principle in anticipation of future satisfactions. The greater the amount bought for future use, the less is the added prospect of satisfaction through an added unit of the commodity.

Some apparent price anomalies resolved. In the concept of relative marginal utility lies the key to certain price relations which sometimes appear to be puzzles. How is it that some of the most useful things in the world are so cheap, while other things that could be given up without any real suffering are so costly? To some the fact that a diamond may be worth a thousand dollars while a loaf of bread is worth ten cents seems strange, inexplicable, or even wrong. The answer is that the value of anything depends, not upon somebody's judgment of the importance or righteousness of the want satisfied, but on demand and supply, and the demand side is controlled by considerations of marginal utility. Marginal utility, we have just seen, varies inversely with the quantity. Diamonds are scarce; their marginal utility is therefore high. Bread is very abundant; its marginal utility is correspondingly low. People are willing therefore to pay high prices for diamonds, but very little for a loaf of bread.

Abundance and scarcity. Some of the most useful things in the world are so abundant that they have no marginal utility at all and therefore no value. Among the prime necessities which man must have in order to sustain life is air; nothing could be more useful. Yet no one will pay for air, simply because it is so abundant that everyone can have all he wants for nothing. Its marginal utility is zero. To have value, a thing must have not only utility, but a marginal utility greater than zero. The only things that can have positive marginal utility are those which both have utility and are so limited in quantity that there is not enough to satisfy everybody's wants. It is for this reason that some of the most important things in the world, as sunlight, air, sometimes water, are not considered "economic goods." They have no value, are not

bought and sold; it is this, rather than any permanent physical characteristic, that rules them out. Water, free in abundance for all in many country districts, is limited in the city; it there has a marginal utility, has value, and is an economic good.

EXERCISES

1. The marginal utility schedule below reflects the subjective estimate of an individual of the importance to him of successive additional pairs of shoes:

Quantity	*Marginal utility*
1	$100
2	20
3	12
4	6
5	2
6	1

(*a*) Following the instructions contained in this chapter, construct a diagram to represent this marginal utility schedule.

(*b*) If the particular grade of shoes sells for $10 a pair, how many pairs will this individual find it worth while to possess? Explain.

2. A college student has $3 to spend for a ticket to a football game or a dance or a book or a new hat. Explain the basis on which he makes his decision between these alternative forms of expenditure.

3. Assuming that people are frequently influenced by shrewd salesmen, misleading advertising, gossip of friends, etc., to buy goods which give them little real satisfaction, is this inconsistent with the principle of marginal utility?

CHAPTER VII

Demand

Individual demand: Marginal utility and demand. It is essential to have clear and precise concepts of demand and supply. For demand the foundation was laid in the chapter immediately preceding, where attention was given to the consumer and his attitude toward the purchase of goods for money. Our starting point is marginal utility. As we have seen, there is no accurate practical measure of absolute marginal utilities. However each person, after giving more or less careful consideration to the amount of his resources and to the prices of the various things which he might purchase, does have some notion of the amount of money he would pay for a unit of any good under a given set of circumstances.

This has already been illustrated by the case of an imaginary purchaser of eggs. The curve *am*, in Figure 2, represents the conditions assumed in that example. It shows the money estimation of the loss which a certain individual would suffer by parting with one dozen eggs, assuming him to have any given number on hand, or the price he would be willing to pay for one additional dozen. Every vertical distance (such as *an*) thus measures the price one would be willing to pay for the additional unit which would bring his stock up to the number indicated on the base line, provided of course he were held up each time for his highest price.

Now this of course is not the way eggs, or other things, are ordinarily bought. There is an established price on the market. No matter how badly the buyer may want the first few units, no matter how great their utility to him, no matter how much he might be willing to give for them if he had to, he does not have to pay more than the market price; and he may ordinarily buy all he desires at that price. The question

he has to determine is, not what price he will pay, but how many he will take at the established price. Suppose the housewife finds that the price of eggs is thirty cents a dozen. How many dozen will she buy? Referring to Figure 2, we conclude that she will certainly buy one; rather than go without one

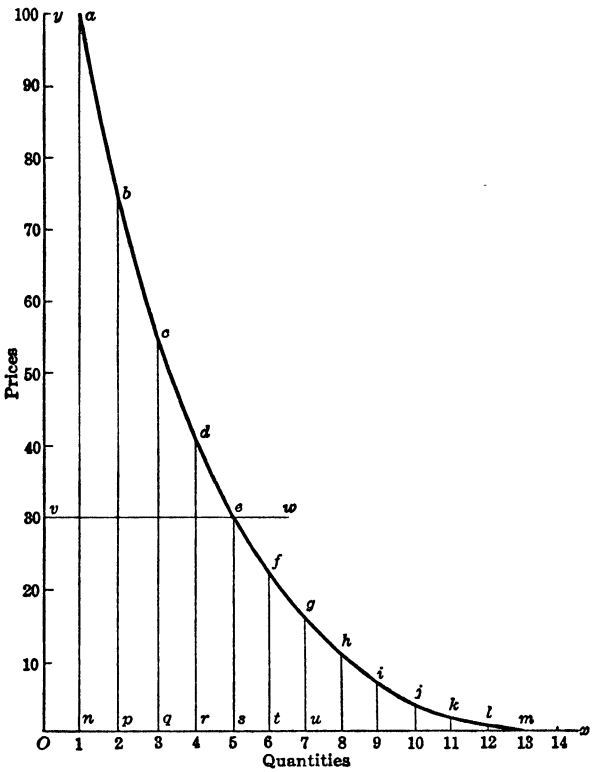


FIG. 2

dozen she would be willing to pay a dollar. Similarly, she would buy a second, a third, and a fourth, since in each case the sum she would give if she had to is greater than the price. For a fifth dozen she would give thirty cents. She will just get "her money's worth" by buying this one. A sixth dozen would be worth only twenty-two cents to her. She would gladly have this one and several more, but not at thirty cents.

We conclude that this person would buy five dozen eggs.

The curve in Figure 2, which originally illustrated the relative marginal utility of a dozen eggs under different circumstances, thus serves equally to illustrate the quantities that this individual would buy at various prices. From the point of the *y* axis which measures thirty cents, we draw a horizontal line, *vw*. This line intersects the curve at the point *e*, indicating that at a price of thirty cents five dozen will be bought. Let us suppose our housewife had found that they were charging seventy cents. She would consider this an exorbitant price and, while she would have to get enough to supply her immediate needs, she would buy fewer than if the price were more reasonable, in this case two dozen. If on the other hand, she had found that eggs were being sold at fifteen cents a dozen — a bargain price — she would have bought a few more; namely, seven dozen.

Of course only one of these price possibilities can be realized at a given time. Suppose the price is actually announced to be thirty cents. This individual will buy five units. This act should not blind us to the fact that consciously or unconsciously there resides in her at the same time both the willingness and the ability to buy eight units at eleven cents, six units at twenty-two cents, and other amounts at other prices as shown on the curve. Such ability and willingness constitute an individual's demand for a commodity. The concept may be defined thus: *An individual's demand for any good (wealth or service) is a schedule of the respective quantities of that good which he is ready to buy at all possible prices.* A curve such as that in Figure 2 is the graphical representation of this schedule and is known as the individual demand curve. It is identical with the relative marginal utility curve. Whichever name it takes, the curve is a picture of possible action in response to possible prices. It is controlled, not by the marginal desirability of the specified commodity alone, but by a balance of the desirability of that commodity against the desirability of other commodities and services.

Significance of this analysis. Hardly anyone would actually take the trouble to make minute calculations such as have been assumed in the foregoing analysis. Nevertheless most of us do have to exercise judgment in the expenditure of money, whether it be only a dollar or ten thousand dollars. Everyone has at least in his subconscious mind some idea of the satisfaction he expects to get from his purchases and can sense to some extent the result of spending a little more on one thing and correspondingly less on others. The given examples are more precise and definite than ordinary human behavior, but such examples are necessary for an insight into the complexities of human problems. A map is far simpler than the country it describes. A globe in the library is smoother than the earth it represents. For the same reason illustrations of economic principles often have to be simplified at the expense of absolute realism. The preceding analysis illustrates (admittedly in ultra-simple and definitive fashion) the manner in which an individual weighs marginal utilities against each other, adjusting his expenditures when he finds marginal utilities out of balance, seeking marginal utilities proportional to the prices he has to pay — all in the effort to make the most of the limited purchasing power at his command. The thought, the feeling, the unseen reflexes which govern this balancing of utilities, may be known only to the individual himself, and even then very imperfectly understood; but out of it all arises visible action, the purchase of economic goods.

How both parties to a trade can gain. People sometimes find it difficult to believe that both parties to a trade can gain. It has been somewhat plausibly argued that, since in any honest trade the values of the two things exchanged must be the same, neither party can have made a gain; one party can gain only by giving less value than he receives; that is, by cheating the other party. It was partly this notion which led the churchmen and lawmakers of the Middle Ages to believe that trade should be strictly regulated and customary prices

enforced in order that neither party to a trade should gain at the expense of the other.

But unless the marginal utilities a buyer has in mind are born in ignorance and false anticipation, a buyer gains by the opportunity to purchase any commodity so long as the amount obtainable for the money spent more than compensates him for the loss of the money. The seller, on the other hand, has received more than enough money to compensate him for the loss of the goods with which he parted. This side of the trade will be more clearly set forth in a later chapter. In the meantime Figure 2 gives us an indication of the buyer's gain. Suppose the price is thirty cents. This individual would buy five units. From the shape of his demand curve, which measures the relative marginal utility to him of various amounts of the good, we can tell that the first unit of the good would compensate him for the loss of one dollar, the second for the loss of seventy-five cents, and so on. We can be certain that each of the five units which he purchases more than compensates him for the thirty cents which he must spend to acquire it and that if his judgment has been sound he is indubitably better off after the transaction than he was before.

Real people. Of course the ordinary person does not use marginal utility curves or schedules or demand curves or schedules in determining what he is actually going to do; they are, for the economist, ways of sketching the various things an individual may do in response to various possible stimuli. The closest approach to rational buying is seen when people budget all their expenditures. A careful family budget is an unconscious illustration of the principle of balancing marginal utilities. But many, perhaps most, people do not go even so far as the family budget. Poor people and even the moderately well-to-do often appear to spend their limited stores of money in the most foolish and irrational manner. A family that is without means to provide warm clothing for the coming winter will cheerfully spend ten dollars on a day's outing at the seashore. The radio, with its monthly installment payments,

may be eating up the money that would seem much better spent for nourishing food. The daily attendance at the "movies" may seem quite superior to any law of marginal utility. Clever advertising and smart salesmanship bank on the careless and irrational way in which people spend their money. All sorts of appeals are used to induce the careless purchaser to buy things whose utility seems absurdly small in comparison with the money expended. To some the appeal of a "bargain" is irresistible, and people come home laden with things for which they have no real need, bought simply because they were cheap; every retail merchant takes advantage of this human characteristic. In any city it is possible to see people buying things in one store when the identical articles are on sale at a lower price only a block or two away. A study of the purchasing habits of the general public is apt to induce a low regard for average human intelligence.

Is all buying irrational? From such facts as these we may be tempted to draw the conclusion that there is no such thing as rational judgment in buying, that the principle of balancing utilities has no existence outside the imagination of the theoretical economist, and that all our talk of demand and supply is pure theory without any useful application to the hard facts of the practical world.

But let us not be too hasty in jumping at such conclusions. For one thing, it will be well to recognize that much of the conduct of other people which we regard as irrational is not really so but is merely at variance with our own judgment. We readily think we could lay out a family budget for our impecunious neighbor which would materially increase the utility of his purchases. But it is his judgment — such as it is — not ours, which decides how he can get the most for his money. The well-fed, well-dressed woman of the "upper class" points with scorn at the "irrational" conduct of the stenographer taking her pathetically inadequate luncheon at the soda fountain in order to save money to buy silk stockings. But this conduct, whatever else may be said of it, is not irrational.

The girl purposely spends more money on clothes, and less on food; this is her way of balancing marginal utilities. Her conduct may be "foolish" according to some other person's judgment. But the economist is investigating how she actually spends her money, not how she "ought" to spend it.

Let us reverse the case. The shopgirl laughs at the wealthy patrons of the Fifth Avenue store who "haven't sense enough" to walk over to Sixth Avenue and buy the same things at lower prices. But it does not follow that the wealthy shopper is irrational. She wants the calm restfulness, the air of luxury and dignity, the deferential personal attention of the exclusive Fifth Avenue shop. These are a part of what she is paying for. She knows she could save money on Sixth Avenue, but the loss of these other advantages would mean a net sacrifice in the utility of her purchases. She also may be "foolish," but that is beside the present point.

Total demand: Individual demand and total demand. Thus far we have considered demand from the standpoint of the individual purchaser, who regards the prices of what he would buy as fixed and not subject to his control. But though the individual's demand for any good has no appreciable effect upon its price, the total demand of all consumers has a profound effect. The total demand is the sum of the demands of all individuals who are contemplating buying the good in question. For the sake of a simple concrete example, let us assume that in a certain village four families only [1] are interested in the purchase of strawberries, and let us assume their respective demands as shown in the schedule on the following page.

Family A is evidently well-to-do; they will take all the strawberries they want, regardless of what the price may be (within the limits of possible prices). To the other families the matter is more serious; their purchases will be governed

[1] The assumption of this small number is of course artificial. It serves to simplify the example without affecting its validity as an illustration of the general principle.

DEMAND FOR STRAWBERRIES, RETAIL MARKET OF X—MAY 1, 1940

Price per quart	*Quantities people are ready to buy*				
	FAMILY A	FAMILY B	FAMILY C	FAMILY D	TOTAL
55 cents	10 qts.	2 qts.	1 qt.	0 qts.	13 qts.
50 cents	10 qts.	3 qts.	1 qt.	0 qts.	14 qts.
45 cents	10 qts.	4 qts.	2 qts.	0 qts.	16 qts.
40 cents	10 qts.	5 qts.	2 qts.	1 qt.	18 qts.
35 cents	10 qts.	6 qts.	3 qts.	1 qt.	20 qts.
30 cents	10 qts.	6 qts.	4 qts.	2 qts.	22 qts.
25 cents	10 qts.	6 qts.	5 qts.	3 qts.	24 qts.

by the price. The schedule in the last column is the total demand for strawberries in this simplified market.

Definition of demand. It is total demand which operates as one of the chief determinants of price. Hereafter the term demand, when not qualified, will be understood to mean the total demand. It may be defined as follows: *The demand for any good is a schedule of the respective quantities of that good which people are ready to buy at all possible prices.*

For the sake of a concrete example, let us expand our illustration of the strawberry market of X by including all of the hundreds of families which may be interested in the purchase of strawberries, giving a demand schedule in general character like the following.

DEMAND FOR STRAWBERRIES, RETAIL MARKET OF X—MAY 1, 1940

Prices	*Quantities* *
55 cents	95 quarts
50 cents	100 quarts
45 cents	115 quarts
40 cents	150 quarts
35 cents	200 quarts
30 cents	300 quarts
25 cents	500 quarts

* Hereafter the word "quantities" in any demand schedule indicates quantities which people are ready to buy at various possible prices.

In Figure 3 (page 120) this schedule is represented in graphical form, the curve *ag* being the corresponding demand curve.

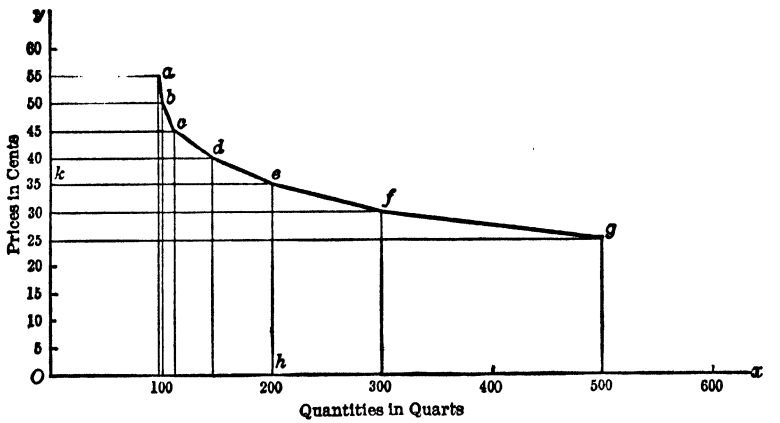


FIG. 3

Time and place of the demand schedule. The demand schedule is a series of simultaneous possibilities, only one of which can be realized at a given time. It relates to a particular date or period — to the day of May 1, 1940, or rather to those hours of that day during which buying can take place. The prices are alternative prices; only one of them can turn out to be the actual price. So of the quantities; only one can prove to be the amount actually bought. Which of the various amounts is actually bought we may not know. If we do, we must not forget that at the same time there is an ability and willingness to buy other amounts at other prices. The demand schedule is neither history nor prophecy. It does not show a series of purchases which have taken place at such and such prices. It does not show how much will be bought at any particular price after the given date or period. Each day will have a demand schedule of its own, which may or may not be the same as this one.

In the strictest sense the demand schedule does not relate to a *period* of time at all, but rather to an *instant* of time. Theoretically conditions of demand may be changing continuously, so that each succeeding instant has its own demand schedule. For most practical purposes however it is suffi-

ciently accurate to consider demand with reference to a considerable period, during which a certain price might hold and a certain quantity be bought. This period, be it remembered, may be quite short, often less than a day or even less than an hour. In the market for certain corporation stocks or certain commodities, such as wheat, cotton, etc., where trading is very active, conditions of demand change rapidly, and there is close approach to the theoretical condition of the constantly changing demand schedule.

Of course such a schedule must relate also to a particular place. Market conditions vary, and a schedule showing the quantities which would be bought at various prices in one market will not hold of any other market, except as a coincidence.

Demand and desire. Demand is not the same as desire. An impecunious student, standing with empty pockets outside a theatre, may have a keen desire to see the show, but we cannot say that he has a demand for a ticket. His desire has no effect upon the demand for tickets, no effect upon the numbers that could be sold at various prices. To have a demand for anything one must not only desire it; he must also be able and willing to buy. The only one who contributes to the demand for a certain make of automobile is the one who is able and willing to buy at least one at a price at which this automobile is likely to be for sale. Though human wants are obviously the origin of the demand for any good, we need to be on our guard against thinking desire and demand are the same thing.

The law of demand. If the reader will glance at Figure 3, he will notice that the demand curve in this chart slopes downward as it goes from left to right. As we move along the curve in this direction, the distances that measure prices grow shorter while the distances that measure quantities grow longer. In graphic form this is an illustration of an important principle known as the *law of demand.* The principle may be stated as follows: *In general, the quantity of any*

good which people are ready to buy varies inversely with the price of that good.[1]

This law has been developed with the aid of numerical schedules and graphs, and these devices will continue to serve us in the further pursuit of the laws of price. The reader hardly needs to be reminded that the exact quantities used in these schedules are a matter of presumption and are not significant except as a means of illustration. The general relation between price and quantity — *i.e.*, that the quantity people are able and willing to buy increases with a decrease of price and *vice versa* — is a truth beyond question. It is only when we undertake to express this in exact numbers that we must deal in presumption. In actual business it is seldom if ever possible thus to give exact values to the demand schedule, although the business man is perfectly certain of the general relation between price and the quantity that people are ready to buy.

Special cases. We have used the phrase "in general" in the above generalization in recognition of certain special cases in which there is a different relation between price and the quantity that people would buy. If the price of diamonds, for example, should fall very low, it is quite probable that sales, instead of being larger or even as large as they are, would actually be smaller. Desire to have diamonds is stimulated to a large extent by the very fact that they are high priced. If they should descend to low prices at which everybody could buy them, probably few people would want them. At times when people have plenty of money, the same perversity is sometimes illustrated in the sale of other luxuries and novelties. These special cases we must recognize when we come upon

[1] As we shall use the terms, "inversely" and "directly" do not imply that the variation must necessarily be proportional. If a increases when b increases, or decreases when b decreases, we say that a varies directly with b, even though the change in a may not be exactly proportional to the change in b; and correspondingly of an inverse variation.

In accordance with strict mathematical terminology it would be more correct to say that the quantity people are able and willing to buy is "a decreasing function" of the price; but this phrase has been found to be awkward and cumbersome in ordinary speech and calculation. The principle as stated above has abundant precedent in good literary usage.

them, but they should not blind us to the existence of the general principle or tendency.

Elasticity of demand. Every household uses a certain amount of salt. It would be most uncomfortable to go without salt or even to cut down appreciably upon the usual amounts consumed, yet there would be little to be gained by increasing the amount used, since everybody already has about all he wants. If the price of salt should be unusually high, people would still buy nearly as much as usual. The amount expended for salt is a very small part of the household budget, and even at a much higher price the cost would be small, and the increase could be made up by a slight saving on

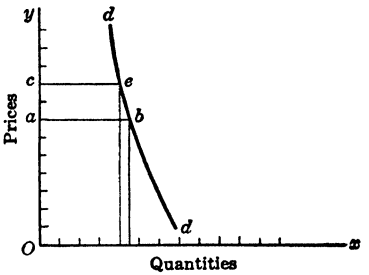


FIG. 4

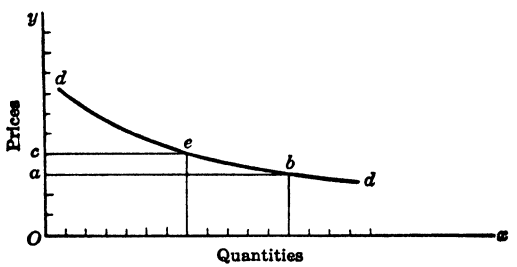


FIG. 5

some other item of expenditure. On the other hand, a very low price of salt would not mean any great increase in the quantity bought, for the simple reason that people were already buying about all they wanted. The law of demand holds good of salt, as of other things; that is, at any given time the quantity that people are able and willing to buy varies inversely with the price. But it varies only slightly. In such a case it is said that demand is very inelastic. The demand curve has a steep slope as shown in Figure 4.

On the other hand, there are things which have an elastic demand. This means that variations in price are accompanied by great variations in the quantities that buyers would take. If the price of a certain high-grade automobile were $500 instead of $3,000, the number of persons able and willing to buy that particular make of automobile would be enormously increased. If the price were $10,000, the number of purchasers would be only a small fraction of those who would buy at $3,000. The demand for any particular kind of car is extremely elastic within a certain price range. The demand curve in such a case has a gentle slope as illustrated in Figure 5. This is because changes in the distances measuring prices (as *Oa*, *Oc*, etc.) are accompanied by relatively great changes in the distances measuring quantities (as *ab*, *ce*, etc.).

Finally it should be noted that the degree of elasticity is not necessarily uniform throughout the whole possible range of prices. It is possible for an article to have an elastic demand within certain price limits while having an inelastic demand at some other level.

The limits of elasticity. The absolute limit of inelastic demand is reached when the quantity people are able and willing to buy is the same irrespective of price. Suppose a state calls for bids to erect a new capitol building. Here there is only one buyer in the market, and there is only one article to be exchanged in any case. The state will take this one capitol building, no more and no less, regardless of the price, up to the limit where the price is so high as to cause the state to give up the building project altogether. The corresponding demand curve will evidently be a straight line parallel to the *y* axis. The same sort of demand appears when a single buyer calls for bids for a fixed quantity of a certain special article for which there is no demand elsewhere. For example a national government might need 500,000 gas masks of a particular style for its army. It calls for bids. The quantity taken will be exactly 500,000, whatever the price may be, and we have again the vertical demand curve.

The absolute limit of elastic demand is reached when buyers are ready to buy any quantity at a given price but would buy nothing at any higher price. In other words, the price people are able and willing to pay is the same irrespective of quantity. As an example we may cite a large oil company which agrees to purchase all the crude petroleum brought to it at a posted price. Demand in this case is represented by a straight horizontal line parallel to the horizontal axis and at a height indicated by the determined price on the vertical axis.

While it is conceivable that these extreme conditions of absolutely elastic or inelastic demand might exist with two or more independent buyers, such cases are seldom found in practice. What makes demand reach either extreme is generally the presence of a single buyer who is alone ready to buy the good in question in the particular market.

EXERCISES

1. Draw up an assumed schedule of the demand for potatoes on a certain day in a small town.
   (*a*) Construct the demand curve corresponding to this demand schedule.
   (*b*) Take any point on the demand curve, and draw perpendiculars from this point to the horizontal and vertical axes. Exactly what does each perpendicular represent?
2. Describe the essential features of a demand schedule. What factors may cause a change in demand from one time to another? Draw a simple diagram to illustrate an increase in demand; a decrease in demand.
3. Draw up a demand schedule for each of two commodities, one of which has an elastic demand, the other an inelastic demand. Test each demand schedule to insure that it conforms to the requirements. Plot the two demand curves, on separate graphs.
4. Explain how each of the following factors tends to affect elasticity of demand:
   (*a*) Necessity or luxury
   (*b*) Presence or absence of substitutes
   (*c*) Habitual consumption
   (*d*) Size of the expenditure for the particular good as compared with one's total expenditure
   (*e*) Advertising warning the buyer to "beware of imitations" or "accept no substitute."

5. In each of the following cases, what condition is implied with respect to elasticity of demand? Explain.

(*a*) On several occasions in recent years the American Telephone and Telegraph Company has reduced its toll rates for evening calls.

(*b*) A small wheat crop may result in a greater income to wheat growers than a large crop.

(*c*) In 1936 passenger rates on American railroads were reduced from 3 cents to 2 cents per mile, at a time when the railroads were in dire need of additional revenue.

(*d*) The Brazilian government has frequently resorted to burning coffee or dumping it in the ocean, in an effort to increase the incomes of coffee producers.

6. (*a*) A city calls for bids to furnish it 50 garbage trucks of a particular type. Draw the appropriate demand curve.

(*b*) The United States Treasury stands ready to purchase all gold brought to it at $35 per ounce. Draw the appropriate demand curve.

7. Does the schedule for Family A on page 119 contradict the principle of diminishing marginal utility? Explain.

CHAPTER VIII

Cost of Production

The different viewpoints of buyers and sellers. The buyers of goods are either consumers, who seek goods to satisfy their wants, or producers,[1] who buy in order sooner or later to sell either to consumers or to other producers. In any case the goods finally reach consumers, and it is this last exchange which is the fundamental one for the study of the demand side of the forces determining price. The demand of manufacturers and dealers is not the ultimate demand; it is an indirect or reflected demand, arising from the demand of the "ultimate consumers." When now we turn to the study of supply, we observe that, with few exceptions, the sellers are not correspondingly parting with goods that might satisfy their wants. There are exceptions, as when a family in reduced circumstances is compelled to sell its household effects, and especially in connection with the sale of non-reproducible goods, such as rare postage stamps, antique furniture, or paintings of old masters. The great bulk of buying and selling however is of goods which are being currently reproduced or are at least capable of reproduction. Here there is an important difference between the ultimate buyers and the sellers. The purchaser buys a set of golf clubs in order to play the game. The merchant who sells him the clubs is not thereby giving up his opportunity to play golf, and the estimate he places upon the clubs has nothing to do with the pleasure of playing golf. The seller's conduct is not explained by a balancing of marginal utilities.

On the seller's side the motive is profit, or other pecuniary gain. This is true not only of the retail merchant from whom

[1] The term, producers, includes of course merchants and all others who have a part in the productive process.

the consumer does most of his buying. It holds also of the wholesaler, the jobber, the manufacturer, the farmer, and all other persons engaged in the various steps of production from the extraction of the raw material to the delivery of the finished good to the final consumer. This generalization applies also to those who sell their personal services, such as laborers, lawyers, physicians, musicians, and actors. The seller makes a profit when he sells for more than his goods cost him. He avoids a loss when his selling price just equals the unit cost of his product. Cost of production thus becomes the most important consideration on the seller's side of the market, just as marginal utility is most important on the buyer's side.

Production to order or for a general market. In the early stages of the medieval town economy most of the products of manufacture were "made to order." The blacksmith, the bootmaker, the weaver did not generally begin the making of anything until there had been an order from a buyer and an agreement upon the price. Since no artisan would knowingly undertake to produce and sell an article for a price that would not cover his costs and a "fair profit," there was here a simple relation between cost and price. Present-day examples of such production are the custom tailor, the custom bootmaker, the house-building contractor, etc.

As a matter of fact a very considerable part of the production of machinery, tools, structural steel, and a host of products that furnish either equipment or raw materials for other factories is made to the customer's orders. The relation between cost and price however is not quite so direct and obvious as in the case of the simpler industries cited in the previous paragraph, owing to the complexity of the industrial processes and the consequent difficulty of knowing in advance just what the cost per unit of product is destined to be, and owing also to the different elements that may be included in cost and to the different ways of measuring cost.

The relation between cost and price becomes far more obscure when production is not to order but for a general market.

Wheat, cotton, and virtually all agricultural products are obviously in this category. The same is true of most everyday articles which are sold for use to the general public rather than to other producers. Ordinary ready-made clothing, shoes, hats, bread, breakfast foods, and household furniture are a few examples from an almost unlimited list. In fact this condition is typical of the greater part of modern industry. Such goods are made up in advance of orders from buyers, and the producer takes the chance of being able later to find customers who will take his product at a profitable price. Under such circumstances we cannot say simply that sellers will not part with their goods except at prices that equal or exceed their costs. The complex relationship between cost and price will appear only after rather searching analysis.

Immediate sales problem: General statement. As the first step in our analysis of costs in their relation to price, we shall need to distinguish between what we may call *immediate* and *remote* sales considerations. The former are those that confront the seller — manufacturer or dealer — with respect to a stock of goods on hand, already produced. Such goods are in a special position, because all of the costs to which their production or acquisition has given rise have already been irrevocably incurred.

The most obvious illustration occurs in the case of non-reproducible goods. The past cost of producing a rare postage stamp, a painting by a famous old master, or a rare and beautiful piece of antique furniture has nothing to do with the price at which it can now be sold. The demand of buyers is such that the possessor's stock can be disposed of at prices far above cost of production. On the other hand, there are non-reproducible goods for which no one would today give as much as the cost — a painting by a past nonentity for example.

A somewhat similar situation exists in the case of articles which, though they could be reproduced, have gone out of fashion and are of so little present use that no one would con-

sider paying as much as it would cost to make them. A search through the barns and carriage houses of the United States would disclose many an old-time sleigh. Sleighing is pretty much out of fashion, and the fact that a sleigh, still in good condition, might now cost three hundred dollars to reproduce will not enable its owner to get any such price for it. If he wants to sell he must take what he can get, regardless of cost.

Even in the case of goods which are being currently produced, it is still true with respect to the present stock on hand that, since their costs have been irrevocably incurred, the seller's sole interest in them is to realize as large a return as possible now that they have come into his possession. Of course considerations of the more remote future will influence the selling policy which he adopts with respect to any particular lot, but the point is clear that for such goods there can be no absolute minimum below which no seller would go. Each seller would like to get back his costs and as much as possible in excess. But the cost — whatever it is — has already been incurred and is generally irrelevant to the seller's decision.

Reservation price. In facing his immediate sales problem, the seller does not of course have to accept a price low enough to take off his entire stock. Generally he will set a minimum which may be called his "reservation price." Various considerations influence sellers in setting their reservation prices. In the fresh fruit and vegetable markets, as the end of the selling period approaches (say late on Saturday evening), the dealers slash their prices ruthlessly in order not to be left with goods on hand which will spoil before the next market opening. Goods subject to the whims of fashion are similar to perishable goods. Thus as the summer wanes the department stores announce "sales," which mean that reservation prices have been cut in order to avoid having unseasonable and out-of-date hats, dresses, shoes, and such goods left on their hands. More durable goods are less subject to this influence. The cotton or wheat farmer can more safely decline to sell at the

present price, or he may sell part of his stock at the market price and hold out the rest for future sale.

Reservation prices are strongly influenced by the seller's estimates of what the future price of his product may be. The higher the expected future price, the higher will be the present reservation price. There will normally be a "spread" between expected future price and present reservation price equal to the cost of "carrying" the merchandise, consisting principally of storage cost and interest on the present value represented by the goods. For example suppose that a lumber dealer believes that the price of white oak flooring next year is likely to be $90 a thousand feet. Assume further that it will cost him $2.00 a thousand for storage for the coming year. If interest is at 5 per cent, he will have a reservation price of about $83.81. If he sells at this price, his receipts plus interest of $4.19 will amount a year hence to $88, which is what he would then receive less $2.00 storage cost. It will therefore be wise for this dealer to sell today as much of his stock of white oak flooring as he can dispose of at a price of $83.81, but not to try to sell more by offering a lower price.

Sometimes a dealer finds himself in financial straits and must turn his stock into cash regardless of cost. He sells at "sacrifice" prices, for which he may have no minimum reservation at all. In times of general financial panic many dealers may find themselves in this position, and a general collapse of prices may result. On the other hand, a sudden upturn of economic prosperity or the general expectation of such upturn may simultaneously influence all dealers to raise their reservation prices and so lead to a general increase in the prices of present stocks of goods. The main importance of these considerations arises from the light they throw upon the forward looking character of sellers' decisions. Once costs of production have been incurred or once a dealer has purchased goods, these past transactions have no significance for his decisions. Of course the price that he succeeds in realizing for a stock of goods already produced may determine whether

he will produce or buy any more goods of the same variety. But this leads us away from the immediate sales problem to a consideration of our next topic.

Remote sales problem: Price governed by cost. Business men, farmers, and producers generally will not be able or willing to remain in business indefinitely if they sell their products at prices which will not cover their costs. And we must note in this connection that costs include not only the payments made by the entrepreneur to other persons but also interest upon his own capital and an adequate reward for his own labor of management. The owner of capital has the choice of employing it in business or lending it to someone else. In the latter case he receives interest, and he will therefore be unwilling to employ his capital himself unless he can expect an equal return from his own business. In like manner the entrepreneur generally has before him the two alternatives of continuing to run his own business or of hiring out to an employer. Unless the income that accrues to him from his business promises in the long run to be at least as great as the wages or salary he could command, the business man will ordinarily not continue as an independent producer.

The producer who is unable in the long run to cover his costs, including interest on the capital invested, whether his own or borrowed from others, and wages of management, whether paid to a hired manager or earned by himself, will sooner or later give up and retire from business. The costs of production thus set a minimum below which producers will not in the long run offer to sell their products.

Production costs are effective also as regards the upper limit of price, provided there is competition between producers. Let us assume that several competing bakeries are offering bread to the consumers of a certain community. Each baker is striving to make his profits as large as possible, which he may accomplish either by a high price or by large sales. If in choosing the first of these alternatives, he sets a price which is substantially greater than the costs of production among compet-

ing bakeries, he will set a margin within which rival bakeries can underbid him while still covering their costs. While there may be a few customers who prefer the bread of this particular baker and will continue to buy it, the majority of his customers will tend to leave him in order to buy their bread more cheaply. It is thus that the costs of his competitors, or the general or "bulk-line" level of costs in the industry, tend, through their effect on the demand for any particular baker's bread, to prevent him from setting his own price too high. The more elastic the demand for a particular seller's product, the more will he be limited in his power to raise prices by the costs of production in the industry.

Future costs control. But the cost of production to which reference is here made is not past cost already incurred. Suppose a new invention has greatly reduced the cost of making a certain product. Since the stock may now be replenished at a lower cost, competition will compel producers to offer the product at a price related to the new cost of production, regardless of what the actual cost may have been. On the other hand, consider the case of a change which increases the cost of a certain commodity, such as a rise in freight rates or an agreement fixing a higher scale of wages. There will now be a new and higher cost scale governing all future production. Those who have stocks of the commodity on hand, knowing that they cannot be replenished except at a greater cost, will find that they can get a price governed by the new cost even for those articles which were produced at the former low cost, and this will control the supply offered by them.

In normal times the prices at which producers decide to offer goods are therefore controlled by future costs of reproduction and little, if at all, by the past costs at which existing stocks were produced. Of course no wise producer closes his eyes to the past or present scale of costs, and if there appears no reason to anticipate a change in the near future, the producer's policy is apparently determined on the basis of past costs. But that is simply because costs already experienced

have turned out to be the same as the later sales price, and it is only as an indication of what future costs may be expected to be that past or present costs enter into the problem.

Analysis of cost: Fixed and variable costs. Our next task is to obtain a definite idea of the meaning of unit cost of production. For this a careful analysis of cost is required. Let us suppose that a certain manufacturer is producing shoes, using a plant (land, factory, machinery, etc.) which cost $1,000,000. The interest on this capital is one of the elements in the manufacturer's cost of producing shoes. It is obvious that, this sum having once been invested in the plant, the interest will go on indefinitely, without regard to the number of shoes produced. Indeed the interest will continue though no shoes at all are produced; it cannot be escaped even by closing down the factory. Even if this particular owner should escape by selling the plant, the purchaser would have to assume the burden. Only by going into bankruptcy can the enterprise shuffle off its contractual interest obligation, and then the creditors suffer a lump sum loss equal to the future interest payments. Interest is therefore called a *fixed cost.*

There is another group of fixed costs, whose amount has little or no dependence upon the immediate number of shoes produced; this group includes such costs as the salaries of the principal officers and department managers of the factory, the payments for insurance, for lighting and heating the factory, etc. These costs are unlike the interest on capital however in that they would generally cease if the plant were closed down permanently.

Finally there are costs which are related closely to the quantity of the product, so that their amount varies with the amount produced. Thus the cost of leather and other materials, wages, and certain other costs will vary according to the number of shoes produced by the factory.

Production costs may thus be classified as follows:

1. Fixed costs: (a) interest, (b) other fixed costs,
2. Variable costs.

In the long run, all costs tend to become variable, for, though certain costs may be fixed within a given range of production, they also are subject to variation if production is permanently increased or decreased by very large amounts.

A numerical example will help us to arrive at the relation between these several groups of costs and the total cost of production per unit of product. Returning to our example of a shoe factory with $1,000,000 of capital invested in plant, let us assume that the annual interest charge, at 5 per cent, is $50,000. Let us assume further that the other fixed costs are $450,000 a year and that the variable costs are $3.00 per pair of shoes produced. What will be the total cost of making a pair of shoes? Evidently it will depend upon the number of pairs produced. The following table shows the results of three separate assumptions as to the output:

COST OF PRODUCING A PAIR OF SHOES

	First case	*Second case*	*Third case*
Amount of product	500,000 prs.	1,000,000 prs.	2,000,000 prs.
Interest	$ 50,000	$ 50,000	$ 50,000
Other fixed costs	450,000	450,000	450,000
Variable costs ($3 a pair)	1,500,000	3,000,000	6,000,000
Total costs	$2,000,000	$3,500,000	$6,500,000
Average cost per pair	$4.00	$3.50	$3.25

It should be especially noted here that this example does not represent a historical sequence; it sets forth rather the results of three simultaneous possibilities as to the quantity that might be produced. The illustration brings out with clearness the principle that, under these assumptions, the average unit cost of production varies inversely with the amount produced. The reason is simple enough; *i.e.*, since the interest and other fixed costs are constant in amount, the part of such costs assigned to each unit becomes successively less as the number of units increases.

Economy of fully utilized plant. In this illustration we see unit cost lowered by a "spreading of the overhead." By this we mean a reduction in average cost through the applica-

tion of fixed costs to a greater volume of production. This is not the only way in which unit cost may be reduced by increased output. Buying raw materials in larger quantities often makes it possible to procure them at lower prices. Thus an important part of the variable cost embodied in each unit of product is reduced. Larger production often makes possible greater economy in the use of raw materials and permits the manufacture of by-products which reduce variable unit costs. In similar manner, freight rates and other shipping charges, if included in total cost of a unit of product, may be reduced by an increase in total production. Within certain limits increase in output also makes possible greater economy in the use of labor.

But there are very definite limits to all these tendencies. Even if one million pairs of shoes can be produced more cheaply per pair than half a million, and two million pairs more cheaply than one million, it does not follow that there would be further reduction for three, four, ten, twenty million pairs. Let us return to our numerical illustration, in which it was assumed for simplicity's sake that variable costs were $3.00 a pair. Evidently when half a million pairs were made, the plant and the services for which fixed costs were incurred were not being fully utilized; otherwise it would not have been possible to increase the output to a million pairs and then to two millions with the same plant and the same general services. Sooner or later, if production is continuously increased, the time approaches when plant and general services are fully utilized. We arrive at the generalization that there is a tendency for each producer's unit cost of production to decrease as his product increases, up to the point where his plant is fully utilized.

Further expansion of output and unit cost. If the firm wishes to increase the output beyond this point, it will have to make a choice. Usually it will be possible to increase output further by resorting to various devices which permit still more intensive utilization of the existing plant. If the firm is already

running two shifts, a third shift can be added, or the plant may be kept open on Saturdays and Sundays and holidays. Sometimes the actual rates at which operations are performed can be speeded up by increasing substantially the labor force. Still another possibility is that production may be increased by making use of old equipment, less efficient than that ordinarily employed, which is kept to be used under exceptional circumstances. If output is increased in this way, it is almost certain to involve a large increase in variable unit costs. If men are hired to work overtime they must usually be paid at a higher rate. As the labor force is increased beyond its normal size, management and supervisory officers are overloaded and function less efficiently. As marginal equipment is called into use, variable costs will be increased disproportionately because of the inferior efficiency of the equipment.

The alternative choice before the business man who seeks an increase in output is to make additional investment in plant and an increase in the staff of executive officers, etc. This will mean an increase in interest and other fixed costs, and so in the total cost per unit. When the capital investment and the executive staff are increased, the increase will at the start almost certainly be greater than sufficient to care for the increased product which can be immediately sold. Therefore there will again be a situation of unutilized plant, with the possibility of another succession of reductions in unit cost as output is further increased.

The limit to large-scale enterprise. But this will not ordinarily go on indefinitely, since there is in most lines at least a limit to the size of the business unit, beyond which efficiency declines. The laboring force may become too large and unwieldy, wage costs may be rising, managerial efficiency may decline through inability to cope with the problems of an overgrown business. The limit of most efficient size for any particular enterprise depends chiefly upon general conditions in the industry which limit the advantages to be gained from large-scale production. In agriculture, for example, expe-

rience has demonstrated the advantage of the small farm. The ordinary farmer soon reaches the point beyond which he does not find it advantageous to enlarge his farm and its equipment. A different situation is illustrated by certain lines of manufacture — such as automobiles, cigarettes, oil refining — where there appear to be advantages of large-scale production which are not exhausted until the business unit attains very great size. This situation is still more extreme in the railroad business and certain other public utilities. The point which especially concerns us here is that each enterprise has a theoretical optimum size, at which under existing conditions it may produce at lowest cost. As demand for its product increases, the enterprise tends to increase its plant until that size is attained. When all the producers in any industry are in this state; *i.e.*, with enterprises of the size of maximum efficiency and plants generally fully utilized, the limit to the advantages of large-scale production has been reached.

Marginal cost: Definition. Application of the principles of cost of production to the problem of price will be facilitated by the introduction of a new concept; namely, *marginal cost.* When a producer is making a decision regarding a possible policy which would lead to an increase of his sales, he must consider two major effects of the proposed step: (1) the change in his revenues and (2) the change in his costs. If the proposed step will reduce costs more than it will reduce revenues or increase costs less than it will increase revenues or reduce costs and increase revenues, then the change will be profitable; otherwise it will not. To arrive at an estimate of the probable effect upon costs, the business man cannot consider only average costs, with which we have been concerned thus far. He needs the concept of marginal cost.

We may define *marginal cost* as *the amount per unit by which a small change in output changes total cost.* [1] The significance of the

[1] Since this amount would be calculated by dividing the increase in cost by the number of units added to output, marginal cost is a *rate*, and it is sometimes defined as "the rate at which a small change in output changes total cost."

concept may perhaps be made clearer if we think of marginal cost as the difference in total cost that will be occasioned by a change of one unit in output. This description makes marginal cost exactly parallel with the concept of marginal utility as we have defined it. It is obvious however that it is generally impossible to measure the increase in cost that will result if just one more unit is produced or the saving in cost if one less unit is turned out. For instance if a shoe factory had been planning to turn out 500,000 pairs of shoes, and we attempted to discover by how much its cost would be increased if it turned out 500,001 pairs, the calculation would be impossible. However the more minutely changes in total cost can be calculated, the more accurate is any measure of marginal cost.[1] In actual practice a fairly rough measure is generally the best that can be obtained.[2]

Marginal cost and average cost. The nature of marginal cost and its relation to average costs may be given more concrete illustration by returning to the example of the shoe factory which has already served us. Let us revise the assumptions however in two respects; in the first place, let us consider this the business unit of most efficient size, and second let us recognize the probability that variable costs will not be $3.00 a pair throughout.

COST OF PRODUCING A PAIR OF SHOES

Amount of product	490,000 pairs	500,000 pairs
Interest	$ 50,000	$ 50,000
Other fixed costs	450,000	450,000
Variable costs	1,967,000	2,000,000
Total cost	$2,467,000	$2,500,000
Change in total cost		$33,000
Change in output		10,000 pairs
Marginal cost		$33,000 ÷ 10,000 = $3.30
Average of variable costs		$2,000,000 ÷ 500,000 = $4.00
Average of all costs		$2,500,000 ÷ 500,000 = $5.00

[1] The pure mathematical concept of marginal cost is based on the assumption of infinitesimal changes in cost and output.

[2] Cost accountants have developed a working concept known as *differential cost* closely related to the economist's idea of *marginal cost*.

In this illustration it is assumed that 10,000 pairs of shoes is the smallest change in contemplated output upon which we can base a calculation of marginal cost. The effect of a change in production from 490,000 pairs to 500,000 pairs is to change total costs by $33,000. Dividing this amount by 10,000, we have a fairly accurate measure of marginal cost. $3.30 is the rate per pair at which total cost changes with an increase in production from 490,000 to 500,000 pairs, or with a decrease in production from 500,000 to 490,000 pairs. $3.30 is said to be the marginal cost of 500,000 pairs of shoes. According to our general definition it would be logical to call $3.30 the marginal cost of either 490,000 pairs or 500,000 pairs; but to prevent confusion it is generally understood that, when a small change in output takes place through addition, the rate of additional cost is to be taken as the marginal cost of the quantity which includes the added units. Thus the same marginal cost is not attributed to two different quantities; and marginal costs through addition and through subtraction are regarded as the same for any particular quantity.[1]

It is important to note that only in special situations is either the average of variable costs or the average of all costs equal to marginal cost. In the example above, marginal cost ($3.30) is less than the average of variable costs ($4.00), which of course has to be less than the average of all costs ($5.00).

Relation of costs to output. The three magnitudes — average of variable costs, average of all costs, and marginal cost — as well as their relations to each other, will vary in accordance with the amount of the output of the business. We may illustrate this by the set of suppositions on the following page (Case 1 being the example we have already considered).

It will be observed that the schedule below is in harmony with principles which we have already developed, in that the average variable cost, as well as the average of all costs, decreases with an increase in the output, but only up to a certain

[1] Compare the similar convention with regard to marginal utility, Chapter VI.

SCHEDULE OF UNIT COSTS IN THE PRODUCTION OF SHOES

	Possible outputs	*Averages of variable costs*	*Averages of all costs*	*Marginal costs*
Case 1	500,000 pairs	$4.00	$5.00	$3.30
Case 2	1,000,000 pairs	3.50	4.00	2.70
Case 3	2,000,000 pairs	2.85	3.10	2.20
Case 4	2,500,000 pairs	2.80	3.00	3.00
Case 5	3,000,000 pairs	3.00	3.16¾	5.00

point, after which each of these averages increases. Marginal cost likewise decreases and then increases, the turning point however coming with a smaller output than for either average variable or average total costs, due to the fact that marginal cost is at first less than the average of all costs. The reader will of course not fail to observe that we are dealing here, not with a historical sequence, but with a series of simultaneous possibilities in the sense that, beginning with a given moment, any of the indicated possibilities might be realized at the expiration of the period necessary for production.

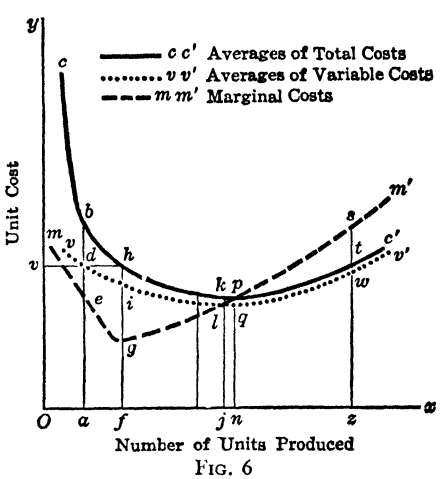


FIG. 6

By the same general methods previously used in constructing utility and demand curves, the schedule of costs can be converted into graphical form. Since the table includes, however, three different kinds of unit costs, its graphical representation is not one curve but three, as shown in Figure 6. The curves *cc′*, *vv′*, and *mm′*, respectively, represent various possibilities as to the average of all costs, the average of variable costs, and the marginal cost, under the given circum-

stances. These curves really show more than the cost schedule. Many intermediate cases are understood to be included, and all three curves have been smoothed. If we were to consider changes in interest and other fixed charges, definite breaks (discontinuities) would have to appear in the curves in Figure 6.

EXERCISES

1. Assume the following conditions of cost for the production of shoes in the X Shoe Company plant:

Output	*Fixed cost*	*Variable cost*	*Total cost*
10,000	$100,000	$ 31,000	$131,000
20,000	100,000	60,000	160,000
30,000	100,000	90,500	190,500
40,000	100,000	123,000	223,000
50,000	100,000	158,000	258,000
60,000	100,000	196,000	296,000
70,000	100,000	239,000	339,000
80,000	100,000	289,000	389,000
90,000	100,000	350,000	450,000
100,000	100,000	429,000	529,000

(*a*) Prepare a cost schedule showing for each output the average fixed cost, average variable cost, average total cost, and marginal cost of producing that output.

(*b*) Plot on a sheet of graph paper the corresponding cost curves.

(*c*) Approximately what is the optimum size, or point of maximum efficiency, for this concern?

Competition and the Market

The market: Definition. The word "market," in the vocabulary of economics, has a somewhat broader content than is given it in ordinary usage. As technically defined, *a market for any good is a place where buyers and sellers exchange that good.* This includes obviously all those markets which we have been studying. It includes also every place where buyers and sellers exchange goods even though there may be no formal market organization. Thus the retail strawberry market of a city need not be any formally organized market in a special building or place; it may be simply the various retail stores and offices where dealers and consumers sell and buy strawberries. It is not even necessary that all the buyers and sellers be physically present at the market. They may be in communication with the market by telephone, telegraph, or mail and represented there by agents or brokers. The market is simply the place at which buyers and sellers make their influence felt, at which exchanges are made, the meeting point of the forces that determine price.

Free market. *A free market is a market in which there is no rationing of goods and no compulsion on either buyers or sellers to go through with transactions which they do not wish to carry out.* In our following discussion it will be assumed, unless otherwise stated, that the term market means a free market as thus defined.

Setting the price. Economists customarily speak of "the prices established by the market," but it may well occur to the reader that a market is a human institution, made up in the last analysis of people, and that when we speak of a price being determined by the market, we mean that the individual or individuals who set the price have been induced to set it

at some particular level by forces which operate on them in the market.

In most markets in modern industrial nations price is set by the sellers. If you, as a consumer, enter either a large department store or a small grocery store, you will find in every case that there is a certain established price for each article in the store. You are free to buy as much or as little at that price as you choose, but it is rare indeed that you will be able by your own unaided action to secure a change in the price.

This kind of market we may contrast with those markets where the prices are named by the buyers. When the large milling companies and grain dealers purchase wheat from the farmers throughout the western United States, it is the buyers who set prices. The farmer coming in with a truckload of wheat to the elevator can either accept the price which he is offered or take his wheat home. He has the option of selling as much or as little at that price as he wishes.

In both of these cases it is of course conceivable that the individual buyer in the store or the individual farmer might be able to secure some change in price and would thus have some effect upon prices through his own actions, but even so it is always the seller, in the first case, and the buyer, in the second case, who actually names the price at which transactions are made.

There is a third type of market which differs from both of these patterns. In the wheat pit or in the stock market or even in dealings between one buyer and one seller, price is apt to be decided by a process of bidding and counterbidding. A buyer may go into the wheat pit and offer to purchase a thousand bushels at 68 cents a bushel. A seller may offer to sell a thousand bushels at 72 cents. Either party may name a price and in such a highly organized market there will always be sellers willing to make sales at a price just above the current market price and buyers willing to purchase at a price just below it. The bidding and counterbidding will of course take a different form if, for example, the president of a manufactur-

ing company is negotiating for the purchase of a piece of special machinery from a manufacturer of machinery. But here again each party will name a price at which he would be willing to make the transaction, and the price finally agreed upon will be the result of their bargaining back and forth. We may thus distinguish three sorts of markets: those in which the buyers name prices, those in which sellers name prices, and those in which either party may name a price.

Competition: Definition. From the seller's viewpoint, the most significant classification of markets relates to the degree of competition which prevails between himself and other sellers. *Competition is the effort of two or more sellers to sell to the same person or persons, each seller acting independently in his own interest without regard to the interests of other sellers.*[1]

Competition and monopoly. Degrees of competition. The antithesis of competition is monopoly, a topic to which we shall give attention in a later chapter. It will be at once evident that the absolute extremes of either perfect competition or perfect monopoly are seldom found in the world of today.[2] The actual condition is almost without exception somewhere between these extremes. In dealing with practical problems it is therefore seldom possible to use the terms competition and monopoly in any rigid sense. As used, these are relative terms, indicating a degree of approach to one or the other extreme. Various qualifying terms have thus arisen in the literature of economics. "Pure competition" is used to describe a state that approaches the competitive extreme but is less absolute than perfect competition or is "per-

[1] This is obviously competition of sellers, and there may be also competition of buyers, with an analogous definition. The latter concept is however seldom employed, and where the term competition is used without qualifying words, competition of sellers is to be understood.

[2] In strict theory there is not perfect competition unless a number of rigid requirements are fully met: (1) The total number of competitors must be indefinitely large. (2) Each competitor must sell exactly the same product, and it must be so regarded by all buyers. (3) Each competitor's stock must be so small in comparison with the whole group of competitors that no change in his output has any effect on the price. (4) Competition must be so keen that if any individual seller raises his price above the others, he will lose all his business.

fect" in only one respect while falling short of perfection in other respects. "Imperfect competition," as the term is generally used, denotes a state which may still be regarded as competition, though it is far "less perfect" than pure competition. For the purposes of this elementary book two degrees only of competition will be sufficient; for these we choose to use the terms *pure competition* and *imperfect competition.*

Pure competition: Market conditions. Definition. Pure competition is characteristic of those markets in which the buyer normally sets the price and of the great speculative markets where prices are determined by the bidding of buyers and sellers. The essential features of this type of market are as follows. The article dealt in is a standardized product. All units offered by the different competitors are exactly alike, and no buyer has any reason for preferring to deal with any particular seller. The number of sellers is very large, and the quantity that could be offered for sale by any seller is a very small part of the total quantity offered.

This situation is illustrated by the regular markets for many great commodities, such as wheat, cotton, and certain other staple agricultural products, and by the stock markets, dealing in corporation securities, government bonds, etc. We do not usually think of individual wheat farmers or cotton growers as competing with one another. Yet inasmuch as they are all selling a standard commodity, they are competitors as truly as two manufacturers each of whom sells electrical refrigerators or automobiles. What obscures the situation is the fact that there are so many of them that each one has a negligible influence on the prices established on the market and so on the fortunes of his fellows. The largest wheat grower in the world does not produce more than a small fraction of the world output of wheat.

Pure competition may be defined as *competition between sellers under such conditions that the individual seller is unable by his own actions appreciably to influence the market price.*

The individual seller meets perfect elasticity of demand. The position of the individual seller under conditions of pure competition may be described in terms of the demand for his product. The farmer comes to market and finds that the price of wheat is $1.25 a bushel. At that price (or by shading it ever so slightly) he can sell any amount of wheat he wishes, and he will not be able to sell any wheat at all at a price even a small fraction of a cent higher. His sales will not affect the price. Whereas the total demand for wheat in the market at

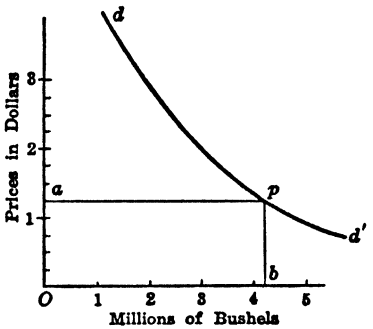


FIG. 7

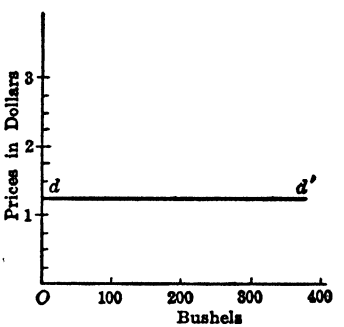


FIG. 8

this time may be represented by the normal downward sloping curve, as in Figure 7, the demand for the wheat of the particular farmer is represented by a horizontal line, as in Figure 8. Such a line is the graphical representation of perfect elasticity of demand. And it is strictly true that the demand for the product of the individual seller in such a market is completely elastic. This is simply another way of saying that the seller has to accept the market price as he finds it. This will be true for any producer who sells his product in a market of which he is so small a part that he cannot influence the ruling price.

It should be noted that the elasticity of demand for the individual producer's output has very little to do with the elasticity of total demand for the commodity which he produces. It is obvious that, although the demand for foodstuffs and

especially for wheat is highly inelastic, yet there is a perfectly elastic demand for the output from any one farm.

The seller has no "price problem." Whereas the manufacturer of almost any branded or trade-marked product, such as a washing machine or a particular kind of corn flakes, must before setting his price think carefully of the probable consequences on the amount of his sales, the wheat farmer has no such "price problem." He does not have the problem of deciding to set a price of, say, $1.27 on the wheat he brings into the grain elevator. Of course, since he sells in a free market, he could be said to be free to set the price at which he would sell instead of the quantity. But this freedom would be rather illusory, because all it means is that he could set his price slightly above the market price if he were willing to sell none of his wheat or below the market price if he were willing to forego voluntarily a part of the money which he might receive. The reason, in other words, that he does not choose to "set the price" is that there is no conceivable advantage to him in setting a price which differs from the established market price.

Imperfect competition: Inertia and ignorance. Once the characteristics of pure competition are understood, it will be easily recognized that in most markets competition is generally not pure. There are many factors which in actual practice prevent the attainment of pure competition.

In the first place, if a market is not well organized, if buyers and sellers have difficulty in coming together and comparing each other's wares and each other's prices, we have imperfect competition. Competition in the retail business is often imperfect to a pronounced degree, because it is so much trouble for buyers to get around and find out exactly what each store is charging for each and every article. Furthermore it is often difficult for buyers to determine how the quality of goods offered by one store compares with the quality in other stores. Competition in the wholesale business and in the great speculative markets much more nearly conforms with the analytical

idea of pure competition, chiefly because the buyers are better informed and more energetic to seize any advantage.

Individuality of products. Competition is imperfect in the technical sense also when the products of competing sellers are not precisely alike. It is typical of the field of industry, as opposed to that of agriculture, that the products of particular producers have considerable individuality. The individual producer is not selling grade A wheat or even a standard quality of cement; on the contrary he is selling a particular kind of machinery or a particular brand of beer or a particular make of automobile. Wherever this is true, individual consumers will have preferences for one make or another. Each producer will have a number of regular and loyal customers who can be depended upon to buy his product even if its price is somewhat higher than others of similar quality. If the price of one make falls a very little below that of another, all buyers will not desert the other makes in a body and concentrate upon the one with the slightly lower price. In the automobile industry, for example, there are many peculiarities in detail among the three close competitors, Ford, Chevrolet, and Plymouth. It is obvious that the demand for Mr. Ford's cars will depend very directly upon the price that is being charged for Chevrolets and Plymouths. But it is unthinkable that a moderate price cut by Chevrolet and Plymouth would completely deprive the Ford Company of its customers. In this respect the position of the automobile manufacturer offers a sharp contrast to that of the wheat farmer.

Trade marks and special services. Even though the products of different manufacturers may be virtually identical, the sellers frequently succeed in creating the impression of differentiation in the minds of the buyers, especially in the retail markets. Each sugar manufacturer puts up his product in a different package and claims for it distinctive advantages. Canned goods have various labels and various contents. Meats, clothes, gasoline, and very many other everyday products in modern America are offered by each competing

producer in a form which he tries to individualize so that it will be known as his own. Brands, forms of packaging, distinctive coloring, slogans, all sorts of advertising are some of the devices used to gain this effect and so to attract a particular following among the buying public.

Consciously or otherwise each individual store develops for better or worse an individuality in its offerings, even though it sells the same brands as another store across the street or another store around the corner or one in the next block. What people pay for in a store is not just merchandise but also the service that goes with it, the general atmosphere of the place, and the convenience of its particular location. These are some of the things which tend to make competition imperfect.

Small number of sellers. Even in the case of standardized industrial products like cement, it is plainly true that the individual producer cannot sell as much as he wishes at a good market price. For in these industries the number of competitors is generally small enough relative to the size of the market so that no single producer can expand his sales substantially without "spoiling the market" for his competitors and probably for himself. In such industries as these a single producer may sell up to ten or fifteen per cent of the total sales in a given market. (Of course this is not the same thing as ten or fifteen per cent of the total sales in the country.) If he attempts, let us say, to double his output, it is clear that the market is going to be disturbed; he will not be able to increase his sales so largely without either engaging in an intensive campaign of advertising and sales promotion or — what is more likely — cutting his prices. Either one of these methods will provoke his competitors to retaliation. If he cuts prices they will be forced to follow, and if he organizes an aggressive sales campaign they are likely to do the same. It is of the very nature of such an industry that no one producer can expand his output very greatly without being willing to sell at lower prices or to incur much heavier selling

costs or both. The individual producer is too large a part of the particular market in which he sells to have it unaffected by the quantity of which he disposes.

Demand not perfectly elastic. We thus observe that competition may be imperfect (1) because of the ignorance or inertia of the customers of the retail markets, or (2) because of individuality of competing products — whether actual or imaginary, or (3) where there are too few competitors to make the influence of any one of them negligible, even though they are all selling a standard commodity. When any one or more of these conditions is present in a market, the elasticity of demand for the output of a single seller will not be infinite or perfect. Henry Ford knows very well that he cannot expand sales without lowering price or increasing his selling outlays. Even the small independent grocer in a large city where the number of sellers is very large and where the products which they are selling are practically standardized is in a different position from the wheat farmer. In all such cases the seller will have a "price problem," and it is notable that virtually all of the markets in which the elasticity of demand for the individual seller's output is not absolute are ones in which the sellers name prices.

Definition of imperfect competition. The competition between sellers under these conditions is *imperfect competition*, which we define as *competition between sellers under such conditions that the individual seller is able by his own actions to exert some influence upon the market price.*

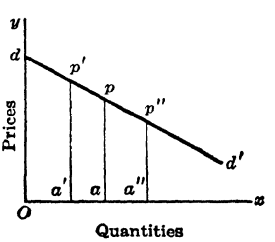


FIG. 9

These matters adapt themselves readily to graphical illustration, as in Figure 9. Let us suppose we have an automobile manufacturer who comes directly into competition with only three or four other makes of car. The average price of these other cars is indicated in Figure 9 by the height *pa*. The number of automobiles he puts on the market is determined,

not by price alone, but also by his interpretation of the conditions of demand. He does not have to charge the same price as rival manufacturers. He can charge a higher price such as $p'a'$, or a lower price such as $p''a''$. At either price buyers can be found, as shown by the demand curve dd'. On the other hand, the presence of these other cars on the market puts a distinct limit on the height to which this particular manufacturer can successfully raise his price. Because of their presence the demand for his particular car is extremely elastic. If he raises the price to Od in Figure 9, he will have no sales. Nevertheless the fact remains that up to this level he can charge any price he wants to in accordance with his own judgment of demand, competitive conditions, and costs of production.

The liberty of imperfect competition. Although the seller's problems are more complicated when he sells under conditions of imperfect competition, another characteristic of such a situation is the comparative degree of liberty enjoyed by individual competitors. This statement may at first appear somewhat paradoxical, because it is the competition between sellers under these very conditions that is usually thought of as being the most vigorous. Farmers never advertise or send salesmen into one another's territory. The economic misfortunes which befall them are usually thought of as attributable in some impersonal way to the market. But when two or three manufacturers sell their products in a market where there are no other sellers, the competition among them takes a very different form. If producer A increases his sales at the expense of producer B, it is clear to everyone what is the source of producer B's difficulties. Because the number of producers is usually much smaller than under pure competition, the competition among them takes on a more personal form and is apt to be more bitterly resented.

Nevertheless in such a market the individual seller has a freedom which does not exist under pure competition. Unlike the wheat farmer, he may raise his price, perhaps sub-

stantially, without losing all of his sales. On the other hand, the seller in imperfect competition cannot expand his sales indefinitely without lowering his price or else spending more money on advertising, salesmen's commissions, etc. As compared with the seller in pure competition, he is in a position where he has greater freedom to raise his price and where the penalties of raising it are less. It is however much harder for him to expand his sales, and the cost of so doing is certain to be greater.

Limitations upon the seller's liberty. Although each individual competitor thus has a degree of liberty in the determination of his own price, this liberty may be so restricted as to be almost negligible, or not worth using. Apparently a business man may have few rivals in his particular field; actually and indirectly, he may have many. Thus in a particular town there may be only very few stores selling a given kind of product, and competition may appear to be solely a rivalry among themselves. All of these stores however may be competing, not only with each other, but also with mail order houses and with larger stores in a nearby city. True it is that not every shopper has time or inclination to price the article he wants in every shop before he buys it; but some people do go to much trouble in this direction. It is true that every alert business man tries to individualize his product and his particular type of service. He may thus succeed in establishing more or less permanent business relations with some of his customers. But at the same time a good deal of his trade may be with casual buyers who care no more for him than for any of his rivals. To them one particular brand of merchandise is about the same as several others, and several possible places of purchase may be equally convenient. In spite of the fact that he could raise prices and still do business, the man with established customers often finds it advisable to keep prices down to a freely competitive level in order not to lose the business of alert and independent people who can be easily attracted elsewhere by lower prices.

Liberty and discretion in big business. In the matter of price it may appear that the larger business enterprises have a greater degree of liberty than is enjoyed by small businesses. Nevertheless there is need for discretion even in the management of an outright monopoly. There is the restraining influence of possible and actual substitutes, a risk which each large concern shares with its rivals; there is the danger of additional competition, the danger of stifling future demand by too high a price in the immediate present, and the possibility of all sorts of popular and political outcries. In addition the large concern which is not a monopoly is held in check by its competitors. It may still have left a considerable degree of liberty in its price policy, but this liberty has to be used with caution.

It is noteworthy that in those industries where competitors apparently have the greatest individual freedom, prices are sometimes subject to the most conservative adjustments. A big concern, though bound to no particular price by external conditions, is often forced by the very complexity of its own internal organization to adhere to a relatively stable price policy. Price changes involve various notifications to the public, the instruction of numerous employees, and all the red tape characteristic of a big organization. Price changes also complicate bookkeeping. It is no wonder then that the management of any big concern, engaged in multitudes of little transactions with the general public, prefers to make as few price changes as possible. Imagine the confusion which would result if Ford or General Motors changed Detroit f.o.b. prices every day or if retail prices in a big department store were subject to hourly fluctuations. The relative stability of prices in industries where competition is concentrated in few hands is *ipso facto* not a demonstration of secret agreements or of arbitrary monopoly power but rather an inevitable result of the complexity of such enterprises and the need for conservative managerial discretion.

Quality competition in place of price competition. The unsettling results of too many changes in prices have a strong

tendency to put competition between large concerns on the basis of quality, salesmanship, and advertising, rather than price change. Instead of selling the same car for less money, automobile manufacturers offer or pretend to offer more car for the same money. Business men generally prefer this sort of competition. It calls for alertness and ceaseless activity and the expenditure of large sums of money in advertising and technical research; but the efforts which each competitor thus puts forth in his own behalf do not provoke the same quick and mutually injurious retaliation as would outright price cutting. For example, if a tire manufacturer has some degree of success in a selling campaign in which he stresses a particular process of treating rubber, he may be sure that his competitors will come forth with something of the same sort; but there the thing will stop until some other new idea crops up in the industry. On the other hand, if a price war is begun, no one can tell where it will stop or how injurious it may be to all competitors concerned.

Advertising. The modern business man is aware that demand may be subject to outside influence, and when such is the case with respect to his own product he does not hesitate to do his utmost to modify demand in his own interest. To this end the business man employs a sales force, large or small as the case may be, to whom he pays salaries and commissions; retailers are given special "margins" as inducement to "push" a particular maker's product; exclusive agencies are set up; and finally there is the whole field of advertising. Whether this is worth attempting or not depends, in the first instance, upon the degree of competition in the particular market. When a product is sold under conditions of pure competition, it will never pay individual producers to employ advertising and other devices to influence consumer taste. It would quite obviously be ridiculous for a Montana wheat farmer to advertise in a national magazine and urge consumers "to use more wheat and eat more bread." The absurdity arises from the fact that he as an individual is sup-

plying so tiny a fraction of the supply of wheat consumed in the United States that, even if his advertising increased the sale of wheat 100,000 bushels per annum, the effect on his individual sales would be negligible. But where a producer is one of a small number of sellers in the market or is the seller of an individualized product, he may find it worth while to advertise.

This is one of the salient features of imperfect competition. We have had occasion before to note the impersonal character of most markets characterized by pure competition; this absence of advertising under conditions of pure competition is one aspect of that same impersonality. We have noted also that the seller who faces imperfect competition, unlike the seller in a perfectly competitive market, has a "price problem." We now see that his situation is further complicated by a "sales problem." In other words, he will have to decide not only what is the most profitable price to set, in view of the fact that a lower price will expand his sales and a higher price will check them, but also how much money he is going to spend in pushing the sales of his product.

The statement that in a market characterized by pure competition advertising does not pay refers of course to the individual seller. It does not follow that, even in case of pure competition, it may not be profitable for many sellers to engage in a joint advertising campaign. Thus the producers of wheat might conceivably, through a national organization, seek by advertising to increase the demand for wheat and so its price, to the benefit of each producer. There is a great deal of national advertising conducted by associations of producers which it would not pay any one producer to undertake. In some of these cases the product is standardized enough and the number of sellers sufficiently great so that there is approach at least to a condition of pure competition.

The endeavor to influence demand through advertising has become an enterprise of stupendous proportions. The following is a partial estimate of the amount of money expended on

three of the principal kinds of advertising in the United States in the first eleven months of the year 1939.[1]

Newspapers (52 cities only)	$1,125,446,000
Magazines	138,698,000
Radio	74,988,000
Total	$1,339,132,000

This is obviously far from a complete account. It does not include such important types as direct advertising, street car advertising, billboard advertising, business papers, etc.; only 52 cities are included in the newspaper advertising, and finally the record does not cover the complete year.

Although few businesses are today able to function without resort to advertising, the volume of advertising naturally varies greatly as between different lines of industry. "The advertising of automobiles; toilet articles and medicinal preparations; food, groceries, and beverages; and cigars, cigarettes, and tobacco constituted 60.5 per cent of the total national advertising in newspapers in these cities (49 cities of the United States) in 1927." [2]

The advertising in the first eleven months of 1939, to which reference is made above, was distributed among certain important objects as follows:

DISTRIBUTION OF CERTAIN ADVERTISING IN THE UNITED STATES
FIRST 11 MONTHS OF 1939 [3]
(In per cent)

	Radio	*Magazines*
Automobiles and accessories	9.2	15.6
Clothing	0.7	5.6
Electric household equipment	0.0	2.9
Financial	1.2	3.1
Foods, food beverages, confections	32.1	15.6
House furnishings, etc.	1.0	5.1
Soap, cleansers, etc.	11.8	3.2
Office furnishings, supplies	0.0	1.5

[1] United States Bureau of Foreign and Domestic Commerce. *Survey of Current Business*, January, 1940, p. 23.
[2] M. T. Copeland, in *Recent Economic Changes in the United States*, 1929, p. 415.
[3] *Survey of Current Business*, January, 1940, p. 23.

	Radio	*Magazines*
Smoking materials	14.0	5.0
Toilet goods, medical supplies	26.8	16.7
All other	3.2	25.7

	Newspaper advertising
Classified	11.4
Display, total	43.7
Automotive	2.4
Financial	1.0
General	9.5
Retail	31.8

EXERCISES

1. Name various ways in which sellers may seek to distinguish their products from those of competitors.

2. Mr. A grows cotton and sells it at 10 cents per pound. Draw the appropriate demand curve for his product. If he should decide to charge a higher price, what would be the effect on his sales? If he should increase his output what would be the effect on price? Would he be justified in advertising his product? Explain, in each case.

3. Mr. B is a manufacturer of toothpaste, which is attractively packaged, nationally advertised, and sold under a brand name. Last year he sold 1,000,000 tubes at 25 cents each. Draw the appropriate demand curve for his product. If he should decide to charge a higher price, what would be the effect on his sales? If he should decide to lower his price, what would be the effect on his sales? Is he justified in advertising his product? Explain, in each case.

4. In retail grocery stores the sale of "bulk" coffee, crackers, and candy has given way to the sale of similar products attractively packaged, nationally advertised, and sold under brand names. Account for this development and explain its significance.

CHAPTER X

Cost of Production and Supply under Pure Competition

Pure and imperfect competition. We have now arrived at a fork in the path of our investigation of the forces determining price, and we must travel successively the two roads. The present chapter and the next one (Chapter XI) will deal with the problem of price under conditions of pure competition. In Chapter XII we shall come back to the fork in the road and proceed to investigate the corresponding problem of price under imperfect competition.

Under pure competition, as we have learned, the producer has complete freedom of choice as to his output in so far as the capacity of his plant permits; he must however accept the market price. Having no price problem, the important question confronting him is how much to produce for sale.

Cost of production and production policy: Price above average cost. In order to be specific, we may return to the shoe factory of the last chapter, positing now a state of pure competition. Figure 10 (page 160), a simplified version of Figure 6 in Chapter VIII, will serve as illustration. Let us assume that the price is equal to *zs* ($5.00). From the marginal cost curve, it is evident that the producer can afford to sell the number of units represented by *Oz* (3,000,000 units). The price *zs* ($5.00) not only covers *zw*, the average of variable costs ($3.00), but also covers *zt*, the average of all costs ($3.17) and permits a clear profit of $1.83 a pair. What is much more important to note however is that when the producer is making *Oz* (3,000,000) units, the price is just equal to his marginal cost (also represented by *zs*). This means that the price covers the additional cost incurred by the last increment of production up to 3,000,000 pairs, and that it more than covers increases from smaller amounts, as indicated by the

marginal costs of quantities less than Oz. If the producer were turning out less than Oz units, he would be in position to increase his profit by expanding output, for any increase in output up to Oz would add less to his costs than it would to his total receipts. Each additional unit that he sells will bring in $5.00, the ruling price. Each additional unit that he produces until he is turning out Oz units will increase his total costs by less than $5.00. Thus it would clearly not pay him to stop short of this output.

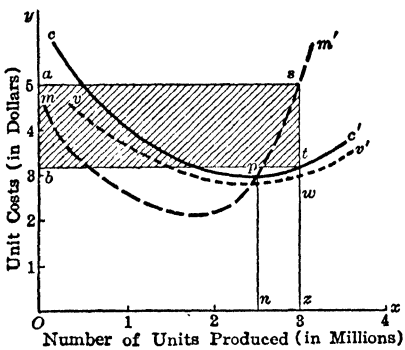


FIG. 10

These conclusions are illustrated graphically in Figure 10. When selling the quantity Oz at price zs, the average profit per unit is represented by ts, being the difference between the price and the average cost for this quantity. The total profit is represented by the shaded rectangle $abts$, which is the product of the quantity by the average profit per unit and at the same time is the difference between the rectangle representing the total receipts, $Ozsa$, and the total cost rectangle, $Oztb$.

The same reasoning demonstrates that the producer will not be able to sell more than Oz units without lowering his profit. From the marginal cost curve it is clear that if he sells 3,010,000 units instead of 3,000,000 units, the additional 10,000 units will increase his total costs at a rate higher than zs ($5.00), whereas his receipts will increase at exactly $5.00 a pair. Therefore it does not pay the producer to sell an extra 10,000 units at price $5.00, and certainly no larger amount. If he does, he may still have a profit on his operations as a whole, but he will be making his total profit less than it need be. The amount which this manufacturer will produce and sell at price zs ($5.00) — that is, the output which will render

his profits a maximum — is the quantity Oz (3,000,000 units). By reference to the rising portion of the marginal cost curve we can determine the amount the producer will sell at any other price, provided this price is greater than np. The reader will note that, whenever the price is above np, the shoe factory will produce an output greater than On and will always be making a profit, since to the right of On the marginal cost curve lies above the average cost curve.

Price below average cost. But what will be the situation if the price is below the firm's average total cost; *i.e.*, below np in Figure 10? Since np is the lowest point on the average cost curve, there will clearly be no output for which the cost does not exceed the price, and the firm is bound to lose money. But if the price is higher than the lowest point on the average variable cost curve, there are certain outputs at which the firm can at least recover its variable costs and have something toward meeting the fixed costs. This, though still leaving him "in the red," is better than closing the factory, since in that case his fixed charges would still go on, with no income at all to pay them. Under such conditions therefore it will still be to the firm's advantage to produce such quantity as will make its marginal cost equal to the price. Though a profit is not obtainable, though a loss must be borne, the producer at any rate will thus minimize his losses, for the time being.

Finally we have to consider the situation in which the price is below the lowest point on the firm's variable cost curve. Now the manufacturer is unable to recover even his variable costs; no matter what quantity he might choose to produce, he would be continually losing all his fixed costs and something more. Even so a producer may sometimes continue to operate — temporarily. Hoping for better times in the future, the producer may choose to bear a present loss in order to keep his plant from deteriorating or to hold together his organization and an experienced labor force. Ordinarily an enterprise that is losing more than its fixed charges will very soon find it best to close down.

Conclusion. We may sum up our investigation to this point in the following three propositions concerning output and price. (1) So long as the firm continues in operation under pure competition it will seek to produce that output for which marginal cost is just equal to price. This output will maximize its profits when price exceeds average total cost and will minimize its losses when price is below average total cost. (2) A firm will not continue production in the long run unless it receives a price that is sufficient to cover average total costs. (3) A firm will continue operations in the short run so long as the price exceeds average variable costs, and it may continue to operate for a short time even if price falls slightly below average variable costs. These three propositions describe the way in which the firm's cost schedules will determine the amount that it will produce, always assuming that it can sell any amount which it chooses at the going market price.

Individual supply. For the study of the seller's actions under conditions such as we are now considering the concept of supply is fundamental. *An individual's supply of any good is a schedule of the respective quantities of that good which he is ready to offer for sale at all possible prices.* The reader will observe that this definition corresponds precisely to the definition of individual demand, which was developed in Chapter VII. The ordinary relation between price and the quantity offered for sale by an individual follows readily from the principles which we have now developed. As long as any industrial enterprise continues to operate under conditions of pure competition, the major portion of its rising marginal cost curve (see, for example, the segment *lm'* in Figure 6, page 141) becomes the basis of its supply curve.

We have learned that when competition is pure each firm is free to produce whatever quantity it wishes. At any given time each firm will therefore be producing the quantity that will maximize its profits or at least minimize its losses, which means that production will always be pushed beyond the

point where the marginal cost curve begins to slope upward. In other words, the firm cannot attain to a position of stability in its production policy until its marginal cost curve is sloping upward. These rising marginal cost curves are, as we have seen, the supply curves of the several individual firms. They illustrate the general principle that, in a perfect market, the quantity of a good offered by an individual seller normally varies directly with the price.

Total supply: Definition. *The supply of any good is a schedule of the respective quantities of that good which people are ready to offer for sale at all possible prices.* This definition evidently corresponds precisely to the definition of demand. So close indeed is the analogy between demand and supply that a few words only at this point will suffice to set forth the formal character of the latter. Total supply is the sum of a number of individual supplies. The phrase "ready to offer for sale" implies both ability and willingness. Supply, like demand, relates to a particular place and a particular time; in strictest theory to an instant of time, though for most purposes the time is an appreciable period — a few minutes, an hour, a day, a week, etc., during which the respective quantities would presumably be offered if various assumed prices should prevail. A supply schedule, like a demand schedule, indicates a series of simultaneous possibilities — on the other side of the market however.

Supply and stock. Much discussion of the various problems of price goes astray through confusing supply with the stock of goods on hand. For example what is popularly known as the "visible supply" of various grains in the United States is reported monthly by the United States Census Bureau. These are important quantities, and they have a real bearing upon supply, but they are not the supply of wheat or corn in any market.

The stock of wheat of various grades in the elevators in the neighborhood of Chicago is not the supply of wheat on the Chicago Board of Trade. The supply is the schedule of the

respective numbers of bushels of wheat that people are not only able but willing to offer for sale at various prices at a particular time. No one of these quantities will be the same as the quantity of wheat in the Chicago elevators. In fact, while there is a relation between the quantity of goods on hand and the supply, this relation will often appear not to be very close. The quantity of wheat offered for sale at Chicago at some particular price may be far less than the wheat on hand in the elevators. Much of that wheat the holders may not be willing to sell at that price. On the other hand, the quantity offered might be more than the total stock in the elevators. The Chicago market may be drawing upon wheat stored in other places. In fact the quantity offered might conceivably be greater than the total stock in existence. This surprising state of affairs occurs now and then in speculative markets. On May 9, 1901, on the New York Stock Exchange, it was suddenly discovered that sellers had contracted to deliver more shares of the stock of the Northern Pacific Railway Company than could possibly be bought or borrowed by them. A disastrous crisis was the result. Of course the people were offering shares which they did not then possess; they hoped to acquire them in time to deliver. But that does not alter the fact that so many shares were actually offered. In 1917, during the World War, there occurred a "corner" in wheat on the Chicago Board of Trade. At one time the price was $3.25 a bushel, and it was stated by a well-known authority that "we are getting into a serious situation. Everybody is long, and there is not available the wheat to deliver to those who have bought it." [1]

Relation of quantity to price: Long and short run. As we have seen, the normal relation between price and quantity offered by an individual seller is that the quantity varies directly with the price. It might appear that the combination of the several individual supplies would lead to a similar relationship for the whole market. This is indeed the result

[1] J. A. Patten, "In the Wheat Pit," *Saturday Evening Post*, Nov. 5, 1927, p. 177.

when we are considering the "short run" only — what we have called, in Chapter VIII, the "immediate sales problem." If we are thinking only of the quantities that sellers could be induced to part with at once from their existing stocks, it is true in practically all cases that the higher the price the greater will be the quantity offered by sellers.

But when we turn our attention to the "long run" or "remote sales problem," the matter is not quite so simple. Here we are interested in the quantities that will be produced for sale over a period of time, involving consideration by the sellers of their future costs of production. The shape of the long run total supply curve of any industry will be affected by the cost conditions that prevail in that industry.

Increasing costs. Let us start our inquiry into the law of supply by supposing that at a given time production had all to come from the establishments then in operation. It is clear that under this assumption an increased product could be obtained only by offering a higher price, permitting each of the individual firms to increase its output at the expense of higher marginal cost, which for all producers would be equated with the price. We should then have an upward sloping total supply curve, indicating that the higher the price the greater would be the number of units offered for sale.

Let us next consider another condition, much more common in actual business. Here production is not limited to the existing firms, but if new firms enter, their average costs, when they are operating at the most efficient size, will be higher than the existing price. In other words, the existing price will be below the minimum points on their average cost curves. That is the reason they are not already in the market. But if prices were high enough they would enter and offer for sale such quantities of the product as would equate their marginal costs with the price. Of course each of the firms already in the market would increase its production till its marginal cost and revenue were similarly equated. We then have a situation with an upward sloping supply curve, the number of

competitors and the amount produced by each varying directly with the price.

The supply situation just described, whether it arises from the lack of new firms able to enter the industry or from the fact that new firms can enter only at higher costs than those of the existing firms, may be described as a condition of increasing costs in the industry.

Diminishing returns. This state of increasing costs is generally brought about by scarcity of one of the factors of production which all firms in the industry require. This condition appears most clearly in the case of agricultural production, where land is the scarce factor. After the early stage of abundant fertile land, which has long since been passed in every well-settled community, a large number of comparatively small farms has generally proved to be the organization of maximum efficiency, and the farms normally have their plants fully utilized. Increasing the product of the existing farms means that more laborers must be employed per acre, more capital must be invested in the form of buildings, farm machinery, tools, drainage systems, etc., more must be spent on fertilizers, and so on. All this means an increased cost per unit of product. On the other hand, there is only a limited amount of idle land available, and this land is of inferior quality to that now used; otherwise it would be already in use. New farms, because of the inferiority of the lands, will from the beginning have to be operated at higher costs than prevail on the existing farms. Since either means of obtaining an increase in agricultural products thus involves increased unit cost, both will be resorted to, and the increase will come partly from the cultivation of additional inferior lands and partly from more intensive cultivation of the lands already in use. Similarly in mining the output may be increased only through resorting to inferior mines and through working existing mines more intensively by going to greater depths, taking out lower grade ores, installing more expensive machinery, and in other ways adding to the cost for capital and labor more than in propor-

tion to the gain in output. Similar conditions appear in lumbering, fishing, and all the extractive industries by which the raw products of nature are obtained. All these industries are thus normally operated, except in their remote earliest stages, under what is known as the *law of diminishing returns.*[1]

In this way the scarcity of one or more factors of production can give a full explanation of the fact that an industry's supply curve is upward sloping. Not only will marginal costs be rising in each individual firm, for reasons we have already studied, but it will be impossible for new firms to undercut those already in the field.

Constant costs. A different supply situation is found in certain lines of industry in which, after a certain price has been attained, it is normally easy for new firms to enter the market with costs no higher, or only a shade higher, than the prevailing price. For an industry to find itself in this situation two essential conditions must be satisfied. (1) The firms composing the industry must have similar cost curves, and it must be possible for new firms, if they enter the industry, to produce under substantially the same conditions of cost. By similarity of cost we mean, more specifically, that the minimum points on the average cost curves of the various firms must be at about the same level. Usually in such an industry the plant of most efficient size is comparatively small. If the most efficient sized plant is very large, the industry is almost certain not to be operating under conditions of pure competition; and even if it is, it becomes difficult and hazardous for new enterprises to enter the field. (2) There must not be heavy reliance upon a scarce factor of production. If there is, the entrance of new firms will raise the general level of costs for all producers so that it will be impossible to expand output except under the stimulus of rising prices.

The separate producer as usual looks upon the prevailing price as a determined thing. Each will furnish that amount

[1] For a more elaborate explanation of the principle of diminishing returns, see Chapter XV.

of product which equates his marginal cost to the price. No firm at present operating will increase its output unless a higher price is offered, because under pure competition the supply curve of any individual firm must always be upward sloping. But since we have assumed that new firms may enter the industry with approximately the same costs as those already in existence, a very slight rise in price will induce a large expansion of output through the entrance of new firms. Conversely a very small drop in price will lead to a heavy shrinkage in output. We have assumed that the firms already in the industry have similar cost curves; if this is the case, competition will have driven the price down to a level where it just covers the average cost (including of course normal profits) for each of the firms in the industry. Each enterprise will have an upward sloping individual supply curve, but the output of the industry as a whole can be enlarged without any increase in the general level of costs. In place of a normal ascending supply curve we have, after a brief initial upward slope, a substantially horizontal supply curve (illustrated by the line *ss'* in Figure 11).

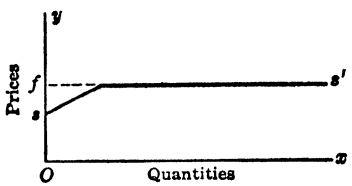


FIG. 11

No supply curve slopes downward. To round out our analysis of the conditions of supply, we must here observe that there can normally be no such thing as a downward sloping supply curve. Barring such exceptional cases as the improvident savages who quit work just as soon as they have received a certain amount in return for their products, how could a lower price possibly induce sellers to offer larger quantities? The relation between price and quantity offered, in a perfectly competitive market, is such that the latter varies directly with the former, thus giving a supply curve that always slopes upward except when it becomes a horizontal line. The rising and horizontal supply curves go respectively with conditions of increasing and constant costs in the industry as a

whole. There may be also, as we shall soon see, a condition of decreasing costs in an industry, but that does not mean that even under such condition there is a downward sloping supply curve.

Elasticity of supply. It is interesting to note that the horizontal supply curve or the condition of constant costs for the industry as a whole represents a state of perfect elasticity of supply. There is variation in the elasticity of the supply of different commodities, analogous to that in the elasticity of demand. If the quantity of a good offered for sale varies a great deal as compared with variations in price, we have an elastic supply. Here the supply curve has a gradual slope. If the quantity offered varies only slightly in response to changes in price, we have an inelastic supply. Here the supply curve has a steep slope. In general, within the range of usual prices, the supply of the products of the extractive industries is inelastic, whereas manufactured articles are elastic in supply, though these rules are by no means without exception.

The law of supply. From the foregoing analysis the following generalization — which we may call *the law of supply* — follows quite obviously. *In a given market at a given time, the quantity of any good which people are ready to offer for sale varies directly with the price.* It should be noted however that at the limit of absolute elasticity of long-run supply a substantial change in price may cause the quantity either to drop to zero or to increase without limit. This law is readily illustrated graphically by upward sloping and horizontal supply curves.

Long run and short run again. Of course the practical interests of business men and of the consumers generally are far more concerned with long-run supply conditions than with the seller's short-run sales problems. And the discussion in this chapter has applied primarily to the long run. If for example we are studying the variations in the output of cotton that occur within a single year in response to variations in price, it is obviously not realistic to assume that large numbers

of new firms will within that period be attracted into the industry by an increasing price or that existing firms will be driven out of production by a decline in price. For short-run considerations we will achieve a much closer approximation to reality by considering the number of enterprises as fixed, giving us a total supply curve based on the individual supply curves of all existing firms. It is only if we are concerned with long-run tendencies that we may legitimately assume that high profits will attract new producers and very low profits will drive existing firms out of business.

Decreasing cost industries. In earlier paragraphs of this chapter we have warned against the fallacy of a downward sloping supply curve. Since a fall in price normally tends to induce existing firms to restrict output and could not conceivably induce new firms to enter an industry, it is quite inconceivable that a fall in price should induce an increase in output except in a very abnormal situation. This does not mean however that there cannot be an industry so constituted that as its output increases the general level of costs will be lower. Specifically if an industry is prosperous so that new firms are attracted into it and output is steadily expanded, it is possible that the very expansion of the industry will have the effect of lowering average and marginal costs for the constituent firms. In such an industry we still could not say that there was a downward sloping supply curve, since a decrease in price would certainly not lead to an increase in output. What we could say however is that if (perhaps through an initial increase in price) output is expanded and costs thereby lowered, the industry could be in equilibrium after the expansion had occurred with a lower price than before.

For example let us consider a condition which is the exact reverse of that which gave rise to increasing costs; that is, a condition in which some element of cost in an industry becomes smaller as the industry expands. Such instances are by no means impossible to find. For instance the rapid expansion of agriculture in the Mississippi Valley eventually reached

the point at which it became practical to use complicated machinery — harvesters, reapers, steam threshers — in place of the simpler tools which were appropriate to small farming. Unit costs were reduced, leading to successive downward shifts of the cost schedules of individual farmers. Under such conditions it is clear that as the output of the industry increases it can afford to sell that output at a lower and lower price.

If now we should undertake to represent this situation graphically, we could draw a downward sloping curve to show for each output the price which would be sufficient to induce the industry to continue producing that output. The reader will of course recognize the fundamental difference between this and a supply curve, which shows for each price the quantity that would be offered for sale.

EXERCISES

1. Assume the following conditions of cost for the production of automobile heaters in the Acme Products plant:

Output	*Fixed cost*	*Variable cost*	*Total cost*	*Average variable cost*	*Average total cost*	*Marginal cost*
10,000	$150,000	$ 75,000	$225,000	$7.50	$22.50	$ 7.50
20,000	150,000	145,000	295,000	7.25	14.75	7.00
30,000	150,000	216,000	366,000	7.20	12.20	7.10
40,000	150,000	288,500	438,500	7.21	10.96	7.25
50,000	150,000	363,500	513,500	7.27	10.27	7.50
60,000	150,000	442,500	592,500	7.38	9.88	7.90
70,000	150,000	527,500	677,500	7.54	9.68	8.50
80,000	150,000	619,000	769,000	7.74	9.61	9.15
90,000	150,000	719,000	869,000	7.99	9.66	10.00
100,000	150,000	832,000	982,000	8.32	9.82	11.30

(*a*) Plot on a sheet of graph paper the curves of average variable cost, average total cost, and marginal cost.

(*b*) Approximately what is the optimum size, or point of maximum efficiency, for this concern?

(*c*) Assume that the product is sold under conditions of pure competition. If the best market price obtainable for the output were $7.10 per unit, would you expect the concern to produce any heaters? If so, how many? Explain.

(*d*) If the best market price obtainable for the output were $8.50 per unit, would you expect the concern to produce any heaters? If so, how many? What would be the loss per unit, and the total loss, sustained? What would be the loss if the plant closed entirely?

(*e*) If a price of $10.00 per unit could be obtained for the output, how many heaters would be produced? What would be the profit per unit, and the total profit, under such conditions? Draw the rectangles which correspond to total cost, total receipts, and total profit, in this case.

2. Assume the following relation between output and marginal cost for the production of potatoes on four different farms:

Marginal cost	*Output (in bu.)*			
	Farm A	*Farm B*	*Farm C*	*Farm D*
$0.20	1,000	400	700	600
.30	2,000	800	1,200	1,200
.50	3,000	1,200	1,700	1,800
.80	4,000	1,600	2,200	2,400
1.20	5,000	2,000	2,700	3,000
1.80	6,000	2,400	3,200	3,500
2.75	7,000	2,800	3,700	4,000

(*a*) Draw up a total supply schedule to include these four potato producers.

(*b*) Construct the supply curve corresponding to this supply schedule.

(*c*) At $1.00 per bushel, what quantity of potatoes would be produced and offered for sale?

3. Describe the essential features of a supply schedule. What factors may cause a change in supply from one time to another? Draw a simple diagram to illustrate an increase in supply; a decrease in supply.

CHAPTER XI

Determination of Price under Pure Competition

Determination of price and quantity exchanged. Equilibrium: Price determined by demand and supply. Demand and supply, hitherto separately examined as representing respectively the buyer's side and the seller's side of the market, may now be brought together. In order to have the assistance of a concrete example, let us assume that we are investigating the price of fresh-laid eggs in the retail market of a small-sized city, which we call T. We assume that on a certain day, October 15, 1939, there were no changes in other prices, but that the price of eggs at the beginning of the day was somewhat uncertain. The demand for eggs and the supply of eggs were as shown in the following schedules. The relations between the quantities and the prices in these schedules conform to established laws of demand and supply. The exact numerical quantities are of course not significant; they are used only for purposes of illustration.

RETAIL EGG MARKET OF T, OCTOBER 15, 1939

Demand		*Supply*	
Prices	Quantities People Were Ready to Buy	Prices	Quantities People Were Ready to Offer
90 cents	160 dozen	90 cents	375 dozen
85 cents	170 dozen	85 cents	360 dozen
80 cents	185 dozen	80 cents	340 dozen
75 cents	200 dozen	75 cents	320 dozen
70 cents	220 dozen	70 cents	300 dozen
65 cents	240 dozen	65 cents	280 dozen
60 cents	260 dozen	60 cents	260 dozen
55 cents	280 dozen	55 cents	230 dozen
50 cents	300 dozen	50 cents	200 dozen
45 cents	320 dozen	45 cents	160 dozen
40 cents	340 dozen	40 cents	120 dozen
35 cents	360 dozen	35 cents	80 dozen
30 cents	380 dozen	30 cents	40 dozen

Now let us plot on one diagram two curves corresponding respectively to these two schedules. The result is shown in Figure 12; *dd* is the demand curve; *ss*, the supply curve.

The curves of demand and supply intersect at the point *p*. The height of this point, *pb*, represents the same price, and its distance from *Oy* the same quantity, on both curves. The point *p*, as a point on the curve *dd*, indicates that at a price of 60 cents people in the city of T were able and willing to buy

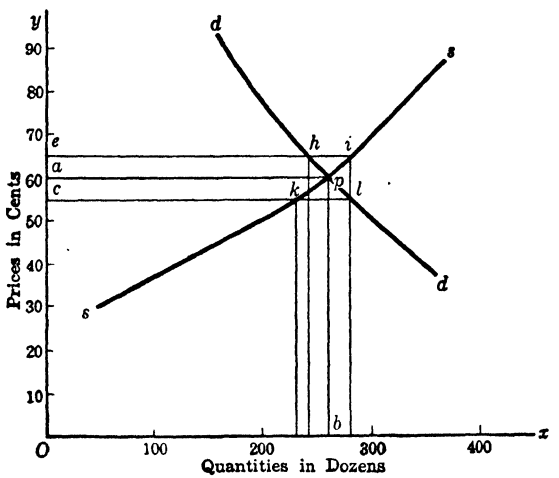


FIG. 12

260 dozen eggs; as a point on the curve *ss*, *p* indicates that at a price of 60 cents other people in the same city were able and willing to sell 260 dozen eggs. At the price of 60 cents quantities demanded and offered were the same.

Since we are assuming a market of pure competition — in which there are many independent buyers and sellers, each familiar with the conditions of demand and supply of eggs on this particular day — this is *the price* of eggs in this market on this day. The price could not be anything else. Any seller who would demand more than this, say 65 cents a dozen, would find nobody willing to buy from him. Rather than be left with his stock unsold on his hands, he would reduce his

price. A seller, on the other hand, who offered to sell eggs for less than 60 cents, say 55 cents, would soon find himself swamped by a rush of eager buyers; he would see that he could easily sell all his stock for more than 55 cents, and he would raise his price. We reach the same conclusion on regarding the matter from the buyer's standpoint. A purchaser who offered to buy any considerable quantity at more than 60 cents a dozen would find all the sellers eager to sell to him and would soon see that he could get all the eggs he wanted at a lower price. He would therefore lower his price. Any buyer, on the other hand, who tried to get eggs for less than 60 cents would find no one willing to sell to him and would have to raise his price or go without. No considerable buying and selling could therefore take place at any price except 60 cents, so long as our curves accurately represent the conditions of demand and supply and buyers and sellers in general are aware of what is going on in the egg market.

Determination of the quantity exchanged. Clearing the market. It will be evident that when demand and supply determine the price they also determine the quantity exchanged. Graphically, the intersection of the curves at *p* (Figure 12) determines the quantity exchanged at 260 dozen, as well as the price at 60 cents. At the price thus determined all sellers of eggs are able to dispose of the entire quantity of eggs which they are willing to sell (at that price) and all buyers are able to get as many eggs as they wish to buy (at that price). No seller in our example who is willing to take 60 cents a dozen will be left with an unsold stock, and no buyer who is willing to pay 60 cents will fail to get what he wants. In technical terms, this price "clears the market." This is not true of any other sum. If, assuming for the moment the impossible, the price were 65 cents, all buyers would obtain all they wanted at that price; *i.e.*, 240 dozen; but sellers, willing to dispose of 280 dozen, could sell only 240 dozen, because buyers would take no more at that price. The difference, 40 dozen (represented by *hi*, the difference between

ei and *eh*), would be left on their hands. On the other hand, consider a price of 55 cents, again assuming the impossible. Sellers would be able to dispose of all they had for sale at that price; *i.e.*, 230 dozen. But buyers, although willing to buy 280 dozen, would be able to get only 230 dozen, since the sellers would part with no more. The difference, 50 dozen (represented by *kl*, the difference between *cl* and *ck*), stands for the unsatisfied requirements of the buyers. Such differences cannot exist at the *price pb*, determined by demand and supply.

The process of adjustment. So far in our investigation of the egg market, buyers and sellers in general have been assumed to be well informed and prompt in action. Such is not always the case. The result is a slowing up of the interaction of the forces of demand and supply. But just as murder will out, so facts regarding the egg market percolate through the community and tend ultimately to have the same effect as they would immediately under more active conditions. On the first day we study the egg market, the price of 65 cents a dozen may be carried over from a previous day; but on October 15 it leaves at least some egg dealers with a surplus on hand which they would have sold at this price if they had had the chance. These sellers may not worry about it because they think it is just an off day, and they hope to make up for it tomorrow. But if other days follow with similar conditions of demand and supply and similar results, someone will eventually react and reduce the price of eggs. His motive is to get rid of his surplus. As soon as any sellers reduce their prices, others, offering the same kind of eggs, begin to reduce their prices also. They may have been lucky enough to sell all they were willing and able at the price of 65 cents; but as soon as competitors reduce their prices, they are in danger of losing their business. They may not indeed lose all of it, but they may lose enough to affect them seriously if they do not meet the reduction in price. Thus if conditions of demand and supply remain the same, the price of eggs will drop toward 60 cents.

On the other hand, on the first day we study this market the price of 50 cents a dozen may have been carried over from a previous day. As a result, some people who want to buy eggs do not get them. This may not have any effect on the price that day, because sellers do not immediately notice the fact. Soon however some seller will notice it and put up the price. In this case he knows that he need not fear the lower price of competitors, because they are not able and willing to sell all that is asked at that price. As soon as one seller is successful in selling at a higher price, others follow suit, seeing the opportunity for additional profit. Thus in this case, if conditions of demand and supply remain the same, the price of eggs will rise toward 60 cents.

The laws of price equilibrium. In previous chapters (Chapters VII and X respectively) we arrived at what we called the law of demand and the law of supply. Observation of the actions examined in the present chapter led men long ago to formulate what are known as *the laws of price equilibrium*, referring to the joint action of forces on both sides of a purely competitive market. These may be stated as follows:

I. *If the quantity of any good which people are ready to buy at the prevailing price is greater than the quantity offered at that price, the price will rise.*

II. *If the quantity of any good which people are ready to buy at the prevailing price is less than the quantity offered at that price, the price will fall.*

III. *At equilibrium, the price of any good and the quantity of the good exchanged are determined by the two forces of demand and supply, at that point at which the quantity which people are ready to buy is equal to the quantity which people are ready to offer for sale.*

The causes of price change: In general. Our excursion into the theory of price has thus far shown us how prices of goods in a market characterized by pure competition are determined, as of a particular time, by demand and supply. Prices in the practical world of today are however seldom stationary but are generally subject to change, more or less

frequent. It is not enough therefore to know how a certain competitive price may be established in a static market. We must also understand the mechanism by which prices rise and fall under the dynamic conditions of modern life.

A glance at the diagram in Figure 12 (page 174) will determine that no other equilibrium price than *pb* is possible with those particular curves of demand and supply, since *p* is

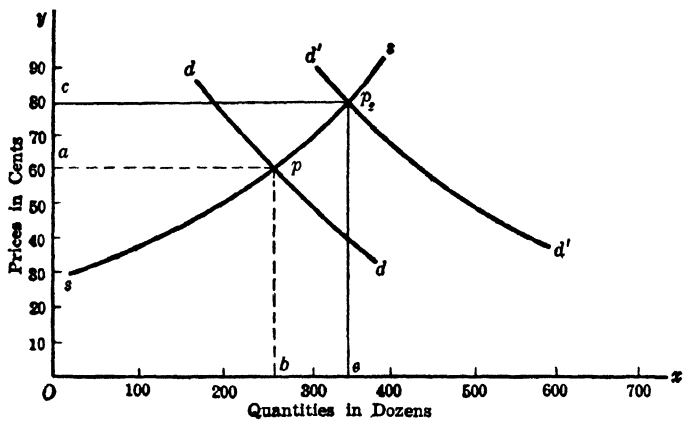


FIG. 13

the only point at which those curves can intersect. A different point of intersection could be obtained only by a shift in one or both of the curves. With the path thus indicated our present inquiry is quite simple. We shall find the graphical aids especially useful in this part of our investigation.

Let us consider an increase in demand. This means that buyers have become more eager. In terms of the definition of demand, it means that people are able and willing to buy greater quantities at all possible prices. Graphically this situation is represented by a new demand curve, each point on which is farther away from the *y* axis than each corresponding point on the old curve; in other words, the demand curve has been shifted to the right.

Returning to our example of the retail egg market of T, let us assume that, as cold weather comes on, the people of T incline to consume more eggs; in other words, the demand increases. Let us suppose that the demand on December 15 would be correctly represented by the curve $d'd'$ in Figure 13. Assuming further for the sake of simplicity that there has been no change in the conditions of supply, the curve ss will serve to represent the supply both on October 15 and on December 15. Compare Figure 12 on page 174. The new demand curve intersects the supply curve at the point p_2, farther to the right and farther up than the point p, which determined the price and the quantity exchanged on the previous date. The price is now 80 cents, and the quantity of eggs exchanged is 340 dozen, an increase in both magnitudes since the earlier date. We conclude then that under conditions such as these an increase in demand tends to cause an increase in price and an increase in the quantity exchanged.

By employing similar reasoning the reader can easily round out this analysis by demonstrating and illustrating these three additional propositions: a decrease in demand tends to cause a decrease in price and in the quantity exchanged; an increase in supply tends to cause a decrease in price and an increase in the quantity exchanged; and a decrease in supply tends to cause an increase in price and a decrease in the quantity exchanged.

The foregoing generalizations have been stated as tendencies, meaning that the results indicated will follow the respective causes unless some other cause interferes; that is, if other things are equal. Thus we have noted the effects of changes of demand or of supply, the other factor remaining unchanged. But it is quite possible and normal that changes in both factors occur at the same time. The hens may lay more eggs, thus increasing the supply, at the same time that the approach of hot weather causes a decline in demand. In such cases the effect is the resultant of the two influences, operating jointly in accordance with the generalizations stated

above. It will be readily apparent that in some cases the two influences reinforce each other while in other cases they neutralize each other. For example simultaneous increases in demand and supply tend to neutralize each other as regards their effect on price. If the increase in demand were the predominant change, then price would tend to rise in spite of the counteracting influence of the increase in supply. If the increase in supply were predominant, the price would tend to fall. Finally the two increases might exactly neutralize each other and so cause no change in price. As to the quantity exchanged, the combined effect of increases in both demand and supply must be an increase, and a greater increase than would have been caused by either force acting alone.

Demand and supply the sole determinants. Price changes are caused by changes in demand and supply, and these are the sole determinants. This may at first seem a sweeping assertion. Is not the price of eggs affected by the laying activity of the hens? Does not the rainfall affect the price of wheat? Are not the prices of many articles affected by changes in fashion? Was not the price of liver raised by general dissemination of knowledge as to its beneficial effects in human diet? And are we not always being told that the "high cost of living" is due to "inflation of the currency"? Such pertinent questions might be multiplied, but their obvious affirmative answers do not imply that here is a host of additional causes affecting competitive price. All these circumstances, and others, do influence prices, but they do so only indirectly. They act only by first affecting either demand or supply. A drought reduces the wheat harvest; the supply is diminished, and the price of wheat goes up. Knowledge of its qualities increased the demand for liver and so raised its price. There is no side door through which any influence can get at price; price may be approached only through one of the specified main entrances, remembering always that we are here concerned only with the prices of goods subject to conditions of pure competition.

The retail egg market. The reader may find it instructive to compare the history of the hypothetical egg market which has served us for illustrative purposes with the records of actual egg markets. The egg market is decidedly seasonal, although the placing of eggs in storage and the withdrawal of eggs from storage in periods of flush and slack production respectively have tended to even out the supply during the year.

The interrelation of these factors of production and price are clearly illustrated by the market record below. These figures relate to a number of markets combined, but they are indicative of what happens in any one market. Receipts of eggs rise during the first of the year until they reach a high

EGG MARKET STATISTICS [1]

Year and month	*Receipts five markets*	*Cold storage stocks end of month*		*Average retail price 51 cities*
	(Thousands of Cases)	Case (Thousands of Cases)	Frozen (Thousands of Pounds)	Cents per Dozen
1934				
January	808	50	49,910	30.1
February	1,165	90	39,181	26.9
March	1,824	1,208	38,679	24.7
April	2,051	4,640	62,632	23.8
May	1,927	7,819	93,947	23.4
June	1,452	8,965	116,058	24.2
July	1,009	8,961	121,564	26.5
August	828	7,938	111,994	31.6
September	665	6,803	99,951	34.8
October	655	4,633	88,715	36.7
November	588	2,380	76,073	39.7
December	642	648	64,879	38.5
1935				
January	750	39	52,726	37.7
February	858	34	39,413	36.7
March	1,488	1,508	39,516	28.6
April	1,866	3,901	59,313	29.6
May	1,963	6,366	84,741	——

[1] Receipts and cold storage stocks from *Survey of Current Business*, Jan., 1935, p. 42, and July, 1935, p. 44; prices are monthly averages of biweekly quotations in various issues of *Monthly Labor Review* (U. S. Bureau of Labor Statistics).

point in April or May. Under the steady influence of increasing supply, egg prices drop off in the spring, reaching a low point at about the same time that receipts of eggs are greatest. Part of the excess production of the spring and early summer is withdrawn from the market for cold storage. Cold storage holdings reach a minimum about the end of February. They accumulate rapidly during April, May, and June, when prices are low, and reach a peak about the end of July. As receipts begin to fall off during the late summer, the price of eggs gradually rises, and storage holdings gradually decline. Prices rise during this period and continue to do so until the end of November. During the winter months the price of eggs gradually declines but stays on a relatively high level as compared with the summer months.[1]

The basic law of price change. The conclusions thus far separately arrived at may now be summarized in the following basic generalizations: *When there is pure competition of sellers,* (1) *the direct effect of a change in magnitude of demand tends to be a like change in price;* (2) *the direct effect of a change in the magnitude of supply tends to be an opposite change in price; and* (3) *the direct effect of a change in the magnitude of either demand or supply tends to be a like change in the quantity exchanged.*

Effect of perfect inelasticity or elasticity of demand or supply. The reader will not have failed to observe that the basic generalizations stated above were derived as direct consequences of the respective slopes of the demand and supply curves. Certain qualifications need to be added to take account of the peculiar conditions which give rise to vertical or horizontal demand or supply curves. Suppose buyers demand a fixed quantity of a given good for whatever price it can be obtained, as shown by the demand curve in Figure 14. Without following through the whole analysis in detail, it should be clear from a glance at the diagram that the vertical position

[1] For another record of the actual egg market, corresponding very closely with our theoretical example, cf. H. B. Vanderbleu, *Economic Principles, A Case Book,* 1927, pp. 110–116.

of the demand curve makes no difference in the conclusions which we have previously reached, save for the one exception that changes in supply alone can obviously not cause any change in the quantity exchanged.

Similarly we conclude that, when the quantity offered is the same irrespective of price (see Figure 15), the general conclusions which we have reached are still valid except that

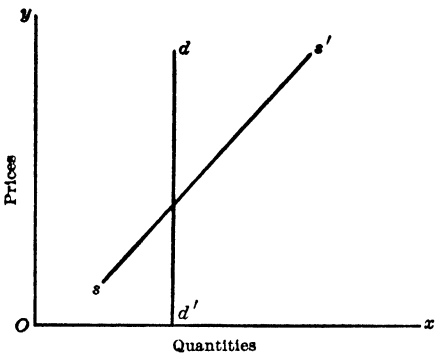


FIG. 14

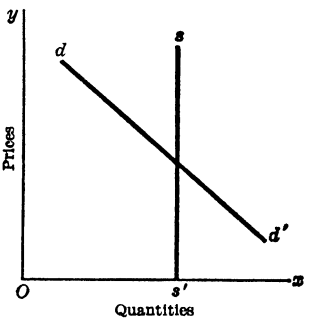


FIG. 15

changes in demand alone can cause no change in the quantity exchanged.

The reader will recognize in Figure 14 an example of a perfectly inelastic demand and in Figure 15, an example of a perfectly inelastic supply. Here the quantity exchanged is determined by demand alone, or by supply alone, as the case may be; but the price in both cases is determined by the combined action of demand and supply.

We must also consider conditions represented by curves of the horizontal type exhibited in Figures 16 and 17 (page 184). Suppose buyers — or more normally a sole buyer — to be ready to buy any amount at a fixed price, as for example, when a local canning factory offers to purchase at a stated price all the wild berries that are brought in. Demand in this case is represented by a curve such as *dd'* in Figure 16. Another possibility is that an industry will have a horizontal

supply curve, which means that sellers are able and willing to offer for sale any amount at a fixed price, as illustrated by the supply curve ss' in Figure 17. Such a supply curve, if interpreted literally, means that nothing will be offered at any lower price. In fact however as we have seen in our discussion of supply, an industry will have a horizontal supply curve only when an increase in output will be provided by the entry of new firms or the reëntry of old firms into the industry, because in any perfectly competitive industry, the supply curves of individual firms must be upward sloping. Therefore a horizontal supply curve really refers only to the long period. In the short run a decrease in price will not mean a complete

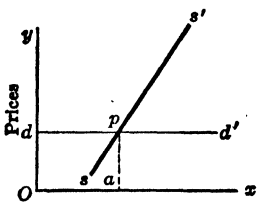


FIG. 16

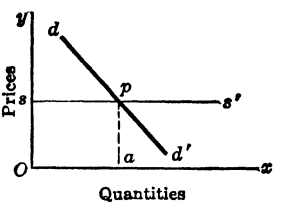


FIG. 17

cessation of production and sales, because a number of existing firms will still be able to cover their variable costs and will continue in production, unless the reduction in price has been drastic.

In these two limiting cases of perfectly elastic demand or supply price is not determined by demand and supply together, but in one case by demand alone, and in the other by supply alone. In both these cases however the quantity exchanged is determined by the interaction of demand and supply.

Decreasing costs. Thus far in this chapter we have studied price determination (always on the assumption of pure competition) under conditions of increasing and constant costs. These conditions permit the use of long-run supply curves either upward sloping or horizontal. There is, as we have

learned, no such thing as a downward-sloping supply curve. Nevertheless, as was developed in Chapter X, there are such things as decreasing cost industries, in which the cost of production for all the firms in the industry will be lowered as output expands and new firms enter. The reader will recollect also that it is possible for us to draw a down sloping curve which shows for any given output the price at which the industry would be in equilibrium. Although this is not a supply curve, it is a convenient way of representing certain significant characteristics of such an industry. For lack of a better name, we may refer to such a curve as a "quasi-supply curve."

Let us suppose that we are studying an industry which is operating under conditions of decreasing costs, as represented

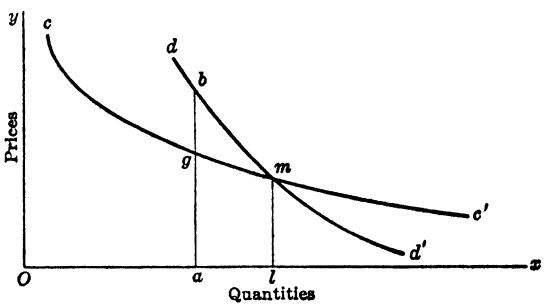


FIG. 18

by the curve cc', in Figure 18. The demand curve is dd'. Our problem is to discover how price will be determined in this situation. As a first approximation, let us assume that the industry is producing the output Oa. Cost conditions are obviously such that, producing this quantity, the industry would be in equilibrium at a price ag. That is to say, if the price were ag and the industry were producing Oa units, the existing firms in the industry would have no incentive to alter their respective outputs, new firms would have no incentive to enter the industry, and existing firms no incentive to leave it. But it is evident that, with demand conditions as indicated, the quantity Oa would command a price of ab. Under our

assumptions then the industry is not in a position of equilibrium. What adjustments are likely to occur?

First of all buyers will tend to bid the price up to *ag*. This will induce an increase of output. Each firm, seeing the price above its own marginal costs of production, will increase its output until its marginal cost becomes equal to the price. Also new firms will be induced to enter the field. The effect of the increased output will be to cause a drop in the price.

If we were dealing with an increasing cost industry, equilibrium would be reached, as we have already learned, at a price somewhere between *ag* and *ab*. But under our present hypothesis of a decreasing cost industry, the increased output of the industry permits the cost curves of all the separate producers to be shifted downward, perhaps because raw material or some other element of cost becomes cheaper, perhaps because the larger scale of operations permits a more efficient organization of the industry. Price therefore may fall even below the initial price *ag* before equilibrium is reached. According to our diagram the industry will be in equilibrium when output has expanded to *Ol* and price fallen to *lm*.

If now we take a fresh start, with an output substantially greater than *Ol*, the reverse reasoning is indicated. Since price is now below the marginal costs of the several producers, each will be induced either to reduce its output or to go out of business. As the industry contracts to a lower level of total production, costs rise; *i.e.*, each firm finds its cost curve higher. Price however rises through the influence of reduced output, and at a faster rate than the rise in unit cost. Equilibrium is reached at quantity *Ol* and price *lm*.

We must not forget that the quasi-supply curve refers only to the long run. In the short run even a decreasing cost industry will have a true upward-sloping supply curve, because in the short run there will not be time for new firms to enter the industry in considerable numbers or for old ones to leave it or even for existing firms to expand plant or alter the scale of their operations. Thus in the short run (as we have had

occasion to note at several points) the supply curve of the industry is based on the upward-sloping supply curves of the firms which compose it. The reactions discussed in the preceding paragraphs refer to developments that take a considerable period of time. If in the above illustration price is initially well above that which would be required to keep the industry in equilibrium, this merely means that profits are being made which will sooner or later induce a substantial increase in output. The reaction would require a considerable period of time; business men would first have to become convinced that the profitable situation was likely to endure, and even then it would require some time before new plant and equipment could be built and new firms organized.

EXERCISES

1. Assume the following demand and supply schedule for fresh eggs, medium size, in a certain city on a certain day:

	Quantity that would be	
Price	*Purchased*	*Offered for sale*
$0.30	1,300 dozen	200 dozen
0.31	1,140 dozen	370 dozen
0.32	1,000 dozen	530 dozen
0.33	880 dozen	670 dozen
0.34	780 dozen	780 dozen
0.35	680 dozen	900 dozen
0.36	600 dozen	1,000 dozen
0.37	530 dozen	1,080 dozen
0.38	475 dozen	1,160 dozen
0.39	410 dozen	1,230 dozen
0.40	350 dozen	1,290 dozen

(*a*) Plot the corresponding demand and supply curves.

(*b*) Draw, on the same diagram, a new demand curve indicating an increased demand and a new supply curve indicating a decreased supply.

(*c*) What is the new price of eggs? Explain the change.

(*d*) What is now the quantity exchanged? Explain the change.

2. (*a*) Plot again the demand and supply curves corresponding to the schedules of Exercise 1.

(*b*) Draw, on the same diagram, new curves indicating an increased demand and an increased supply.

(*c*) What is the new price of eggs? Explain the change.

(*d*) What is now the quantity exchanged?

3. What changes in demand alone or in supply alone or in demand and supply simultaneously would lead to the results indicated below? Draw simple diagrams to illustrate in each case.

(*a*) An increase in price.

(*b*) A decline in price.

(*c*) An increase in quantity exchanged.

(*d*) A decrease in quantity exchanged.

4. What simultaneous changes in demand and supply would fail to cause (*a*) any change in price? (*b*) any change in quantity exchanged?

5. Although the demand for farm wagons has fallen off enormously in the last thirty years, a wagon costs somewhat more today than it did in 1910. What is the explanation? Draw a diagram to indicate what has happened.

6. How would you explain the following price comparisons? Draw simple diagrams to illustrate in each case.

(*a*) The price of cement was the same in 1930 as in 1900.

(*b*) The price of a standard grade of lumber was higher in 1930 than in 1900.

(*c*) The price of gasoline was much lower in 1938 than in 1908.

(*d*) The price of an ice refrigerator was the same in 1938 as in 1900.

(*e*) The price of sugar in 1938 was much lower than in 1860.

7. A manufacturer asks for bids to furnish him 100 machines of a particular type. He is prepared to pay any price up to $150 apiece.

(*a*) Draw a diagram illustrating this situation. What price does he have to pay?

(*b*) Draw supply and demand curves to indicate a condition under which there could be no sale.

CHAPTER XII

Cost of Production and the Seller's Policy under Imperfect Competition

The seller's problem. We must now go back to the fork in the road at which we had arrived at the beginning of Chapter X and, setting our feet on the second of the alternative routes, proceed to investigate the forces that determine prices under conditions of imperfect competition.

As we have learned, the distinguishing mark of pure competition is that the individual seller is unable by his own actions appreciably to influence the market price; or differently stated, the demand for the individual producer's product is absolutely elastic. He can sell all he chooses at the market price and none at any higher price. In contrast, the producer under imperfect competition has freedom to set the price of his product (within limits), since the demand for his product is not perfectly elastic; the demand curve has the normal downward slope.[1]

As we have observed, this is the situation of most industrial producers in the United States; either the competing products have certain individualities which appeal to different customers, or the number of competitors is so small that the individual's selling policy may have an appreciable effect on the prices and the sales of other competitors. The demand for his product does not present itself to him in the form of a going market price at which he must sell or refuse to sell. Instead he faces some sort of demand schedule, which is a way of saying that the number of units he will sell depends upon the price which he sets. He has a price problem.

The producer who is selling in an imperfect market has a position of privilege, but it relates to the seller's side of the

[1] See Chapter IX for definitions and descriptions of pure and imperfect competition.

market only. Since he usually sells in what we have defined as a free market, he may fix either the price at which he will sell or the quantity. In case he takes the first course, demand will determine the maximum quantity he can sell. In the other case, demand will determine the maximum price he can charge. In any case his problem is that of discovering the most profitable price or quantity to set in view of the shape of the demand curve for his individual product and of his various cost schedules, such as we studied in Chapter VIII.

The maximum profit. Suppose the demand for the product of the firm is as represented by the curve *dd'* in Figure 19.

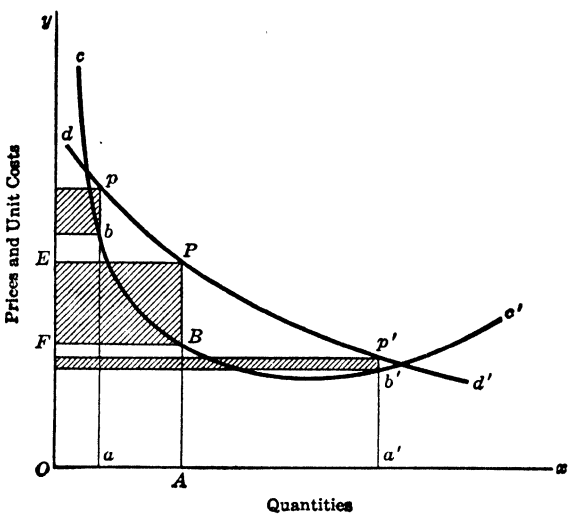


FIG. 19

cc' is the representation of the schedule of average total costs. If this producer charges a very high price, such as *pa*, his sales will be only *Oa*. If he hopes to sell the very large amount *Oa'*, he must be prepared to accept the very low price *p'a'*. In the first case, his total costs are represented by the rectangle with its base on the horizontal axis and its upper right-hand corner at point *b* on the average cost curve, for the area of this rectangle is equal to *Oa* (total quantity) times *ba*

(average cost). Under the same circumstances his gross receipts are represented by the rectangle with right-hand corners at points a and p, this area being equal to Oa (total quantity) times pa (price). His net profits then are represented by the comparatively small shaded rectangle with corners at p and b, for this area is equal to the difference between the gross receipts rectangle and the total cost rectangle.

Sizing up the areas of rectangles with upper right-hand corners at b' and p' respectively, we see that selling the quantity Oa' at price $p'a'$ would again likewise result in comparatively small profit, as represented by the narrow shaded area with right-hand corners at p' and b'.

But now let us consider an intermediate price such as PA. Here the profit rectangle is much greater than in either of the previous cases. If the profit of the firm is to be made a maximum, the problem, translated into graphical terms, is to select a price such that the area of the profit rectangle, $EPBF$, will be as great as possible. In numerical terms, this means finding a price such as $1.00 in the following schedule:

Price	*Sales*	*Gross receipts*	*Total cost*	*Net profit*
$1.20	6,000	$ 7,200	$ 3,000	$4,200
1.10	8,000	8,800	4,000	4,800
1.00	10,000	10,000	5,000	5,000
0.90	12,000	10,800	6,000	4,800
0.80	14,000	11,200	7,000	4,200
0.70	16,000	11,200	8,000	3,200
0.60	18,000	10,800	9,000	1,800
0.50	20,000	10,000	10,000	0

The question now arises: Is there not some general rule or principle which will describe the attainment of maximum profit? It must be recognized that the producer needs to know the demand for his product. Sometimes his knowledge is very vague — a matter of guesswork. But assuming that he has some idea of the demand for his product, or that he can find out through experiment, how can we describe the most profitable price adjustment that he can make? An answer is found

in a very simple relationship between *marginal revenue* and marginal cost.

Marginal revenue. The concept of marginal cost has already been discussed in Chapter VIII; marginal revenue is a complementary idea. Suppose 90,000 pairs of a certain make of shoes can be sold at a price of $4.50, but to sell 100,000 pairs the price has to be reduced to $4.25. The total revenue or gross receipts from the sale of these respective amounts are as follows:

100,000 pairs @ $4.25	$425,000
90,000 pairs @ $4.50	405,000
Difference in revenue	$ 20,000

Rate of difference in revenue:

$20,000 ÷ 10,000 = $2.00

In this example $2.00 is said to be the marginal revenue derived from the sale of 100,000 pairs of shoes, in accordance with the following definition: *Marginal revenue is the amount per unit by which, under given demand conditions, a small change in volume of sales changes total revenue.* The phrase "a small change" implies the smallest change in sales for which, as a practical matter, a corresponding change in price can be calculated or estimated. As our mathematical illustration clearly shows, the marginal revenue is calculated by dividing this "small change in sales" (10,000 units) into the corresponding increase in revenue ($20,000). Marginal revenue is therefore strictly a *rate;* and it is sometimes defined as "the rate at which, under given demand conditions, a small change in volume of sales changes total revenue." Obviously this definition and the one we have chosen are identical in meaning.

The pure mathematical concept of marginal revenue is based upon the assumption of infinitesimal changes in sales, price, and revenue. For purposes of exposition it may sometimes be convenient to consider theoretical changes of one unit. Marginal revenue is then thought of as the amount added to total revenue by the sale of one additional article. Of course in practice this result can seldom if ever be meas-

ured. To complete the concept of marginal revenue, we must add that if the small change in sales takes place through addition, the change per unit in revenue is said to be the marginal revenue of the quantity which includes the added units.

If a small change in sales can be effected without a change in price, and without change in the general conditions of demand, price and marginal revenue are the same. For example, if 100,000 pairs of shoes can be sold at the same price as 90,000 pairs, we may have:

100,000 pairs @ $4.50	$450,000
90,000 pairs @ $4.50	405,000
Difference in revenue	$ 45,000

Rate of difference in revenue:

$45,000 ÷ 10,000 = $4.50

The price is also $4.50

Since there is no change in price, the sale of each additional pair brings in an additional $4.50, and the rate of change in revenue is $4.50 throughout. This is obviously the situation under pure competition; marginal revenue is always equal to the price, since any change in the individual's offerings has no appreciable effect upon the price.

In the case of a producer selling in an imperfectly competitive market however a change in sales has to be accompanied by a change in price. Thus the change in total revenue is made up of two component parts; it will be increased by the revenue from the sale of the additional units, but it will be decreased as a result of the reduction in price of the number of units which was previously being sold.

Marginal revenue and marginal cost: In general. Price above average cost. Whenever a producer discovers that his marginal revenue is greater than his marginal cost, it is clear that he will gain by an increase in production, because such an increase will add more to his revenue than to his cost. As a result of this adjustment, marginal revenue itself will decline and marginal cost may rise so that a further adjustment will presumably add less to his profits and may eventually

even reduce them. But small adjustments which involve a cut in price and consequent increase in sales will always offer the possibility of additional profits so long as marginal revenue exceeds marginal cost. In graphical terms there is the possibility of increasing the area of the profit rectangle (see Figure 19, page 190).

When marginal revenue is equal to marginal cost, so that any further increase in production would increase cost as much as it would increase receipts, additional sales will no longer increase the producer's profit. By the same reasoning, if marginal revenue is less than marginal cost, an increase in production will decrease profits, and total profits can be increased only if production is reduced. In this situation it is to the producer's interest to raise his price (if he sells in a market where price is set by the seller) or directly to limit production. Thus we reach the conclusion that *profits are at a maximum when marginal revenue and marginal cost are equal.*

For the sake of a simple numerical illustration let us assume that marginal revenue is based upon changes of 2,000 units in sales and that marginal cost will increase gradually with increased product, as in the schedule below.

Price	*Sales*	*Gross receipts*	*Marginal revenue*	*Marginal cost*
$1.20	6,000	$ 7,200	—	—
1.10	8,000	8,800	$0.80	$0.50
1.00	10,000	10,000	0.60	0.51
0.90	12,000	10,800	0.40	0.52
0.80	14,000	11,200	0.20	0.53
0.70	16,000	11,200	0.00	0.54
0.60	18,000	10,800	−0.20	0.55
0.50	20,000	10,000	−0.40	0.56

Suppose the producer considers selling 6,000 units at price $1.20. It pays him to increase the amount to 8,000 units even though the price has to be reduced to $1.10. His gross receipts with this reduction in price increase $1,600; *i.e.*, at the rate of $0.80 for each additional unit. This more than covers the marginal cost of $0.50. For similar reasons it pays him to

increase production from 8,000 to 10,000 units. In this instance marginal revenue is $0.60 as compared with marginal cost of $0.51. It does not pay him however to reduce the price to $0.90 and sell 12,000 units. Here marginal revenue is only $0.40.

If price were reduced to this figure, the final units sold would be adding $0.12 less to revenue than to cost. The producer would still be making a profit to be sure, but not so much as he might. On the other hand, selling 10,000 units at price $1.00, he has a marginal revenue a little greater than marginal cost. If he were disposed to alter his sales by smaller jumps, he would probably find the ideal output around 11,000 units. With finer calculations of marginal items, he would be able to find an exact equality between marginal revenues and marginal costs somewhere in this neighborhood.

Price below average cost. In the examples we have so far discussed under imperfect competition, we have assumed that it will be possible for the firm to earn a profit over and above its total costs for a very considerable range of outputs. In order therefore to round out our consideration of the behavior of producers under imperfect competition, we must take account of the possibility that they will not always be so favorably situated. Suppose for instance that the demand for the product is such that there is no amount of product for which the average cost would not exceed the price at which such quantity could be sold. In graphical terms the demand curve lies below the average total cost curve throughout its whole length. Although the firm might lose more money for some outputs than for others, it could not under any circumstances cover its costs.

A moment's consideration will make it clear that, although the firm in this situation has no possibility of maximizing profits, its losses may be minimized by equating marginal cost and marginal revenue. If the firm reduces its output below the point at which marginal revenue and marginal cost are equal, then it will reduce its revenues by more than its total

costs; consequently its losses will be increased. If it expands output into the region where marginal cost exceeds marginal revenue, then more will be added to costs than to receipts and again losses will be increased. We may safely conclude then that so long as the firm remains in business it will tend to produce the output at which marginal cost is equal to marginal revenue so as to minimize its losses.

As to the rest, the situation is similar to that under pure competition. A firm operating under imperfect competition may remain in business for a considerable period of time operating at a loss if variable costs are covered; that is to say, if it is earning something, however small, toward meeting its fixed costs. But it cannot thus continue indefinitely.

Conclusion. We may now sum up the results of our investigation of the relationship between cost and the price and output policies of a firm in three propositions, which are analogous to those which emerged from our study of pure competition. (1) So long as a firm continues in operation under imperfect competition it will tend to produce that output by which marginal cost is just equal to marginal revenue. (2) The firm will not continue production in the long run unless it receives a price that is sufficient to cover average total costs. (3) It may continue operations in the short run so long as price exceeds average variable costs.

Influence of elasticity of demand. We have previously learned that under conditions of pure competition, each producer tends so to regulate the amount of his product as to make his marginal cost equal to the price. In other words, no producer is normally charging a price that is greater than his marginal cost of production. Under imperfect competition, on the other hand, each producer ordinarily equates his marginal cost with his marginal revenue. But under imperfect competition marginal revenue is always less than the price; this is because an increase in product can be sold only by a reduction of price, and the added revenue from the additional units sold at the new price is reduced because of the lower

price at which the number of units formerly marketed must now be sold; the addition per unit to total revenue is thus less than the price. It follows that under imperfect competition producers are normally charging prices that are greater than their marginal costs of production.

How much above marginal cost the price will be depends chiefly upon the degree of elasticity of demand for the particular producer's product. If demand is very inelastic, it will require a large reduction of price to secure any given increase of sales, and marginal revenue for any particular output will be far below the price. In different words, the gain from the added units sold will be small compared with the loss due to the heavy decline in the price received for the units that could have been sold without lowering the price. The producer has comparatively little inducement to lower price and is led to seek the maximum profit by a price far above his marginal cost. On the other hand, if demand is elastic, a small decline in price will add largely to sales and lead the producer to seek his maximum profit by large sales at a price only a little above his marginal cost. The more elastic the demand, the less is the difference between price and marginal cost, and the less is the difference between imperfect competition and pure competition, under which there is no difference between price and marginal cost.

Ascertaining the facts of demand. Surely no further demonstration is needed of the vital interest of the ordinary business man in the demand of consumers for the product he makes or sells. Owing to the general prevalence of imperfect competition, the salesman occupies the key position in modern business, and a knowledge of the nature of the demand for his goods is a prime qualification of salesmanship. The salesman is aware that, other things being equal, the higher his prices, the greater the income and the profits of his business; but his knowledge of the law of demand tells him that other things are not equal, that a high price may defeat its end by reducing sales, while the true road to profit may be large sales

stimulated by low prices. The business man would like to know the precise form of the demand curve of his product. What will be the response of the buying public to his price policy? Are people already using about as much as they want, or is there possibility of bringing in a crowd of new users by the announcement of low prices? Is any large part of his present custom likely to drop his product if the price is raised, or may he count upon a condition of inelastic demand?

It is of even more importance to the business man to know the absolute demand for his product; within any given range of prices are people ready to take large or small quantities? The fate of his business may very well turn upon his knowledge or ignorance of the amount of his goods which buyers would take at various prices. And it should be noted that the enterpriser wants knowledge of the state of demand, not only as it is at the present moment, but as it will be in the future, both immediate and more remote.

The task of discovering the actual facts of demand is anything but easy; it is in truth perhaps the most difficult part of the technique of modern business. In spite of the difficulty however, the importance of the knowledge is such that vigorous efforts are made to obtain facts throwing light on the conditions of demand. Demand is a resultant of two factors, desire and the purchasing power to make desire effective. The attempts of business men to measure and foretell demand revolve about the analysis and interpretation of business statistics which bear upon these two factors.

As indicating the desires of consumers, the business man avails himself of the statistics of consumption of various products, showing both amounts consumed at particular times and trends. For example, the per capita consumption of beef declined from 67.8 pounds in 1900 to 58.0 pounds in 1927, while the consumption of pork was increasing from 64.7 pounds to 68.5 pounds per capita. Ability to buy is indicated by regularly published statistics of incomes, showing the sum

paid in any year for salaries, wages, and pensions, dividends, interest, rents and royalties, etc., and the corresponding indexes showing trends.

A longer range indication of future demand is provided by study of the flow of goods from producers to consumers. The flow from manufacturers, farmers, etc., to the wholesale and retail markets is indicated by statistics of car loadings, exports and imports, and wholesale trade. The flow of goods from these markets to the consumer is shown by such measures as department store sales, life insurance, and advertising. Going still further back toward the sources of income, the business man studies the records of production, such as the statistics of crop production, receipts of goods at export ports, building construction, etc.

These data are supplemented by indexes of general banking and credit conditions, stock exchange transactions, postal receipts, electric power production, business failures, velocity of bank deposits, new incorporations. All these facts contribute to the picture of general economic conditions and so help business men to form their estimates of the future demand for their products.

EXERCISES

1. Assume the following conditions of cost for the production of automobile heaters in the Acme Products plant: [1]

Output	*Fixed cost*	*Variable cost*	*Total cost*	*Average variable cost*	*Average total cost*	*Marginal cost*
10,000	$150,000	$ 75,000	$225,000	$7.50	$22.50	$ 7.50
20,000	150,000	145,000	295,000	7.25	14.75	7.00
30,000	150,000	216,000	366,000	7.20	12.20	7.10
40,000	150,000	288,500	438,500	7.21	10.96	7.25
50,000	150,000	363,500	513,500	7.27	10.27	7.50
60,000	150,000	442,500	592,500	7.38	9.88	7.90
70,000	150,000	527,500	677,500	7.54	9.68	8.50
80,000	150,000	619,000	769,000	7.74	9.61	9.15
90,000	150,000	719,000	869,000	7.99	9.66	10.00
100,000	150,000	832,000	982,000	8.32	9.82	11.30

[1] The student will note that these are the same conditions as were assumed as the basis of Exercise 1 at the close of Chapter X.

(*a*) Plot on a sheet of graph paper the curves of average variable cost, average total cost, and marginal cost.

(*b*) Approximately what is the optimum size, or point of maximum efficiency, for this concern?

2. Assume that the product of this concern is not homogeneous with that of its competitors, but is distinguished through branding or special features of construction. The table below represents the demand schedule for Acme heaters, sold under conditions of imperfect competition:

Prices	*Quantities*
$15.00	10,000
14.00	20,000
13.00	30,000
12.00	40,000
11.00	50,000
10.00	60,000
9.00	70,000
8.00	80,000
7.00	90,000
6.00	100,000

(*a*) For each price and amount of sales in the demand schedule, compute the total receipts that would be obtained from selling that amount at the corresponding price.

(*b*) By reference to the figures of total cost in Exercise 1, compute the net profit or loss at each possible price and output.

(*c*) At what price and output are the profits of the concern maximized?

3. (*a*) From the demand schedule in Exercise 2, compute the figures for marginal revenue.

(*b*) Plot on a sheet of graph paper the demand curve, and the marginal revenue curve.

(*c*) From the cost schedule in Exercise 1, plot again the curves of average total cost and marginal cost (four curves on the same graph).

(*d*) What is the best output for the concern to produce? Reading from the chart, at what price can this output be sold? What is the average total cost of producing this output?

(*e*) What is the profit per unit and the total profit under these conditions? Draw the rectangle which corresponds to total profit.

4. Assume now a different demand schedule for Acme heaters, as indicated below:

Prices	*Quantities*
$11.50	10,000
11.00	20,000
10.50	30,000

Prices	*Quantities*
$10.00	40,000
9.50	50,000
9.00	60,000
8.50	70,000
8.00	80,000
7.50	90,000
7.00	100,000

(*a*) For each price and amount of sales in the demand schedule compute the total receipts, and the marginal revenue, that would be obtained.

(*b*) Plot on a sheet of graph paper the demand curve, and the marginal revenue curve.

(*c*) From the cost schedule in Exercise 1, plot again the curves of average total cost and marginal cost (four curves on the same graph).

(*d*) What is the best output for the concern to produce? Reading from the chart, at what price can this output be sold? What is the average total cost of producing this output?

(*e*) What is the loss per unit, and the total loss under these conditions? Draw the rectangle which corresponds to total loss. What would be the loss if the plant closed entirely?

CHAPTER XIII

Determination of Price under Imperfect Competition

The mechanism of price determination. In the immediately preceding chapter we studied the way in which the individual producer, selling in an imperfectly competitive market, adapts his policy to a given condition of demand for his product, as represented by a particular demand curve. Given the demand, his policy is governed by his cost conditions. This analysis corresponds logically with Chapter X, in which we made similar study of the way in which the individual producer under pure competition adapts his policy to a given market price, governed likewise by his cost conditions.

When now we proceed to the next step and, abandoning the assumption of a fixed demand, inquire into the combined forces which determine price in a market where competition is imperfect, it may seem at first sight as though conditions of demand and supply exerted no very obvious control over the actions of individual sellers. What we see in a large department store or in the market for automobiles or in any other market where products are differentiated and sellers are few is that the sellers themselves set prices. In the first instance, at any rate, the suit which you buy in a store costs fifty dollars because the owner of the store chooses to place that price upon it. In other words, imperfectly competitive markets are usually those in which prices are named by the sellers, though this is not always the case, and the laws which govern price in such a market are exactly the same whether or not the seller sets the prices.

Our study of the forces that determine price under conditions of pure competition was made comparatively simple by the possibility of representing the behavior of buyers by a demand curve and the behavior of sellers by a supply curve.

Putting these two curves together, we could then show how the actions of all the individuals composing the market determine the price. Under imperfect competition, on the contrary, the forces on the two sides of the market cannot be so simply represented. Instead of one demand (schedule or curve) representing the buyers interested in a certain product offered by many sellers, we have a separate demand for the product of each seller. Nor can we employ the concept of supply to represent the behavior of the sellers.

Supply in fact has no meaning except when the amount of a commodity that will be offered in the market depends exclusively upon the market price; *i.e.*, the quantity offered is a function of price alone. This is a condition which is to be found only when competition is pure. In any imperfect market the amount that a given firm will decide to produce and sell will not depend simply on the price at which it can dispose of its output. It will depend also upon such other factors as the selling costs which the firm must incur in order to sell a given output at a given price and upon the possibility of selling a smaller output at a higher price or a larger output at a slightly lower price. Clearly under these conditions it would be quite impossible to draw up a supply schedule showing the various amounts that will be offered for sale at different prices. The possibility of selling a certain output at a certain price will not induce individual firms to establish that price and sell the output which the market will absorb if it is also possible for them to sell practically as large an output at a much higher price or to sell a much larger output at a slightly lower price. In fact it is impossible to speak without qualification of a certain "market price" as the sole force inducing producers to "offer" a certain output if they are selling in an imperfectly competitive market. For in the typical market of this sort the producers themselves set the prices of their products, and in so doing they will have to take account of the whole demand curves for their own products rather than an established market price. Thus in such a market the

notion of a supply curve has no meaning. Whereas the supply of wheat is an intelligible and useful concept, there is, in the strict sense in which we define the term, no such thing as the "supply" of cheap automobiles.

In spite of all this it remains true that prices in an imperfectly competitive market, no less than those in a market where there is pure competition, are determined by the joint actions of buyers and sellers. We shall only have to be prepared to find the relationships somewhat more complex and less capable of statement in simple and precise terms.

Interaction of sellers' policies. Our first step is to recognize and formulate two kinds of consequences which may follow action by a producer. In the first place, having some knowledge of the demand for his particular product, the producer knows that the amount he can sell will depend on the price he fixes or the price which he can get will depend on the quantity he puts on the market. These relationships were investigated in Chapter XII, using the assumption of a given demand.

Now we have to take account of the further fact that under imperfect competition the individual seller cannot take the demand for his product as given; *i.e.*, as independent of his own price-setting acts. In fact the business man must always take account of a factor which does not appear in a demand curve; namely, the probable behavior of competitors. Especially is this so if the demand for his own product is highly elastic, so that a small cut in his price would (other things being equal) largely expand his sales. For then that same action would be bound to do considerable injury to his competitors and provoke retaliation from them. Such retaliation, in the form of price cuts by competitors, would reduce the demand for his own product. It is common knowledge that the most difficult problems which a business man must solve in determining his price and sales policy concern the changes in the demand for his product to which his own actions may indirectly give rise by causing retaliatory action on the part

of competitors. In general any one producer feels the impact of the activities of his competitors, because their policies determine in considerable measure the demand for his own product. It is impossible to analyze the functioning of an imperfectly competitive market unless we take account of the fact that sellers base their decisions, not only upon estimates of the demands for their products, but also on some knowledge of their competitors' behavior, in other words, upon estimates of the indirect effects of their own actions.

The rôle of substitutes: Their significance. We have earlier noted that a consumer's demand for any one commodity depends not only upon his tastes and income but also on the prices of other commodities and the quantities of them which he already possesses. Although a demand schedule or curve always shows the relation between the price of a product and the quantity of that product which people are willing to buy, this does not imply that the quantity of a product people are willing to buy depends exclusively upon the price of that one product. Indeed the reader will recollect that it is impossible to draw up a demand schedule for any one commodity except upon the assumption that the prices of other commodities are given. From a demand schedule it is possible to discover at any given time how much more of the product people would buy if its price were, let us say, ten per cent lower, but only if it can be assumed that other prices remain unchanged.

Our present inquiry is from the seller's viewpoint, and by substitutes we mean things which buyers might use in place of the products of a particular seller. At the one extreme it must be recognized that almost all commodities are substitutes for one another, since they are all competing with one another for the consumer's income. If he decides to spend more on his house, he will have less to spend on clothes and a car. Obviously however it is the closer substitutes that most directly affect the conditions of demand for any single product. These close substitutes may be other brands or makes of a

similar product, as in the example of the automobile industry which we have frequently cited, or they may be physically dissimilar commodities which nevertheless render a similar service, such as gas and coal. What is significant for an analysis of price is the closeness of the substitute in this economic sense.

The limit of closeness of substitutes is reached when the products of the individual sellers are identical. In the market for grade A wheat, for example, each seller is aware that each of his competitors is offering a product which is a perfect substitute for his own. This the reader will recognize as one of the conditions of pure competition; it is in consequence of this and the large number of sellers that, under pure competition, each seller is confronted with a perfectly elastic demand for his own product, represented by the horizontal demand curve.

When the closeness of substitutes is less than complete — *i.e.*, when the products of the individual sellers are not identical — we enter the realm of imperfect competition. The manufacturers of the three principal competing cars in the low-price field, Ford, Chevrolet, and Plymouth, are selling products which, however closely they may approach each other, are still not identical. Each producer has a demand for his own product which is represented not by a horizontal line but by the normal down-sloping demand curve.

Influence of substitutes on size of demand. Let us consider the example of the competition between several grocery stores in a neighborhood. Each store will have certain regular customers attached to it by preference or habit, and it will also have an advantage over its competitors so far as the people who live in its immediate neighborhood are concerned. For the people within one block of grocery Store A but five blocks away from grocery Store B, Store A will be the more convenient. Because of these consumer preferences, the stores are selling in an imperfect market, that is to say, Store A could charge a slightly higher price for a certain grade of coffee, let us say, than Store B without losing all its customers and

could cut its charge slightly below those of Store B without depriving the latter of all its coffee buyers. In a sense, even though the actual physical goods which the several stores sell may be the same, they are to the buyers economically different goods because they are not perfect substitutes for one another.

Now let us suppose that Store B sharply cuts the price of its coffee. Clearly B's action will have the effect of attracting a number of A's customers. Some people who live closer to A will now be induced to buy their coffee from B because of his lower price, even though it will be slightly less convenient for them to do so. The demand curve for A's coffee will be shifted downward and to the left; that is to say, A will sell less at any given level of prices after B's action than he would have before. He will be compelled either to reduce his own price to meet that of B or else to sell less coffee.

These effects are readily illustrated graphically. Figure 20 relates to Store B; Figure 21, to Store A. Store B reduces his

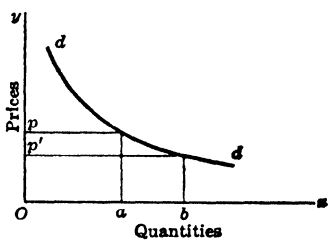


FIG. 20

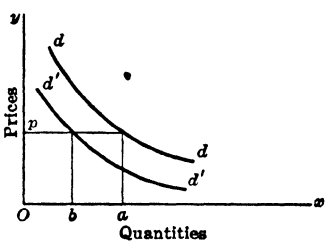


FIG. 21

price from *Op* to *Op'*. This, in accordance with the shape of the demand curve for his coffee, induces new customers to come to him. These customers come from Store A, thus reducing the demand for his coffee, as shown by a new demand curve *d'd'*. The shifting customers increase B's sales from *Oa* to *Ob*, and the reduced demand for A's coffee reduces his sales, assuming he keeps to his price, from *Oa* to *Ob*.

The appearance of a new competitor in the neighborhood will affect Store A in the same way as if B had cut prices. Suppose a new store, C, is established only three blocks away

from Store A. Store C will now be the most convenient for a considerable group of people who formerly traded with either A or B, and in time Store C will build up a clientele of customers who regularly deal with it. Clearly the positions of the demand curves for A's coffee and B's coffee respectively will be altered. Both will be shifted downward and to the left. We may generalize and say that, from the viewpoint of any seller, a reduction in the price of available substitutes for any given product or an increase in the number of substitutes will lower the demand for his own product.

Influence of substitutes on elasticity of demand. To illustrate this influence of substitutes, let us assume two other grocery stores, both selling coffee, but A sells a superior quality to generally well-to-do customers, while B sells a cheaper grade and draws upon a poorer class of the community. It would require a considerable cut in B's prices to draw any great number of customers from A, and a considerable increase in his prices to drive many customers to A. The demand for B's coffee is therefore quite inelastic. Store A also faces an inelastic demand, for similar reasons. Now suppose a new store, C, enters the neighborhood and sells low-grade coffee similar to what B has been selling. The new store may have little effect on the trade at Store A. But Store B will immediately feel the competition as a new limit on his power to raise prices without losing trade. After the new store becomes established, a small cut in price by either B or C will attract many customers from the other. Our conclusion is that an increase in the similarity or "closeness" of the nearest substitute will increase the elasticity of demand for the product of an individual producer, and vice versa.

Price determination in the short run: Price cutting. What distinguishes the short-run problem is the fact that each competitor, in shaping his price policy, assumes that the number of other firms and the scale of their operations may be treated as fixed. Let us assume again that we are considering two firms which produce close substitutes and sell

them in the same market, such as the two grocery stores of the previous example. Suppose that when we first turn our attention to these two firms, A has succeeded in adopting a price which just equates marginal cost with marginal revenue. He has therefore no incentive either to raise or to lower his price. But let us suppose that B has not succeeded in adopting so successful a price policy; he finally arrives at the conclusion that if he were to cut his price five per cent he would expand his sales substantially and increase his revenue more than his costs. In other words, he discovers that his marginal revenue is still substantially in excess of his marginal cost. Presumably when he has come to this decision he will lower his price, let us say five per cent, in the hope of thereby increasing his profits.

Whatever the wisdom of the price cut from B's point of view, it immediately has certain repercussions upon A. As B reduces the price of his product, he will begin to cut into A's trade, and the demand curve for A's product will shift downward and to the left. A now finds himself out of adjustment. So long as the price of his product remains unchanged, the decline in demand will sharply reduce his sales. With this smaller output A's marginal costs will be lower, while his marginal revenue is likely to be the same or higher. So it is very likely to be profitable for him to meet at least some part of the decline in demand with a price reduction. Especially will this be true if the elasticity of demand for each producer's product is high, since in that case B's price reduction will attract a large number of A's customers. When A cuts his price, he will in turn have upset B's original calculations, because after A's price cut B's demand curve will have been shifted downward. By precisely the same reasoning we conclude that B may then have an incentive to cut once again. For if his initial price cut would have enabled him just to equate marginal revenue and marginal cost, and if A's action has now reduced the demand for B's product, B will presumably find marginal revenue still in excess of marginal cost.

This process of competitive price cutting is familiar enough to business men, and once started it may continue for a considerable time.

Limitations. But there are certain limits to such a process of competitive price cutting. In the first place, the cutting of prices of a certain article by all the competing sellers will tend to induce people to buy more of this product, with money withdrawn from the purchase of other goods. For example if competition of the producers of radios should materially reduce their prices, total sales of radios would tend to increase, at the expense, let us say, of electric ice boxes, household furniture, etc. The producers of radios would thus find their sales expanding as the process of price cutting went on. In consequence their marginal costs would rise, and each producer would be approaching more and more closely a position where marginal revenue was equal to marginal cost. The process of price cutting will thus be limited as the firms approach a condition where no one of them has any incentive to carry the matter further. The final situation is one in which there is no possibility of further price cutting, because no firm, even disregarding the actions of its competitors, has any incentive to lower its price.

However it is highly probable that some producers at least will give thought to the retaliation of their competitors and will therefore not carry this price cutting to the bitter end. The fall of prices is thus halted before the limit described in the previous paragraph is reached. If, on the other hand, the limit has been reached, there is likely to be a reaction. Some producer will advance his price, expecting his competitors to follow his lead. This they are likely to do, since the price increase by the first producer has the effect of increasing the demand for the goods of his competitors. Thus the prices may react to a level somewhat above the limit at which all producers find their marginal costs and marginal revenues equal.

Finally there is always the possibility of agreement among

the competitors to halt price cutting. In an industry where competitors are few and where the marginal cost curve throughout a large part of its length lies well below the average cost curve, there is constant danger of what is known as "cutthroat competition," under which the lower limit to price cutting may not be reached although all of the firms in the industry are suffering losses. This will be especially true if the demand for the output of the industry as a whole—that is, for the whole series of substitutes produced by the various firms involved—is highly inelastic. Under these circumstances the producers are very likely through tacit agreement or understanding to strive to avoid price cutting. Each realizes that, if price cutting once starts, the process may be difficult to stop. Although marginal revenue may be well in excess of marginal cost for everyone in the industry, there will thus be a strongly developed "public opinion" (perhaps fostered by a trade association) to prevent the initiation of a price war.

But though an uneasy truce may be maintained in this fashion, it is by no means accurate to call the position one of equilibrium, and indeed it is bound to be highly unstable. For until the limit at which every firm has equated marginal revenue with marginal cost has been approached, at least a few of the competitors in the industry would profit by a price cut and an expansion of output if only they could be sure that other firms would not follow suit. A change in the policy of one firm or the impact of a depression is all that is needed to start the price cutting process.

Conclusions. Where competitors are comparatively few and where they sell individualized products each one of which is in some degree a substitute for the others, their price behavior will tend to have its well-defined characteristics. (1) Prices change infrequently, because of the general dislike and fear of price cutting. (2) Prices tend to be stabilized somewhat above the lower limit to which price cutting could carry them; *i.e.*, the point at which prices have been cut until every firm has equated marginal revenue with marginal cost.

(3) There will be an inherent instability in the price structure of an industry unless prices have been stabilized near this lower limit.

The relative importance of these conclusions will depend upon the type of imperfectly competitive industry under consideration. In the cement and steel industries, where competitors are few but the products sold by the different producers are very close or even perfect substitutes for one another, the probable actions of other sellers will have great importance. A very small cut in price by one cement company virtually compels the others to follow. Therefore no intelligent executive would cut the price of cement without taking account of the inevitable reaction of his competitors. Conversely in such fields of business as retail trade, competition is imperfect mainly by reason of the individuality of products and the element of consumer preference for a particular seller, and not by reason of the fewness of competitors. Even though the number of sellers is large, the demand for an individual seller's product may be so inelastic that he can wisely disregard the reactions of competitors when considering price changes. For instance a price cut by one grocery store in a neighborhood may not have a great effect on the demand of the corresponding product of any one other grocery, because the store which has cut its price will draw only a few customers from each of the other stores in the district. Here a moderate price cut might provoke little or no retaliation. The business man in this type of industry is not likely to base his actions to so great an extent upon estimates of what his competitors will do.

Price determination in the long run. Our previous conclusion that prices would not be stable if they were far above the level which would equate marginal cost and marginal revenue does not mean that established firms will not be able to earn very substantial profits in the short run. If at any given time high profits thus prevail in an industry, the mere fact of imperfect competition does not mean that any par-

ticular firm is free from outside competition. Unless the industry of which it is a part is highly organized to prevent the entrance of new competitors, or unless the profits that the firm is earning are abnormal and not expected to endure, new competitors will be attracted into this line of business by the high level of profits. Since we are studying an imperfect market, it is probable that our producer's product is individualized and protected by brand name or patent or trade mark, so that he will not face the competition of other producers turning out products which will be perfect substitutes for his own. Nevertheless new competitors will appear and will offer to the public products which presumably will be close substitutes. Some of them may also offer closer substitutes than have hitherto been available. It is probable therefore, in accordance with the two propositions which we derived from our study of the significance of substitutes, that the demand curve for our firm's product will shift downward and to the left and will at the same time become more elastic. As this process continues the firm's profits will diminish. Though striving always to maximize its profits, its best efforts will be unable to prevent this decline.

A position may finally be reached at which the firm in question, by producing and selling the output at which its marginal cost is equal to its marginal revenue, can just cover its average costs. By the time the firm we are studying has reached this position, it is probable that the general level of profits earned by firms producing substitutes will also have been forced down to a very low level. There will therefore be no incentive for new firms to put still other substitutes on the market. In this particular sector an equilibrium will have been reached, similar to the equilibrium achieved in an industry under pure competition when the entrance of new competitors has driven the market price down to the point where it just covers the average costs of the various firms in the industry.

This discussion shows that even where competition is im-

perfect, in the sense that there is a down-sloping instead of a horizontal demand curve for the product of each individual producer, the established producers in a field will not be safeguarded from competition. They will, it is true, have greater freedom to determine at any particular time what prices to charge. But if they earn substantial profits, the profits will attract new competitors into the industry. The appearance of new firms will lower the demand for the products of those already in the field, force them to push their sales more intensively, and ultimately make it impossible for them to continue to earn very high profits, just as surely as if they were selling in a market where pure competition prevailed.

Examples of this limitation upon the freedom of the producer in an imperfectly competitive market may easily be found. There are only three established makes of low-priced automobiles, and there is very considerable individuality to each. Therefore as we have several times had occasion to note, the automobile industry provides an excellent example of operation under conditions of imperfect competition. Yet no one would deny that competition among the three companies is very keen. Except in boom years when output is enormous they do not earn an abnormally high rate of profit. Furthermore in the periphery of the industry there are a number of marginal concerns which have been through bankruptcy or are close to it. If the "Big Three" were able to make very substantial profits year in and year out, it would be possible for one of these smaller competitors to put out as good a car at a lower price and thus to move the demand curves for the three major makes downward and to the left so that their profits would in time be reduced.

Ascertaining production facts. In the chapter immediately preceding we saw how the producer, seeking to adapt his policy to a given demand for his product, attempts by every possible means to acquaint himself with the facts of demand. We can now appreciate that the producer, especially when selling in an imperfectly competitive market, is

equally interested in a second body of information, having to do with the actions not of his customers but of his competitors. And he is likely to regard as competitors, not only those engaged in his own line of production, but also those engaged in other lines which may compete with his product for the money of the consumer. Business men are thus interested in average and marginal costs of production, in capital charges and overhead, in wage scales and costs of materials, in stocks of merchandise on hand, in the progress of goods in process of manufacture, in freight charges, and so on. And all these matters interest them, not only in connection with their own enterprises, but with respect to industry as a whole. In short they desire to know all they can about production, realizing that these are factors which react upon the demand for their own products. In recent years well-organized efforts are being made to ascertain the significant facts bearing upon the seller's side of the market of all important products.

Sellers' decisions are in the long run predicated by certain definitely measurable conditions surrounding any particular commodity. The most important of these conditions is the *potential supply*, or *stock*, of the commodity which is available for sale. This potential supply is the resultant of two factors; *viz.*, current production, and stocks of the commodity produced in past periods but still unsold, known as *stocks on hand.* Current production may be measured in physical units, such as barrels of crude petroleum, bushels of wheat, board feet of lumber, etc. It is sometimes more suggestive to present the changes in current production by means of index numbers indicating by what percentage current production is greater or less than the production of some selected past period. Similarly stocks on hand are measured both in terms of physical quantities and by index numbers.

But even these data fail to present the complete picture. Properly to interpret potential supply, it is necessary to know whether stocks on hand are larger than normal and whether

current production is running above or below normal. To permit the presentation of these data clearly, statistical technique has developed the concepts of long-time movement or secular trend, seasonal variation, cyclical fluctuations, and residual fluctuations. These concepts make possible definite conceptions of what is normal.

The growth of population and the increase in productive output, with the resulting elevation in standards of living, provide the basis for the statistical assumption that the long-time movement in the production of any commodity will be continued. The recurrence of seasons with their influences on production provides the basis for the assumption that these seasonal variations will persist in marked degree as long as the influences of weather on production continue. Thus combining the expected effects of long-time movement and seasonal variation on the production of a particular commodity, we derive a measurable concept of normal. It then becomes possible to discover whether current production at any particular time is above or below this normal.

By the use of these devices, the seller is enabled to know whether production is outrunning consumption. If production is exceeding normal expectations and stocks on hand are increasing, potential supply is increasing, and decreases in price threaten. If, on the other hand, production is above normal but stocks on hand are declining, consumption has exceeded past expectations, and prices can be expected to remain stable or even increase. For example the following figures show a condition of increasing production, increasing stocks, and falling price in the gasoline market (the quantities being monthly averages for the respective years, expressed in barrels): [1]

	1921	*1928*	*1938*
Production, raw at refineries	10,225,000	31,432,000	46,321,000
Consumption	8,950,000	27,403,000	43,471,000
Exports	1,058,000	4,302,000	3,478,000
Stocks at end of month	15,018,000	33,759,000	73,194,000
Price, wholesale at New York, per gal.	$0.261	$0.174	$0.127

[1] *Survey of Current Business*, March, 1939, p. 45.

While the seller is primarily interested in his own particular commodity, he is not immune from developments throughout the whole range of business. There are available for his purposes elaborate indexes of production, summarized by large classes, such as iron and steel, textiles, food products, automobiles, mineral production, etc.

Decisions of sellers are predicated, not only on the potential supply that may be thrown on the market in the immediate future, but also on the supply in the more extended future. Sellers accordingly look to more general indicators than those of current production and stocks on hand. Thus the expansion of bank deposits, the import or export of gold, the policy of the Federal Reserve Banks, the activity in security markets, the productivity of labor, and the policy of labor organizations represent influences acting upon interest, cost, wages, prices of raw materials, freight rates, etc.

Monopoly: Imperfect competition and monopoly. The foregoing discussion of imperfect competition leads naturally to the subject of monopoly. There are monopolies controlling the sources of certain products of nature, such as the Carlsbad mineral waters, the diamonds of Africa, and the anthracite coal of Pennsylvania. The possessor of any unique non-reproducible article, such as a famous work of art, is of course in a monopoly position. A famous actor or musician has a monopoly in the personal services which the public demands of him alone. The United States government has a monopoly of the service of carrying letters. Patents and copyrights give monopolies to inventors and authors, and city governments grant monopolies to corporations in the form of exclusive franchises to operate street railways, to furnish water, gas, electricity, etc. Finally there are those great corporations and combinations, in fields in which the advantages of large-scale production and combination are especially strong, which have come to positions at least approaching monopoly in their particular fields.

Our investigation of imperfect competition readily suggests

that actually the monopolist differs in degree rather than in kind from the ordinary industrialist. The local electric light company may be a monopolist because no one else may sell electric current in the same district, but there is a demand curve for the product of the local utility just as there is a demand curve for the output of Mr. Ford's company. Furthermore recent experience would appear to indicate that the demand for electric current is highly elastic. Therefore the president of the local utility has to make a decision of precisely the same kind as that which Mr. Ford must make. The most difficult question which each of these men must answer concerns the increase in sales which will result from any given reduction in the price of what he sells. Each must decide what price to set or what output to throw on the market with reference to exactly the same criteria.

The essential distinction between a monopolist and a business man selling in an imperfectly competitive market is that the monopolist sells a product which has no close substitute, whereas the ordinary competitive business man, although he may sell a highly individualized product produced by no one else, nevertheless must as a rule meet the competition of close substitutes. For instance although there is only one manufacturer of the Ford car — the Ford Motor Company — there are other cars which are nearly its equivalent in quality and in price. The fact is that these substitutes also limit Mr. Ford's power with regard to the setting of price. On the other hand, the local water company does not have to compete with any close equivalent and may thus properly be called a monopoly.

From this distinction we may infer that the demand curve for the monopolist's product is usually less elastic than that facing the business man who has competitors but sells in an imperfectly competitive market. If the power company increases its rates, its consumers generally have no alternative but to economize. If Henry Ford raises his prices, his customers may turn to Plymouth and Chevrolet. Though still

only a difference in degree, this is a real and significant distinction between the two for the purposes of economic analysis.

Monopoly and the public interest. The seriousness of a monopoly situation is mostly dependent upon whether the monopoly controlled commodity is a luxury or a necessity. We have already had occasion to note that the demand for necessities is usually very inelastic. If therefore the necessity is produced or sold by a monopoly, it will pay the monopolist to set his price extremely high, because marginal revenue will be much lower than price (in fact it will probably be negative for most outputs) and thus, when the monopolist has equated marginal cost and marginal revenue, price will be high. On the other hand, monopoly control of a luxury is no great menace. The monopoly will, in its own interest, generally refrain from charging an exorbitant price. And if its price is excessive, the people can do without this particular article.

Anticipating our later study of the subject of monopoly, it may be stated here that in general the public interest is better served when production is carried on under the régime of competition. There are certain exceptions, some self-evident, others not so obvious. Personal monopolies, such as those enjoyed by the popular singer or actor, are inevitable and can do the public little harm. The monopoly granted by the patent or copyright is generally in the public interest. And there are certain lines of industry, such as railroading, telephone, telegraph, the furnishing of water, gas, electricity, etc., which from their economic nature appear destined to be monopolies whether or no. In such cases it is futile to put our trust in competition which cannot endure; monopoly must be accepted and made the best of. This topic will present itself for our further study in a later chapter.

Public opinion today generally holds that abuse of monopoly power, particularly in the necessities of life, should so far as possible be prevented in order to safeguard the public in-

terest and that when monopoly is inevitable government regulation should be relied upon to ensure to the public good service and reasonable prices. For this general opinion there is firm foundation in economic principles.

EXERCISES

1. Assume the following conditions of demand and cost of production for a low-priced automobile:

Price	*Sales*	*Marginal revenue*	*Average total cost*	*Marginal cost*
$875	100,000	$875	$1,310	$310
780	200,000	685	813	315
700	300,000	540	650	325
635	400,000	440	573	340
584	500,000	380	530	360
540	600,000	320	508	400
500	700,000	260	500	450
465	800,000	220	503	520
435	900,000	195	516	620
410	1,000,000	185	542	780

(*a*) Plot on a sheet of graph paper the four curves of demand, marginal revenue, average total cost, and marginal cost.

(*b*) What is the best output for the manufacturer to produce? Reading from the chart, at what price can this output be sold? What is the average total cost of producing this output?

(*c*) What is the profit per unit and the total profit under these conditions?

2. A new manufacturer of automobiles enters the low-priced field. The additional competition reduces the demand for the product of the first manufacturer, described in Exercise 1. Assume that the new condition of demand for his product is as follows:

Price	*Sales*	*Marginal revenue*
$785	100,000	$785
700	200,000	615
630	300,000	490
572	400,000	398
524	500,000	332
485	600,000	290
450	700,000	240
418	800,000	194
390	900,000	166
365	1,000,000	140

(*a*) Using the chart constructed in Exercise 1, plot the new demand curve and the new marginal revenue curve.

(*b*) What is now the best output for this manufacturer to produce? Reading from the chart, at what price can this output be sold? What is the average total cost of producing this output?

(*c*) What is the effect of the appearance of a new competitor on the price of the product of the first manufacturer and on the volume of his profits?

3. The Federal Trade Commission, a government agency charged with the suppression of monopoly and the preservation of competition, has devoted considerable effort to the prevention of "brand imitation." Does an increase in the similarity or "closeness" of available substitutes tend to make competition more or less perfect? Explain.

4. (*a*) The Harvard Stadium has a seating capacity of about 60,000. The Athletic Committee sets a price of $2.00 per ticket for the game with Shady Rock and the stadium is half-filled. Draw a diagram to illustrate this situation. Reading from your diagram, at what price could the stadium have been sold out?

(*b*) The Committee sets a price of $5.00 for the Yale game and is unable to fill all orders for tickets. Draw a diagram to illustrate. Reading from your diagram, how many tickets could have been sold at $5.00, if the seating capacity of the stadium had been large enough? What price could the Committee have charged and still have sold out the stadium?

CHAPTER XIV

Interrelationship of Prices. Price Theory in Practice

Interrelationship of prices. We have been investigating the way in which prices are determined by the forces of demand on the buyers' side of the market and supply and cost of production on the sellers' side, under various conditions of pure and imperfect competition. It has been recognized throughout this investigation that these immediate forces which directly determine price are themselves the resultants of a great variety of forces. Among these forces, acting thus indirectly upon the price of any good in any market, will always be found the prices of other goods or other markets. The limits of this book do not permit full investigation of the innumerable price relationships here suggested. We may however profitably give special attention to some of the more direct and fundamental ways in which prices are related to each other.

The present value of future income: Capital and income. The theory of price is incomplete without inclusion of the most fundamental of all price relationships; namely, that between the value of future income and the present value of capital goods. We have learned that all production requires capital.[1] To obtain income man must have capital goods, and it is the production of income which alone gives significance to capital. In a primitive society this effort to secure capital takes the direct form of devising and constructing the weapons, tools, buildings, and other instruments by means of

[1] In the present discussion we shall use the term capital in the broadest of the several meanings which have become irretrievably attached to it; *i.e.*, as a synonym for wealth, thus including land as well as man-made instruments. The term is generally so used in discussions of investment and when it is desired to give prominence to the distinction between wealth, the fund, and income, the flow. Cf. Chapter I and the footnote on page 47 (Chapter III).

which the productivity of man's labor is multiplied and his enjoyable income increased. Under such simple conditions therefore it is customary for the same man both to bear the final costs of capital — the labor of making the instruments and of keeping them in repair — and to consume its benefits. Neither capital nor income is frequently bought and sold; hence neither acquires an exchange value, and there is no tendency to dissociate capital from income and regard it as an end in itself. Capital is viewed in its proper light as a means toward an end. Like other means, each unit of capital is more or less prized according to its efficiency in achieving the end in view; that is, its power to produce income for its owner.

In the social order of today there are to be found many instances of this direct and visible relation between capital and income. Men sometimes gather materials and with their own labor construct some simple capital instrument whose benefits will accrue to themselves directly. The farmer may cut and shape a handle for his axe or build a fence or wagon shed. A blacksmith may make some of his own tools. In such cases the relation of capital to income is clearly grasped; one sees that it is the expected income which bestows on capital its importance and that the owner's appraisal of the capital is but a reflection of his evaluation of the benefit which he hopes to obtain from it.

Property and capital. But these simple cases are not typical of present-day society. So complex has our economic structure become that we often lose sight of the fact that capital has no significance when divorced from the benefits which it produces. Chief among the attributes of our social system which have this effect is the fact that the ownership of capital has been subdivided into a large variety of property rights, which are bought and sold in great numbers. Those property rights are usually represented by documents of different kinds — stocks, bonds, mortgages, contracts, and the like — which frequently have stamped upon them some nom-

inal valuation (for example the $100 "par value" on a stock certificate), which misleads the owner into the belief that these documents have a "normal" or "true value" in themselves. The development of many agencies with the purpose of facilitating trading in these property rights — such as banks, brokerage and investment houses, stock exchanges, and so on — has aided in beclouding the true relation of capital to income.

As we have learned, property is but a right to the future benefits of wealth (capital). The benefits are produced by the capital which underlies the property; it is the function of the property right to divide this income among the owners of the capital. It is apparent therefore that the principal reason for the existence of the certificate of property is that it affords a convenient method of transferring ownership in capital from one man to another and of subdividing the ownership. These certificates are bought and sold as representatives of the valuable thing (capital) in which they transfer ownership; hence the action of the buyers and sellers in the market will result in giving the certificates the same value that in the opinion of these buyers and sellers should be placed on the capital which they represent. If for example a manufacturing concern has issued one thousand shares of stock and these shares sell in the market for $150 each, there can be but one reason for this; namely, that in the opinion of investors a one-thousandth share of the owner's equity in this corporation is worth $150. The fact that the stock has stamped upon it a nominal value of $100 or $200 or $50 or any other figure is not competent to determine the value of the share.

Present worth. The real problem therefore is to discover how the value of capital goods is determined. Here we must repeat that capital is desired only because of the income it produces and is bought only as a means of obtaining this income. Let us suppose that a given machine will last ten years and produce a net income of $100 a year. If this machine is offered for sale to the owner of a factory, he will con-

sider that by owning it he can increase his income during the next ten years by the amount of $100 a year, or $1,000 in all. Whatever price he is willing to pay for the machine will be a measure of the present value he puts upon this thousand dollars, distributed as it will be through the ten years to come. To change the illustration, we may suppose that an acre of land is offered for sale to a man who has reason to believe that it can be made to produce a net income worth $10 a year forever. In this case too the price offered by the buyer for the land will be his estimate of the *present worth* of an annual income of $10 which continues indefinitely. Or again an investor may be offered a bond which will be redeemed for $100 at the end of five years and in the meantime pay $5 a year interest. Such a bond would be viewed as an opportunity to receive a series of money payments during the next five years: $5 one year from now, another $5 two years from now, another $5 three years from now, $5 more four years from now, and a final payment of $105 after five years have elapsed. The amount of money that the investor is willing to part with today in order to buy the bond must be an indication of what he considers this series of future payments to be worth at the present moment. This is the essential thing in calculating the value of any capital good or of any property right in capital goods. It amounts to placing a present value on an expected future income of a certain size and form and duration.

Present goods more desirable than future goods. It is a well-known fact that men do not consider a sum of money which will accrue in the future to be worth its face value now. The investor referred to in the preceding illustration will not calculate the present value of the bond by simply adding together the different payments which will accrue in the future to its owner. Such a calculation would give the bond a present value of $125; and the buyer paying this price for it would eventually get back just the amount of money that he paid, though he would have to wait five years for most of it. If one

offered for sale a promissory note which would pay its owner $100 twelve months from today, he would find no one willing to pay $100 for it now. Nor would this reluctance to buy this future income at its full face value in present money be due only to risk of loss. If the $100 were locked up so safely that all doubt regarding its security were removed, still no one would be willing to pay out $100 now and be content to get it back after twelve months had passed. In the mind of the ordinary man the present value of the future dollar is less than that of today's dollar; hence sums of money which accrue in the future will be bought only if they can be had for less than their face value. And the price that will be paid for capital which gives its owner the right to receive income in the future will be lower, by some substantial amount, than the value of this future income when it accrues.

Interest. If asked why he will not pay $100 now for the right to receive $100 one year hence, the practical man will probably answer that he is unwilling to lose interest on his money. Because of the existence of interest, it lies in the power of any man who is willing to exchange present wealth for future wealth to make the exchange on terms which will bring him a future return of larger value than his present payment, and as long as this opportunity to gain interest on the money they lend is open to all men, no one will consent to pay its full future value in present money when buying income. *Interest* may be defined as *the premium which present goods command over similar goods due in the future.* The man who lends $100 for one year at 5 per cent interest exchanges $100 worth of present goods for $105 worth of next year's goods, the premium of $\frac{5}{100}$ being required to give the future goods a present value equal to that of the present goods which the lender surrenders. As the interest rate rises and falls the premium which must be added to make these two amounts — one present, the other future — of equal present value rises and falls accordingly. It is, therefore, the *rate of interest* which determines how much larger the future return must be than

the present sacrifice in order to cause the exchange of the two to occur.

The link between future and present values. We need not at this point attempt to explain why interest exists or how the rate of interest is determined; these are problems to be solved later in our study. For our present purpose it suffices to accept the existence of interest as a feature of the present business world. Our immediate concern is to show that the rate of interest supplies the link between future and present values which enables us to translate the one into the other.

Given the rate of interest, it is a simple matter to calculate the equivalent future value of any amount of present money by means of the following formula:

$$A_n = P(1 + r)^n$$

in which:

A represents the future value, any number of years hence,
P represents the present value, or "principal,"
r represents the rate of interest,
n represents the number of years.[1]

For example, when the interest rate is 5 per cent, \$100 invested now will be worth \$105; *i.e.*, \$100 × 1.05, one year from now; \$110.25; *i.e.*, \$100 × $(1.05)^2$, two years from now; \$115.76; *i.e.*, \$100 × $(1.05)^3$, three years from now; and so on.

[1] The demonstration of this formula is as follows:

At the end of the first year, the amount (A_1) is the sum of the present value, or principal, and one year's interest; that is,

$$A_1 = P + Pr = P(1 + r).$$

That is, the amount at the end of the first year is the principal multiplied by $(1 + r)$.

This amount, $P(1 + r)$, is the new principal for the second year. The amount at the end of the second year (A_2) will be found in exactly the same way as for the first year; that is, by multiplying the principal at the beginning of the year by the factor $(1 + r)$. Thus

$$A_2 = P(1 + r)(1 + r)$$
$$= P(1 + r)^2.$$

In the same way, we find at the end of the third year the amount is $P(1 + r)^3$, at the end of the fourth year, $P(1 + r)^4$, and so on for any number of years.

This same formula may be adapted to the reverse problem of determining the present worth of a given future value by merely rearranging its terms, as follows:

$$P = \frac{A_n}{(1 + r)^n}$$

the symbols having the same meaning as in the preceding formula. To illustrate, assuming again an interest rate of 5 per cent, \$100 due one year hence is worth now \$95.23; *i.e.*, \$100 ÷ 1.05; if due two years hence, it is worth now \$90.70; *i.e.*, \$100 ÷ $(1.05)^2$; if due three years hence, it is worth now \$86.31; *i.e.*, \$100 ÷ $(1.05)^3$; and so on.

These numerical examples assume that the interest rate is so much *per annum* and the interest is "compounded" annually. The formula holds however for any unit of time. All that is required is that the rate of interest be stated with respect to the same period as is used in compounding. For example it is required to find the amount of a loan of \$5,000 for two years at the rate of 1 per cent a month, interest compounded semi-annually. Since the interest is compounded every half year, we must take a half year as the unit of time. Then we must call the rate of interest, instead of 1 per cent per month, 6 per cent per half year. Instead of saying the loan is for two years, we must say it is for four half years. Then our formula becomes $A = 5{,}000(1.06)^4$ and, when we solve, we find that $A = \$6{,}312.38+$.

The formula which determines the present worth of a given future value will serve likewise to place a present worth upon a series of future values. What is the present worth of a series consisting of \$200 due one year hence, \$500 due a year later, and \$100 due at the end of the third year, if the rate of interest is 5 per cent? By means of the formula, $P = \frac{A_n}{(1 + r)^n}$, we find the present worth of these future amounts to be respectively \$190.48, \$453.52, and \$86.31. The total of these

three amounts, \$730.31, is the present worth of the series of future values. By this process, one evaluates the ordinary bond. Thus a 5 per cent \$1,000 bond due in 7 years is valued as a series of 6 annual payments of \$50 and a final payment of \$1,050 at the end of the seventh year.

It is frequently important to evaluate a uniform perpetual series of future values, as in the case of a perpetual government bond. The general formula for the present worth of a uniform perpetual income is $P = \frac{i}{r}$, in which P stands for the present worth, i stands for the periodical payment, and r stands for the rate of interest.[1] Thus, assuming an interest rate of 3 per cent, the present worth of a 4 per cent \$1,000 perpetual bond (like the British consols) would be \$40 $\div \frac{3}{100}$ = \$1,333.33.

The discount process. Although it is possible by means of the interest rate to calculate either future values from present or present values from future, and although in the business world there is continual use for both of these calculations, it is the latter process which has the greater importance. This is true for the reason stated at the beginning of this chapter;

[1] This formula is derived, as a special case, from the general interest formula, remembering that, in this case, the interest is not compounded but paid each year, so that the principal is brought back each year to the original amount. According to the general formula,

$$P = \frac{A_n}{(1 + r)^n}.$$

Assuming a single interest period, $n = 1$. Then

$$P = \frac{A}{1 + r}.$$

But $$A = P + i.$$

Hence $$P = \frac{P + i}{1 + r}$$

$$P + Pr = P + i$$

$$Pr = i$$

$$P = \frac{i}{r}.$$

namely, that the economic activity of mankind is directed toward the obtaining of income and all income which has power to give value to capital lies in the future. The valuation of capital instruments and property rights is incidental to this struggle for income, being a product of the incessant buying and selling of the means by which income is produced. Hence the calculation of the present values of future incomes of every conceivable variety as regards size and duration is one of the commonest events in the business world. The name given to this calculation of present values of future incomes is the *discount process.*

Summarizing the preceding discussion, we may say that this process requires these two fundamentals: (1) a knowledge of the probable future value of the income, (2) a knowledge of the interest rate. The first of these fundamentals is the basis of the calculation — the future value whose present worth it is desired to obtain. The second supplies the mechanism for determining the present worth; the process being to divide the future value by one plus the rate of interest raised to a power equal to the number of years which must elapse before the future income accrues.

Capital and income. The relation of capital to income may be looked at from either of two points of view; namely, that of the physical relationship and that of the value relationship. From the physical point of view it is capital which produces income; an orchard produces a certain number of barrels of apples; a factory, so many pairs of shoes or other units of output; a house, so many years' shelter, and so on.

But the value relationship is the opposite of this. The value of the capital does not produce the value of the income. The value of the apples gathered from an orchard will be established in the market in which the apples are sold after they have been produced, in accordance with the laws of market price as developed in preceding chapters. Setting a value upon the capital will not avail to dictate the terms upon which its product can be sold. On the other hand, a given set of

capital instruments may be very productive, in the sense that it turns out a large stream of products, but if these products cannot be sold at profitable prices, the value of the capital instruments will be very low, or even non-existent. This is clearly evident of capital when the demand for its products has declined or disappeared. Abandoned farms, factories, and railroads bear testimony to the fact that the value of capital goods is derived from that of their income.

Discount and the laws of price. It is essential that we understand exactly how this process of discount is related to the general laws of price which we have previously developed. We have found that under conditions of pure competition prices are in general determined by demand and supply and that prices under imperfect competition are determined by decisions of the sellers in conjunction with the conditions of demand and with consideration of actual and potential costs. We have learned that back of demand is marginal utility, and back of supply and seller's actions, cost of production. Now it appears that values are determined by discounting future expected incomes. Are we faced with two conflicting principles?

To make our question concrete, suppose we ask: why does a particular tract of land sell for $200? Is it because this is the price determined by the forces of demand and supply in the market, or is it because this is the present worth of an expected perpetual income, say $10 a year discounted at 5 per cent? The answer is "Yes" to both questions, and the absence of any conflict will be apparent as soon as we inquire into the part which discounting plays in the market. Values are not determined directly by discounting. In the simple numerical examples which we have thus far employed, we assumed certain future values and we assumed a certain rate of interest. In the practical world, neither of these essential factors is objectively determined. The future incomes — just because they are future — are uncertain, more or less. The rate of interest is the rate at which any given person will discount the future.

Every potential buyer and every potential seller must come to his own decision as to what these factors are in any given case. A, on the basis of the information available to him and his best judgment, decides that the tract of land in question will probably bring forth such and such future incomes, that these incomes will have certain values to him, that he is prepared to discount the future at a certain rate of interest; with these factors thus estimated, he goes through the discounting process and determines how much he is willing to pay for this piece of land. B makes similar estimates, arriving possibly at different figures from A at every point, and finally emerging with the price he is willing to pay. So of all other possible purchasers. And the present owner of the land must himself go through a similar process in order to determine at what price he will be ready to sell. When the potential buyers and sellers meet in the market, the discounting process has already taken place in the mind of each. If the tract of land in our example sells for $200, it is because the seller considered it worth something less than that to him, and the purchaser considered it worth something more. In arriving at his judgment, each employed the discounting process. The discounting process is itself one of the forces that determine demand and supply, which in turn determine price. Discount, so far from being in conflict, is itself included in the principle of demand and supply.

We recognize of course, as in all such theoretical analyses, that every person does not consciously go through the mental process here visualized. Some buying and selling, as we have seen, is more or less irrational. But back of demand and supply will always be found human judgment discounting expected future incomes. Even where — as appears to be the normal case on the speculative stock and produce markets — dealers seem to make their decisions, not with any regard to discounted future incomes, but on the basis of expectations as to what prices will be in the future, the fundamental principle is still in operation, though somewhat obscured. For in the

last analysis, back of all estimates of future prices, there is always someone's forecast of future income and its discount at a rate of interest. The dealers are simply cutting corners to arrive at the result. The market rate of interest itself is a composite judgment of the relation between future and present values.

Discounted income and cost of production. There are a great many goods that are staples in the market. They are constantly being reproduced as a matter of routine business practice; day after day they appear in a steady stream upon the markets of the world. The producers of these goods, after they have offered their supplies for sale, must be content with a price determined by the laws of market price, in the operation of which the discount process will govern the demand of the buyers. But what if the estimate thus placed by the buyers upon any given quantity of the goods differs from the cost of producing this quantity? The discrepancy will be resolved, as we have learned, through the power which the producers have either to avoid future costs by giving up or diminishing the production of these goods or, on the other hand, to increase future costs by increasing their output. Their decisions on this point will, after an interval of time, exert an influence upon the value of the capital goods they are producing. If they produce an increased quantity of the capital goods, the services rendered by them will increase in quantity and command a lower price; hence the discounted value of the capital goods will fall. On the other hand, if the producers reduce their output of the capital goods, the peculiar benefits which flow from these goods will become scarcer and their price will rise, hence raising the discounted value of the capital goods. In other words, it is the willingness or unwillingness of the producers to bear future costs which, by affecting the value of the income from capital, brings marginal utility and marginal cost into harmony.

In the case of standard and staple goods, the markets for which have settled to a point of relatively stable equilibrium,

the producers may confidently take the present price as an indication of what the future price will be. If the present price of the good stands very much above the cost of making it, existing producers will be tempted to expand their output; while at the same time other enterprisers will be tempted by the prospect of large profit to engage in its production. As the number of instruments in use increases, the value of the income which one instrument can earn will decrease, and consequently the discounted value of the instrument will fall. The marginal cost of producing the instrument will rise as attempts are made to supply it in larger quantities. As a result the discounted value and the marginal cost, which at the outset were different from each other, will tend to draw together.

On the other hand, if the present price of the capital instrument had been below the marginal cost, enterprisers desirous of avoiding the losses caused by this situation would abandon the production of the instrument entirely or contract their output of it. This restriction in the supply of the instrument would cause the income it produces to grow scarcer and higher in price, with the result that the discounted value of the instrument would rise, and the diminution in the production of the capital instrument would reduce the marginal cost of production. Again there would be a tendency for the two — marginal cost and discounted value — to approximate each other.

In the case of capital goods which are not being currently reproduced, discounted value may part company from cost of production indefinitely. Then it is discounted future income, not cost of production, that determines the value. The construction of dwellings, office buildings, hotels, and factories will not be undertaken unless the builders expect the selling price to be as high as the cost of production. But after the costs involved in the erection of a given building have been borne and the building is offered for sale, the price will be determined by the buyers' estimate of the present worth of its future products. A summer resort constructed at great

expense in an unpopular place will command a price lower than its cost of production because the discounted value of the net income to be derived from its operation is low.

Prices of the same good in different markets. Turning now to a different type of price relationship, let us suppose that the price of wheat in Chicago is $0.80 a bushel and in New York $1.05 and further that the entire cost of transporting a bushel of wheat from Chicago to New York is 15 cents. There is evidently a profit to be made by buying wheat in Chicago, shipping it to New York, and there selling it; and there is a class of business men ever on the alert to take advantage of such opportunities. Such dealers will immediately enter the market, seeking the profit of 10 cents a bushel which the existing conditions indicate. But these very operations will inevitably change the conditions. The dealers enter the Chicago market on the demand side, increasing the demand for wheat and so raising its price. In New York, on the other hand, they appear upon the seller's side and tend to increase the supply and lower the price. Even so, the margin of profit, though declining, continues for a time, and a considerable quantity of wheat will be transferred from Chicago to New York. The prices on the two markets will approach each other until finally the difference is only 15 cents; let us suppose the Chicago price has risen from $0.80 to $0.84 while the price in New York has fallen from $1.05 to $0.99. The two prices are now in equilibrium with respect to each other, differing by exactly the cost of transportation. In such manner the prices of the same good in different markets may react upon each other.

Such reaction is obviously possible only where there would otherwise be a difference in price greater than the cost of transportation from the low-price market to the high-price market. If the price of brick is $10.00 per thousand in New York and $8.00 per thousand in Chicago, while the cost of transportation between the two cities is $20.00 per thousand, these prices are evidently independent of each other, since no

one will be tempted to buy brick in one market for sale in the other.

The extent to which prices in a given market are affected by the prices of the same good in other markets thus depends upon the cost of transportation in relation to price. Why were the commodities which entered into the long-distance trade of the medieval merchants so generally articles of luxury — jewels, silks, fine wines, spices, perfumes, etc.? The answer is that heavy transportation costs made it unprofitable to bring to market from remote places anything that could not be sold at a high price. Cheap and bulky commodities, such as farm produce, could be sold profitably only in the markets close to the place of production. Since that time the transportation system has become so extraordinarily efficient and the price of transportation so low that the province of trade has been enormously extended. Cheap and bulky goods, such as wheat and bananas, are shipped thousands of miles over land and sea to be placed at the disposal of distant consumers. With each step in the progress of lower transportation costs, the realm of these price reactions is widened. For many products, such as wheat, cotton, the precious metals, etc., the markets of the whole world are thus tied together. This subject will be further investigated when we undertake the study of interregional trade, at which point we shall better realize how vitally these price relations affect the well-being of mankind.

Prices of competing goods. In the winter of 1925–26 there was a long continued strike of the miners in the anthracite coal mines of Pennsylvania, which for a considerable time virtually stopped the flow of such coal to the markets. The consequent decline of supply raised greatly the price of anthracite. As a consequence, people sought other kinds of fuel, such as bituminous coal, coke, etc. This meant an increased demand for such fuels and in consequence higher prices. Suppose there is in a certain market a sudden increase in the supply of beef. Beef prices will decline in consequence, and more

people than usual will buy beef. This will entail a decrease in the demand for lamb, let us say, and the price of lamb will therefore go down. These events illustrate another sort of price relationship; *i.e.*, the relation between the prices of substitutes or competing goods, with which the reader has already become acquainted through earlier chapters. We have the general principle that the prices of competing or substitute goods are so related that they tend to rise and fall together.

Prices of complementary goods. Suppose some change in conditions of production should bring about a material increase in the price of bread. People would naturally incline to diminish their consumption of bread, seeking to find similar nutritive satisfaction from other kinds of food. The result might be, in accordance with the law of competing prices, a rise in the price of meat or vegetables. But what, if any, would be the effect upon the price of butter? Butter is chiefly used in connection with bread. Using less bread, people would want less butter. The demand for butter would decline and its price would fall. Similarly a fall in the price of bread, by causing its increased use, would increase the demand for butter and tend to raise its price. This is one example of a pair of complementary goods, meaning goods which are used only or chiefly in conjunction with each other. There are plenty of other examples. In the extreme case, the use of the one good is possible only in connection with the other; for example tennis rackets and tennis balls are completely complementary goods. It is evident that the prices of complementary goods tend to move inversely to each other, the reverse of the relation between the prices of competing goods. For example horses and mules are substitute or competing goods; their prices tend to rise or fall in unison. Horses and wagons are complementary goods; their prices tend to move inversely.

Prices of competing cost goods. The price relations of competing goods and complementary goods arise out of the reactions of the buyers; they are on the demand side of the

market. There is another relation between competing goods (substitutes) which we have not as yet encountered. Labor is normally devoted to those pursuits which offer the greatest return. Consider the case of a small town in which there is one important manufacturing industry making, let us say, grindstones. A rise in the price of the product of this enterprise will cause its owner to seek to expand his output; for this purpose he will require more labor, which he will draw, by offer of higher wages, from the dairy farms and market gardens of the neighborhood. The resulting shortage of agricultural labor will cause a decreased supply of milk and vegetables in the local market, with a consequent rise in their prices. Thus a rise in the price of grindstones has caused a rise in the price of vegetables. Products such as these, both of which require the same cost element, make competing demands for the factors of production — competing in this example for the use of labor. They may be called "competing cost goods." Their prices evidently tend to move up and down together. This relation is quite distinct from the one which arises from the demand side; that is, from the fact that a rise in the price of one good increases the demand for all substitutes.

Prices of joint cost goods. Where cattle are slaughtered for beef, hides are made available, as well as horn, tallow, and a great number of other products. These products appear jointly as the results of the same production costs. Having determined to produce a certain quantity of one, the industry will inevitably bring forth the corresponding quantities of the other products. Where the products are of unequal importance, the less important is called the *by-product.* For example the growing of cotton necessarily yields a crop of cotton seed The latter has its uses, but in general cotton would not be grown for the sake of the seed; that appears incidental to the production of the primary crop. Such products as these are said to be the results of joint costs, and their prices have a definite relation to each other. Let us suppose that some-

thing causes the price of cotton to be especially high. To take advantage of this, the planters will increase their production of cotton, thereby inevitably increasing the supply of cotton seed. The price of the latter will consequently be reduced. A fall in the price of cotton would discourage production, thus leading to a decline in the supply of cotton seed and a rise in its price. The prices of joint cost products thus tend to move inversely, being the same relation as was found to exist between the prices of complementary goods. In recognition of this analogy, joint cost goods are sometimes called "complementary on the supply side."

Prices of "tandem" goods. Where one good is the material from which another good is made, there exists still another sort of price relation. If the price of wool rises, there is caused thereby an increase in the marginal cost of producing woollen cloth. This will cause a reduction in the supply of woollen cloth; *i.e.*, at any given price the quantity offered for sale will be less than before. Hence the price of woollen cloth rises. This is a simple example of a principle that is a commonplace of everyday observation; *i.e.*, that the prices of manufactured goods tend to vary directly with the prices of the materials that enter into their production.

Price theory in practice: Practical aspects of the theory of value. The very considerable time which we have devoted to study of the theory of price is justified by the importance of these concepts in every department of business and practical affairs. Inasmuch as the modern social system rests upon private ownership of wealth and free exchange, problems of value are at the heart of the practical enterprises of mankind. This appears nowhere more clearly than in the field of investment. The ownership of corporations (in the broadest sense) is expressed in their stocks and bonds. These securities are bought and sold, in some cases daily or even hourly on the exchanges, in other cases more or less rarely and privately. Values are determined thus by the action of demand and supply, and back of demand and supply is always the process

of discount or capitalization of income. Consciously or unconsciously, every buyer or seller or owner of corporate securities is making practical use of value theory in reaching his decisions to buy or to sell or to hold.

That the values of corporate shares sometimes appear to be quite out of harmony with the current dividends, so far from being denial of the validity of our theory of value, is the best proof of its truth. Before the sensational stock market crash in the United States in the autumn of 1929, the prices of American stocks generally were very high. For example the common stock of the United States Steel Corporation reached its high point of 261¾ on September 3. The annual dividend rate being 7 per cent, purchasers at this price and owners who held at this price were apparently satisfied with a return of only 2⅔ per cent. Now the fact is that American stockholders are not satisfied with any such return as that, and the explanation is that these investors were not so naïve as to discount only present income; they were acting in accordance with sound theory and discounting the future incomes expected to flow from shares in the United States Steel Corporation, based upon surpluses already accumulated and expected future earnings. September, 1929, witnessed the culmination of a long period of buoyant optimism among investors, and the general anticipation of future earnings of American corporations was high. The reaction is seen in high values and low apparent yields all along the line.

That these rosy views of the business future were exaggerated and not warranted by the facts was the opinion of some of the more conservative observers as early as the summer of 1929, and the correctness of this opinion was demonstrated by the collapse of values that occurred in October and November. United States Steel dropped from its high price of 261¾ to 150, on November 13, at which price it stood to yield on the basis of the past year's dividends 4⅔ per cent.

The important lesson for our present purpose is that it was the investors' and speculators' estimates of future facts, not

their theory of value, that was wrong. And this brings out clearly the principle that all value is the resultant of human judgment, human estimates of future events. Obviously no one is sure of the future. In any broad market, such as the stock market, there are dealers of all degrees of knowledge and judgment, from those who by resort to all possible sources seek to inform themselves of conditions and exercise the best judgment that is humanly possible, to those who act, without knowledge or intelligence, on the basis of rumors or "tips" or mere guesswork. The resultant demand and supply are the consequences of the composite judgment of these buyers and sellers, weighted in proportion to their respective resources. Knowledge and ignorance, sound judgment and recklessness, conservative investment and pure gambling, all combine to determine prices, through the chain of cause and effect which we know as the laws of price or value.

The overworked term "value." Although the concept of value as used in the science of economics is thus basic in practical affairs and is so recognized by practical men, there have developed in certain administrative fields other concepts, closely related to value though not precisely the same, which generally go by the name of "value." Thus the accountant finds it necessary to assign values to the various items in his statements. Starting with the basic notion of value, accounting practice has evolved certain concepts which serve its purposes but depart more or less from the economic notion of value. Thus items of wealth, such as land, buildings, machinery, etc., may be carried on the books on various bases of cost, cost less depreciation, replacement cost less depreciation, etc., even though these "values" may be known to be higher or lower than the considerations for which the articles could be sold. "Goods in process" of manufacture are "valued" on the cost accountant's books on such bases as cost of materials plus direct labor cost, with or without a share of overhead, even though in certain cases there might be no possibility of sale of such partly finished products except to another man-

ufacturer in the same line, in the absence of which these goods might have no real value except as scrap. In such cases as this the term value is frankly used to mean something different from the economic concept of value. It is always unfortunate when the same term is used with diverse meanings, but a clear understanding of the usages employed will generally suffice to avoid confusion.

The actual value of a corporation to its shareholders may generally be best shown by the total market value of its stock. This may however differ widely from the net value of its assets as shown by the balance sheet. The discrepancy will be due to a variety of causes. In the first place, the accountants' valuations may depart from market value, as already seen; and, such as they are, they may not be kept strictly up to date. Secondly, the balance sheet may not include certain property rights of an especially intangible or elusive nature, such as the character of management, certain franchise or contract rights, the "good will" of an established clientele, etc. — factors which are combined under the title of "going concern value" and upon which the future earnings of the enterprise may in large part depend. The balance sheet valuation, though serving effectively its own purpose, may thus depart materially from the economic concept of value. That the latter is the fundamental one is demonstrated by the very fact that buyers and sellers of the corporation's securities arrive at values different from those shown by the balance sheet.

In the practical task of regulating the rates of public utility corporations, such as the railroads, the telephone and telegraph companies, gas and electric companies, etc., it is found necessary to ascertain "values" of the capital invested in order that the legally determined rates may permit a "fair return" on the investment. In this task the various official regulatory bodies, such as the United States Interstate Commerce Commission, the state railroad and public utility commissions, etc., and the courts, have developed a technique of "valuation," which scarcely pretends any longer to seek to

arrive at the economic concept of what the property would sell for. What is actually sought is a "rate base" which shall effectively serve the purposes of regulation. That such rate base continues to be called "value" should not confuse the well-informed.[1]

A modified concept of value is likewise used in taxation, especially in America, but unfortunately with less of candor than prevails among those who deal with public utility rate making. Examination of the property tax laws of the forty-eight states would show that these taxes are imposed universally upon the value of the taxable property and that, with few if any exceptions, the lawmakers had in mind the concept of value as used in economic science. The legislatures have gone out of their way to specify "market" value, "cash" value, "actual," "fair," "true and just" value, what the property would bring if sold "by a willing seller to a willing buyer," etc. — terms which indicate their adherence to the economic concept of value. Property tax administration, as the result of circumstances which will be examined later, has not always succeeded in carrying out this idea of value, though still generally maintaining the pretense of a true valuation. Assessed values are very often not the same as true values, and they vary all the way from considerably above the true value to as little as 5 or 10 per cent; as a rule assessments are below true value. It is safe to say that tax administration will never be competent to place true values upon all the complex kinds of property which in America are taxable. For certain classes of property it may become necessary to emulate the frankness of the rate-making authorities and set up a more or less arbitrary "tax base" which will be workable.

The use of the term value in these varying meanings has been the source of much confusion — confusion which is to be escaped only by recognition that book values, rate-base values, tax values, etc., are generally neither the same thing

[1] See G. G. Tunell, *Value for Taxation and Value for Rate Making*, Proceedings of the Twentieth National Tax Conference, 1927, pp. 263–279.

as value in the economic sense, nor consistent with each other. It is not, as is sometimes assumed, indisputable evidence of fraud or corruption when a railroad reports one "value" to the rate-making commission and another to the tax assessor. Nor is there the supposed fine logic in the proposal to stop tax evasion by permitting the state to buy in any property at the valuation on the owner's tax return. Discrepancies between market and book values do not necessarily demonstrate either foolish investing or crooked accounting. And finally the situation does not demonstrate hopeless confusion as to the meaning of value.[1]

EXERCISES

1. Suppose $100 is deposited in a savings bank which pays 3 per cent interest on such deposits.

(*a*) What would this deposit be worth in three years' time if no money were withdrawn in the meantime?

(*b*) What would it be worth in six years?

(*c*) What would it be worth in fifteen years?

NOTE. Solve (*a*) exactly. For (*b*) and (*c*) work out the formula only.

2. Suppose the interest rate is 6 per cent. What is the present value of a promissory note for $1,000 due in two years?

3. A stand of growing timber will be marketable three years hence at an estimated value over and above the cost of cutting of $5,000. What is its present value if the interest rate is 5 per cent?

4. Suppose the interest rate is 5 per cent. What is the present value of a $1,000 4 per cent bond maturing three years hence? This description means that the owner of the bond will receive $40 per year for three years and a redemption payment of $1,000 at the end of the third year.

5. A certain piece of land can be made to return a perpetual net income of $600 a year. What is its present value, (*a*) at a 5 per cent rate of interest? (*b*) at a 4 per cent rate? (*c*) at a 6 per cent rate?

6. In general, if income value remains unchanged what is the effect on the value of capital, (*a*) of a rise of the interest rate? (*b*) of a fall of the interest rate?

7. Draw diagrams illustrating the relation between the prices of competing goods.

[1] For an interesting discussion of these matters, written from a somewhat different viewpoint, see L. D. Edie, *Economics: Principles and Problems*, 1932, especially pages 190–201.

8. Draw diagrams illustrating the relation between the prices of complementary goods.

9. Draw diagrams illustrating the relation between the prices of joint cost goods.

10. Draw diagrams illustrating the relation between the prices of "tandem" goods.

CHAPTER XV

Economic Rent: The Income of Landowners

The meaning of distribution. As used in economics the term "distribution" denotes the division of society's total income among the members of the social group. The study of distribution may be pursued with two different objects in view. We may set out to discover the actual incomes of the different individuals and to explain the inequalities which appear among them; this would be called a study of *personal distribution.* Or we may distinguish between types of income, without reference to the total receipts of individuals, and investigate the forces which determine the size of these different kinds of income. This would lead us into a study of the several factors of production and of the economic laws which control their remuneration. Such an inquiry would be called a study of *functional distribution.*

We shall study both aspects of the problem of distribution in the following chapters, postponing our study of personal distribution until we have made a preliminary study of the functional types of income. There are four of these functional types: economic rent, interest, wages, and profits. Although there are similarities between these types of income when considered from certain points of view, each of them when viewed as a separate category of income is affected as to its amount by independent social forces. It is our purpose to examine these distinctive blocks of income separately, beginning with economic rent.

Definition of economic rent. As used in the discussion which follows, *economic rent* is a technical term denoting the income which accrues to landowners from the use of their lands for productive purposes. Before phrasing the definition of this type of income it will be well to call attention to certain assumptions on which it is based.

In the first place, it must be understood that land is a generic term embracing not only the soil but all the resources of nature afforded by man's physical environment, the mineral deposits, the forests, the waterfalls, the fisheries, etc. Income from any of these things, strictly defined, is economic rent. Thus we may have rent from soil under cultivation, from untilled land used as a building site in a city, from a mine, from a hydro-electric power station — in short from all enterprises exploiting natural resources.

Secondly it should be noted that an income to be properly called rent must be attributable to the land itself and not to some other productive factor. In ordinary business practice when land is leased by the owner to someone else the lease covers articles of wealth imbedded in the land which are not a resource of nature but the product of man's labor. The tenant in the city, for example, commonly obtains the use of buildings and other improvements as well as the site; in the country districts fences, barns, drainage systems often go with the land. The lessor commonly receives a single payment for the use of this composite unit of land and capital and usually speaks of his receipts as "rent." From the standpoint of economic science however this income is made up of two distinct elements: interest on capital and economic rent of land. Frequently the case is even more complex than this. For example, a man may cultivate his own farm with the aid of his own tools and other capital instruments. His income will accrue to him in a single stream, and he will be at a loss to distinguish different elements in it. But in theory only that part of his receipts can be called economic rent which is attributable to the land, as distinguished from wages for his labor and interest on the value of his capital instruments.

In harmony with these assumptions we derive the following definition: *Economic rent is a net income from the use of land, distinct from all income from labor or the use of capital.*

The specific income from land. The first step in the study of economic rent is to ascertain whether these incomes from

land alone exist and if so how they can be measured. Given a plot of ground, a hypothetical method of discovering its specific income is to assume the land withdrawn from use and to note the effect on the total income of society. In this illustration it must be assumed that the labor and capital formerly employed on the land are transferred to other uses, so that there has been no change in society's productive resources save that this one plot of ground has been withdrawn. If the social income remains undiminished, it may be said that this land has contributed nothing; and on this assumption its owner could not induce anyone to pay him an income for the privilege of using it. If, on the other hand, the social income is reduced by the withdrawal of this land from use, this portion of income may justly be attributable to the land and not to labor, capital, or any other productive agent, since these other agents have remained unaltered in amount. Also in this case the landowner can demand that income be given to him in exchange for the use of his land. Other members of society can afford to pay him this income and still be as well rewarded for their labor and capital as before. And if no one is willing to contract for the use of his land, the owner can gain this same income — in addition to payment for labor and capital — by exploiting it himself.

Diminishing returns. As a general rule labor and capital cannot be withdrawn from a part of the land and concentrated on the rest without causing a reduction of social income. The reason is that the exploitation of land and other resources is governed by the *law of diminishing returns*, to which we have already given some attention in connection with our study of price.[1] This law may be illustrated by reference to any situation in which a natural resource of fixed quantity is exploited by increasing applications of labor and capital. Suppose for example that we have a farm of 25 acres producing corn and that additional men are put to work upon it, one after another, each equally well equipped with capital instruments. It is

[1] See Chapter X.

conceivable that two men will produce twice as much as one; but this rate of increase in the total product will not go on indefinitely. The case may very well be as follows: one man produces 500 bushels; a second man adds 500 bushels to the output; a third man, 400 bushels; a fourth, 300 bushels; a fifth, 200 bushels. The total product continues to increase, but the marginal product — the increase due to the labor of the last man — diminishes as the third, fourth, and fifth man are added. The declining marginal product illustrates the principle of diminishing returns. In its most general form the law of diminishing returns may be stated as follows: *after a certain point has been reached in the exploitation of a productive factor of fixed amount, additional applications of other factors to its exploitation, though they will increase the total product, will not increase it in proportion to the increase of the variable factors.* As thus stated, it will be seen that the law applies when any one of the factors is fixed in amount while the others increase. A unit of capital of given size — say ten machines — may be operated at greater intensity by increasing the amount of labor and land combined with them; at some point in this process the output of the enterprise will begin to increase at a rate less than proportionate to the additions of labor and capital. Or if the abor supply is the fixed factor and capital and land the variables, the same phenomenon will appear: sooner or later successive additions to the increasing factors will cause less than proportionate increases in the total output. This law of diminishing returns is of utmost importance in determining economic rent in situations where land is the fixed factor.

It follows that if the lands in society are being exploited at any time beyond the point of diminishing returns, labor and capital cannot be withdrawn from an area of land and concentrated on other areas without diminishing the social income. In the new location these productive agents must be added to the labor and capital already at work and they will return diminished marginal products — in other words, a smaller product than formerly. In this situation a given area

of land will have power to add to the social income, and its owner can exact this quantum of income in exchange for the use of his land. If the point of diminishing returns has not been reached, the withdrawal of one unit of land will not affect the social income. But under this condition there would be no economic rent on any land. It is due to the operation of diminishing returns that economic rent exists as a distinct type of income. The principle of diminishing returns applies to all types of land — urban sites, mines, forests, etc.; but for purposes of further analysis we shall continue with our simple example from agriculture.

Market measurement of economic rent. When we come to measure in dollars and cents the amount of economic rent accruing from a given area of land, we have to do with the operations of the market for goods and services. All incomes are distributed in the process of normal business dealings and take the form of money receipts drawn from the market. If any individual gets an income merely as an owner of land, in addition to normal returns on his labor and capital, it must be because of an automatic response of the market to certain conditions of demand and supply. We shall now inquire how this happens, in the course of the inquiry demonstrating again the existence of economic rent as well as discovering how the size of this income is determined.

As a first step we must see that diminishing returns is only another name for increasing marginal cost. Returning to our example of the 25-acre farm producing increasing quantities of corn by means of increased applications of labor and capital, let us translate this expanding working force into terms of monetary cost. A compensation must be allowed the working proprietor sufficient to repay him for his efforts; the cost must also include a normal interest return on money invested in capital instruments; and expenditures for labor, seed, and other materials must be added. In short all expenditures required to produce the grain must be accounted for in our estimate of costs. We shall assume that these costs of produc-

tion can be apportioned equally among the workers employed in cultivating the farm.

INCREASING MARGINAL COST IN AGRICULTURE

Number of workers	*Expenditure in dollars*	*Total product in bushels*	*Marginal product in bushels*	*Marginal cost per bushel*
1	$100	500	500	$.20
2	200	1,000	500	.20
3	300	1,400	400	.25
4	400	1,700	300	.33⅓
5	500	1,900	200	.50

An equal expenditure for each worker — $100 per man — resolves the increasing labor force (column 1) into terms of total monetary cost (column 2). The third column shows the increase of total product; the fourth column, the marginal product, diminishing after the addition of the third man; the final column translates the diminishing marginal product into increasing marginal cost. The first marginal product is 500 bushels; its total cost is $100; its marginal cost per bushel is 20 cents. The marginal cost when two men are employed is the same as before, since the marginal product remains unchanged. Diminishing returns set in with the addition of the third man, and the marginal cost rises to 25 cents, because the third man adds only 400 bushels to the total product with an expenditure of $100. Marginal cost rises to 33⅓ cents when production is expanded by the addition of a fourth man and to 50 cents when a fifth man is employed.

In our study of value we learned that goods produced under conditions of increasing cost tend to sell for a price equal to their marginal cost. In the case of any given category of goods identical units will of course command the same price. The price of each bushel of a certain grade of corn in a given market at a given time will be the same as that of every other bushel; and so it is with cotton, wool, wheat, and other products. If the production of these things is governed by in-

creasing marginal costs, all units will tend to command the same price as that paid for the most costly unit exchanged in the market at the time in question. These principles of economics are pertinent to the problem of economic rent.

Consider now the relations of our farmer with the market. He will be governed as to the extent of his operations by the expected price of corn. He will produce 1,900 bushels only if he expects the price to be at least 50 cents; or 1,700 bushels for a price of $33\frac{1}{3}$ cents; or any other quantity only if the price expected is high enough to cover the marginal cost of this quantity. The producer of any increasing cost good reacts to price change by expanding his output when prices rise and reducing it when prices fall, tending at all times to equate his marginal cost to the price of his product. These are normal, business-like operations; they result in giving our farmer a surplus income above costs after production on his land has passed the point of diminishing returns. Note that when he produces 1,700 bushels of corn at a total cost of $400 and sells his product at $33\frac{1}{3}$ cents a bushel his total receipts are $566.66, or $166.66 in excess of his costs. An output of 1,900 bushels selling for 50 cents a bushel and costing $500 will bring in $450 in excess of cost.

Producer's surplus. This excess of income above costs is economic rent in the guise of *producer's surplus.* It is a true surplus over and above all expenditures required to produce the quantity of grain offered for sale. In other words, it is an income distinct from wages, interest, and profits, since these latter incomes were all accounted for in the allowances made for cost of production. Our illustration deals of course with hypothetical conditions and may for that reason appear unconvincing as a proof of the existence of rent. But close examination of the circumstances under which an increasing cost good is produced and sold will disclose the existence of this surplus income. Taking care to set down under cost all items of expense necessarily involved in the process of production — including a normal reward for the entrepreneur — money

receipts must exceed this total expenditure after marginal costs have begun to rise. If our farmer's product sold for no more than enough to cover total cost — including a normal reward for himself — this income would be sufficient to keep his farm in operation. The fact that it sells for more than this is the result of the operation of diminishing returns and an expected market price high enough to warrant cultivation beyond the point where diminishing returns set in. Thus a surplus income is inevitably derived under normal market conditions from the cultivation of the soil after production has been pressed beyond the point of diminishing returns.

Explicit and implicit economic rent. In the preceding illustration the economic rent accrued to the farmer as an implicit income, a part of his general money receipts from the market. But it came to him in his character as an owner of land, not as a worker or manager or owner of capital instruments, for he received in addition a normal reward for these other functions.

Suppose now that our farmer wishes to retire from active labor and so offers his land for lease. He will of course expect his earnings as a worker to cease. We may assume that he turns his supply of capital instruments into cash and invests the proceeds in good bonds, thus perpetuating that part of his income which consisted of interest. In these circumstances his real loss of income upon retiring from the land will be equal to the producer's surplus or economic rent which he has been receiving; this is what the land is worth to him as income-bearing wealth, and he will ask a contract rental of this amount from any prospective tenant. The tenant will be able to pay a rental of this amount; for after subtracting this rental from his gross income he will have enough left to cover all costs of production and still retain for himself a normal reward for his labor and management. Under competition therefore contract rent will tend to equal the surplus gain to the owner implicit in his cultivation of the soil. When the land has been leased on these terms, the owner will occupy a

peculiar position as a sharer in the social income. A share will accrue to him merely as a landowner quite in addition to any income he receives as a worker or supplier of capital. In a well-developed society a multitude of these lease contracts exist, and many people draw incomes from them. In this group belongs also the large number of people who are drawing implicit economic rent from their own exploitation of land. Here is economic rent presented in the broad social sense as a distinct functional type of income.

Lands of different quality. We have dealt thus far with a single plot of ground of uniform quality. If the earth's arable land were all of one quality, its cultivation would give rise to economic rent as soon as consumers' demand for the product had so stimulated cultivation that a part of the supply must be produced under conditions of rising cost. But lands, as we know, vary in productivity because of differences in natural fertility, in proximity to the market, in access to transportation facilities, and so on. In business practice these differences in quality take the form of larger costs of production on some lands than on others. Whatever the crop in question, there are some lands whose combination of location and soil properties enables them to return the largest receipts for a given expenditure; others are somewhat less productive of income from the same expenditure; others still less productive. This is true of different parts of a single farm. Exhaustive analysis of the entire land area would permit its classification into many varying grades of productivity. The principle of diminishing returns will apply to the exploitation of each of these grades of land, and each will return an economic rent if certain conditions obtain with respect to the relation of marginal cost to market price.

A simple illustration will make these relationships clear. We will take corn lands of three grades, their relative quality being shown by differences in their initial marginal costs of production. Grade A is the farm used in our previous illustration. The inferiority of Grade B is shown by the higher

marginal cost of the first units of output. Grade C is inferior to Grade B for the same reason.

GRADE A SOIL

Expenditure	*Total product*	*Marginal product*	*Marginal cost*
$100	500	500	$ .20
200	1,000	500	.20
300	1,400	400	.25
400	1,700	300	.33⅓
500	1,900	200	.50

GRADE B SOIL

Expenditure	*Total product*	*Marginal product*	*Marginal cost*
$100	400	400	$ .25
200	700	300	.33⅓
300	900	200	.50
400	1,000	100	1.00

GRADE C SOIL

Expenditure	*Total product*	*Marginal product*	*Marginal cost*
$100	300	300	$ .33⅓
200	500	200	.50
300	600	100	1.00

It is obvious that unless the price of corn is higher than 20 cents a bushel only Grade A soil will be cultivated, since marginal costs on the other lands are above this price. At 20 cents a bushel there will be no rent even on the best land, since the quantity produced at this price would return no producer's surplus. Moreover it is clear that except for diminishing returns the price would never rise above 20 cents; inferior soils would never be cultivated; no economic rent would ever appear. These statements hold good because in the absence of diminishing returns unlimited quantities of corn could be produced at the lowest cost shown in the table — 20 cents a bushel. Let us assume now that the little society pictured here experiences an increase in the demand for corn because of an expansion of population. This greater demand, together with the appearance of diminishing returns when Grade A soils

are called on for greater production, will eventually cause the price to rise — let us say to 25 cents. At this price cultivation will begin on Grade B soil, and economic rent will appear on Grade A; Grade C land will still be idle. If the price rise continues until corn sells for 33⅓ cents, economic rent will appear on soil of the second quality; economic rent will rise on Grade A land; and the poorest of the three grades of soil will be drawn under cultivation. The price must rise to 50 cents a bushel before this third grade of land returns an economic rent.

Margin of cultivation. The foregoing analysis introduces a concept of importance in the doctrine of economic rent; the concept of the *margin of cultivation*, which may be defined as *that stage of exploitation at which a unit of expenditure just pays for itself.* This last, or marginal, expenditure is applied under such conditions that it produces an income only equal to the cost. That this is true may be due to either of two causes: the expenditure may be the last of a series applied to a given plot of ground, its product being restricted by the law of diminishing returns; or it may be the first expenditure made on lands so lacking in fertility or so remote from the market that, at the existing price of the product, they are barely worth cultivating at all. It is useful to distinguish these two cases by calling the one the *intensive*, and the other the *extensive, margin of cultivation.* The intensive margin of cultivation on any grade of land is reached when the product of the final unit of expenditure sells for no more than it costs to produce. The extensive margin appears on that grade of soil which it is just profitable to cultivate at a given time. Referring again to our illustration, with corn selling at 33⅓ cents, Grade C soil lies on the extensive margin; the intensive margin is represented by the second expenditure on Grade B and the fourth on Grade A.

It is important to note that no producer's surplus is derived from the expenditures at either the extensive or the intensive margin. This is because marginal cost equals market price. The illustration makes it clear that the position of both mar-

gins at any given time depends upon the selling price of the product. If anything occurs to raise this price, the intensive margin on all lands under cultivation tends to be raised and the extensive margin to be pushed out to lands whose lack of productivity made them extra-marginal at the old price. A fall in the price tends to cause a lowering of the intensive margin on all lands and a narrowing of the extensive margin by the process of throwing the poorest lands out of use. It should be noted that these statements hold good only on the assumption that costs of production remain unchanged as the selling price fluctuates.

Alternative uses of land: Agrarian lands. It is obvious that we have simplified the problem by assuming that only one crop is produced on the land. But it is really unnecessary to consider all the complexities caused by the vast number of different crops actually under cultivation at a given time, for this situation does not alter the principles of rent. For each of these uses the different parts of the land area vary in profitableness because of differences in fertility, in location, etc. Under given conditions of cost and price there will be for each piece of land one most profitable crop and an intensive margin measuring the largest amount of that crop which it pays to produce. Each farmer will try to employ his lands so as to expand his producer's surplus to the maximum. He will plant that crop which at the time offers the largest producer's surplus and will continue this line of production until some other crop becomes more profitable. In this way the existing area of agrarian lands is allocated among the different lines of cultivation so as to produce the largest economic rent under prevailing conditions of cost and market price.

Urban lands. In the city, as in the country, we find lands devoted to different uses and unequally productive. Different sites vary in profitableness from the point of view of any type of business that we might choose to consider. The retail stores, the theatres and hotels, the office buildings and apartment houses, the banks and wholesale establishments in any city of

considerable size occupy sites which, even to the most casual observer, differ in respect of their adaptability to the purposes of the business. In most cities, this difference is so noticeable that it is not difficult to point out the enterpriser who has surpassed his competitors in the scramble for desirable locations. It is evident that each of these leaders in his own field would sacrifice income if he were forced to move to a less desirable location, even if he applied his labor and capital as abundantly and efficiently as before. There is no way to explain these differences in the income-producing powers of enterprises differently situated, when managerial ability and other factors are equal, except to attribute them to the relative productivity of the lands on which they stand.

The principle of diminishing returns operates on each plot of urban land whatever the use to which it is put. For example in retail trade an increased expenditure of capital and labor on a site of given dimensions takes the form of additional stories to the building, a larger area of floor space, an increased stock in trade, increased working force, a more complex and unwieldy organization. As the building grows in size, the costs of maintenance, of transportation of goods and passengers from floor to floor, of supervision, will eventually rise. Many small sources of waste will pass unnoticed by the managers which in a smaller establishment would be discovered and corrected. For many reasons the costs of rendering the service will rise, or to use the correlative expression, the returns from the last unit of expenditure will fall. Hence the time must come when additional expenditure on the best sites no longer affords a larger profit than the same expenditure on lands not so well situated.

The exploitation of each parcel of land tends to be pushed to the limit — typified by an added story, department, operating device — which it is worth while to operate at a given time on any particular site of given size. This is the intensive margin. The extensive margin of land use is not to be found in retail trading or in any other use of city land. There is

usually no no-rent land within the limits of the ordinary city. What occurs is that the business in question, as it expands, pushes constantly out from the focus of the most desirable land, taking in locations successively less desirable till it reaches those parcels of land which are no more useful for this business than for other uses. Beyond this the business cannot at any given time go, because there now appear other uses for the land which will yield a higher rent, this rent being determined by the net income to be derived from some other use of the land and resting finally upon the prices obtainable for the services of the land in this other — most productive — use. There is equality of return between the intensive margins on all the different parcels of land devoted, at any given time, to the particular business, and as conditions change this equality tends to maintain itself. With increased demand for retail services, for example, the intensive margin is raised by more complete — and more costly — utilization of sites already in use, while at the same time the business takes over certain sites which were formerly more profitably devoted to other uses but which the increased returns of the retail business now make it possible to wrest away through the offer of higher rents.

When economic rent from urban lands is an implicit income — that is, when the owner of the land is also its user — the problem of estimating its amount on a given site is extremely difficult. By comparing one site with another, as we have done, we can assure ourselves that rent exists, but in any given case it is so inextricably merged in the total net income which the owner derives from the operation of his business that he himself is not able to mark it off to the last dollar. In actual practice however its appraisal is not left to the owner to be worked out by an analysis of his money receipts. In every city of average industrial activity the amounts of economic rent on various grades of land are measured through the activities of speculators in real estate and the competition of tenants for the use of land they do not own. These activities result in a sched-

ule of explicit rents, or contract payments for the use of land, which very closely approximate the amounts of producer's surplus which the different plots of ground would return if exploited by their owners.

The value of land. It is the specific income from land — the economic rent — which together with the interest rate determines land values. At an earlier point in our study we considered the discount process as it applies to the valuation of durable capital instruments. It was there observed that, whereas the demand schedules for capital goods are always derived by discounting expected future income, cost of production operating on the supply side is also a factor in the value of such categories of these goods as are reproducible. In the case of capital instruments which cannot be reproduced, it is the discounting of future income which alone operates on both demand and supply to determine value. With trifling exceptions of insufficient moment to affect the general rule, land is to be classified with this group of goods. Its value varies directly with economic rent and inversely with the interest rate. It is of course understood that if government takes a part of the economic rent through taxation it is the net economic rent remaining to the landlord that governs the value of the land. This is not to say that the sale price of any piece of land is determined precisely by the present rent capitalized at the present rate of interest. Considerations as to the probable future rise or fall of rent and the rate of interest always enter into these calculations of the present value of future income; but alterations of land values resulting from these considerations are all in harmony with the basic principle, which may be stated as follows: *The value of land (and other natural agents) is derived, through demand and supply, from buyers' and sellers' estimates of the discounted value of expected economic rent.*

Relation of rent to prices. There is a good deal of popular confusion regarding the causal relation existing between the rent of land and the sale price of the product of the land, it

being commonly believed that high rents cause high prices. But this popular notion confuses cause and effect; it is the price of the product which affects the economic rent of land, not the reverse. In the case of rural lands the true relationship between rent and the price of the product becomes quite clear when we review our previous analysis. The price-determining costs of production in agriculture are the marginal costs. But on the margins, as shown in our preceding discussion, cost is equated with price, allowing no producer's surplus or economic rent from the sale of these marginal units of supply. In other words, economic rent is not included among the price-determining marginal costs. Surplus receipts are obtained from the sale of the units of supply other than marginal because the existing market price makes production profitable beyond the point of diminishing returns. If rents are high at any given time it is because prices are high.

As to urban lands, the fallacious belief that high rents cause high prices may be traced to a misconception of the nature of the product of these lands. The product of land devoted to retail trading is obviously not the articles offered for sale in the shops. It is the service rendered by the merchant to the social group by promoting the ease and convenience of supplying human wants. If we wish to speak of the price of this service, it should be clear that we refer to the reward of the merchant, not the sale price of his stock in trade. Any attributes of a site which increase the merchant's gain on a given expenditure of capital and labor in the retail trade beyond the amount received by the users of marginal land will bestow a rent on the site. In some cases a given site will increase the merchant's profit for no other reason than that it places him in a position to overcharge his patrons. Here there is a relation between rent and the selling price of goods, but the causal connection runs from price to rent, not from rent to price. More frequently it is the effectiveness of the site in reducing operating costs which increases the business man's net return. Among mercantile enterprises large volume of

turnover in comparison with necessary overhead expense is the primary cause of a relatively low cost of operation. Sites on crowded thoroughfares and at the transfer points of traction lines permit a volume of business so much in excess of that done by shops in less effective locations that the merchants who possess them can gain larger returns at prices no higher than those charged by their competitors. In the one case, high rents will be conjoined with high prices; in the other case with low sales prices for goods; but in both the high rent is itself caused by those attributes of the site which enable its possessor to derive a large income from its exploitation. These illustrations have been drawn from retail trading, but the reader will have no difficulty in applying the principle they embody to lands in other uses.

Quasi rent. It will be recalled that the principle of diminishing returns, as we have stated it, is applicable under appropriate conditions to the exploitation of any factor of production. The condition essential to diminishing returns is that one of the factors be held constant while those which cooperate with it are increased. As the variable factors increase in amount the output of the combination, while increasing absolutely, will diminish in proportion to the increasing factors. As we shall see in subsequent chapters this principle is of great importance in determining the other functional types of income — interest, wages, and profits — as well as rent.

We should note at this point that whenever a situation arises in which either labor or capital is a fixed factor, with the other factors variable, its income will be of the same nature as the economic rent of land. A labor group, if it is both fixed in number and indispensable to productive enterprise, will derive a producer's surplus determined by forces analogous to those discussed in the preceding pages. Similarly the owner of capital instruments, if their supply cannot be increased and they are in great demand for productive purposes, can exact a payment of this kind for their use. Ordinarily the competitive forces of the market fix the rates of

return for labor and capital in a quite different way, because ordinarily and in the long run the supplies of these factors can be and are increased in amount. It should be carried in mind however that when a segment of the labor or capital supply is placed in favoring circumstances — limited in amount and combined with increasing supplies of the other factors — its income will diverge from the returns normally allocated to that factor through the market and partake of the nature of economic rent. To such incomes, secured by factors of production other than land, it is customary to give the name *quasi rent.*

Economic rent distinguished from other income. The condition of fixed supply is much more important in determining the income of land than that of the other factors. Every productive factor, other than land, can be increased in amount over long-run periods, either in its entirety or in its special forms. Hence the owners of these other factors of production can enjoy surplus incomes from their use only in exceptional circumstances and for short-run periods. But the income from land is normally a surplus income — producer's surplus.

Let us illustrate this distinction by comparing the income from land with the income from capital in a selected situation. Suppose that an office building is erected at a cost of $100,000 upon a certain city lot. A year's operation shows the following results: total income, $15,000; operating costs, $5,000; interest on capital (at 5 per cent), $5,000; net income, $5,000. This last item is true producer's surplus; it is, in fact, economic rent, being due solely to location on a superior piece of land. Since the number of these superior sites cannot be increased the surplus income will tend to persist.

Suppose now that a shrewd enterpriser builds, at a cost of $100,000, a ship to carry passengers from the city to a nearby shore resort. The results of his first year's operations are exactly the same as those of the office building; *i.e.*, total income, $15,000; operating costs, $5,000; interest on capital,

$5,000; net income, $5,000. Here also is a surplus income, the reward of superior business foresight. But it does not rest on the firm foundation of economic rent. Very soon other enterprisers, seeing the extra profits offered, will build other ships. Competition will reduce the rates charged passengers; the gross income per ship will decline in consequence; and finally there will accrue to the owner only the $10,000 necessary to cover operating costs and interest on his capital investment. Thus does competition tend to make short-lived the surplus gain arising on occasion from shrewd applications of capital.

"Unearned increment." Because of its character as a surplus receipt, economic rent by private individuals is peculiarly hard to justify on grounds consonant with the common man's impulsive sense of justice and fair play. Economic rent is not a payment for any service rendered by the recipient, for it exists only when the recipient's income is in excess of adequate remuneration for all his labor as a worker, all his service as a saver and investor of capital, all his utility to society as an enterpriser. A service, it is true, is rendered in exchange for economic rent, but it is rendered by nature, not by the rent receiver. In other words, society, in permitting the existence of economic rent as a private income, is making payment for a service which could be had for nothing.

For this reason economic rent has often been described — and condemned — as an unearned increment. As the words imply, this name refers particularly to the tendency of rent to increase as population grows. But there is no reason in principle why distinctions should be drawn between different cases of economic rent. Whether rising, falling, or remaining constant, all true economic rent is, in the common meaning of the term, "unearned" by the landowner; it is an addition or increment to his income in excess of adequate payment for every productive service he performs.

It is of course true that the present recipient of economic rent may have invested his savings in the purchase of the land

and is therefore receiving only a justifiable return on his investment. For example assume that someone pays $40,000 for a site which returns an economic rent of $2,000. Clearly there will be no unearned increment in the income of this individual — at least until the economic rent increases — for the yearly rent represents only the normal return on his financial outlay. But the former owner is still receiving $2,000 yearly from the investment of his receipts from the sale of the land, and if we agree that part or all of this income was unearned increment before the sale, it is still unearned so far as he is concerned.

Rebelling against the apparent injustice of the economic rent income, people with a bent toward social reform have frequently proposed that this income be taken out of private hands. These programs of reform have taken many different forms, from outright nationalization of land, as an extreme example, to milder proposals to appropriate a part or all of the economic rent for social purposes. The method suggested for such appropriation of rent is usually taxation; and this proposal too has sometimes taken the extreme form — as in the case of the single tax movement — of demanding that the state tax all landowners the full amount of the economic rent. The condemnation of rent as a phase of their general condemnation of all inequalities in the distribution of income is an element in the reform programs of the socialists, whose proposals we shall consider in a later chapter.

EXERCISES

1. The figures below show the results of successive added expenditures on different grades of land devoted to the production of wheat:

Grade A Land		*Grade B Land*		*Grade C Land*	
Added Expenditure	Added Product	Added Expenditure	Added Product	Added Expenditure	Added Product
$100	200 bu.	$100	100 bu.	$100	80 bu.
100	150	100	90	100	70
100	125	100	80	100	60
100	100	100	70	100	50
100	80	100	60		
100	60	100	50		

(*a*) Convert the figures for diminishing physical product with additional outlays on different grades of land into figures expressing rising marginal cost.

(*b*) Plot the curves of marginal cost for the production of wheat on the three grades of land (on one chart).

(*c*) How far would cultivation be carried on each grade of land, if the market price of wheat were $1.00 per bushel? If the market price were $1.25 per bushel?

(*d*) Locate the intensive and extensive margins of cultivation when the price of wheat is $1.00 per bushel; when the price is $1.25.

2. A piece of land occupied by a warehouse is leased for $12,000 per year. Taxes, insurance, maintenance, and depreciation consume $5,000 per year. The building cost $100,000. The rate of interest is 5 per cent. Explain the two elements of which the owner's net income is composed, and determine the amount of each.

3. With the construction of the Erie Canal, the price of wheat at New York City declined. How did this affect the intensive and extensive margins of cultivation in New York and New England and the rent of land devoted to wheat production in this area?

4. Explain how each of the following factors may influence the use of urban sites:

(*a*) Convergence of subway lines.

(*b*) Proximity to schools.

(*c*) Easy access to railroad terminals.

(*d*) Freedom from noise.

(*e*) Proximity to tidewater.

(*f*) Migration of population to suburban areas.

5. (*a*) How would the movement of the New York Stock Exchange to New Jersey affect rents in the New York financial district?

(*b*) How would the replacement of an elevated street car line by an underground transit line affect rents and the use of urban sites along the route?

(*c*) How would the development of the St. Lawrence Waterway affect urban rents in Boston? in Chicago? in Baltimore?

CHAPTER XVI

Interest: The Income of Owners of Capital

Time preference and capital formation. It is an attribute of human nature to be impatient for present enjoyment. Man postpones enjoyment only with reluctance and with a sense of sacrifice. Today's goods are in general preferred to future goods of like kind and amount. Conversely one naturally feels a distinct sense of gain when, having reconciled himself to waiting for a pleasure which has been promised in the future, he is enabled to abbreviate the period of waiting. Although the pleasure is the same in magnitude, with its nearer approach it grows in desirability and is more highly prized. This being true, identical goods when separated by intervals of time do not have identical present desirabilities. To this human trait is given the name *time preference*, defined as *the inclination to place a higher estimate upon present goods than upon identical goods in the future.*

In the case of every capital instrument there is an inevitable waiting period before costs are recovered through sale of products. The first income from a new railroad is sometimes secured long years after raw materials were bought for its roadbed, terminals, and rolling stock. With other capital instruments — *e.g.*, simple tools — these temporal considerations may be of slight importance; but the time factor in supplying capital instruments is never entirely absent. Capital formation therefore always involves the expending of present wealth and income in consideration of a future return. If capital goods are to be produced, there must be remuneration, not only for the ordinary costs of production, but also for the cost of waiting for deferred returns. In other words, it takes additional compensation to overcome time preference. No man will voluntarily defer present consumption unless the

future return appears at the present moment at least equal in desirability. But since a dollar on hand is more potent in its present appeal than a dollar a year hence, the two will not be exchanged unless a premium is added to the future dollar. This surplus or premium is interest — the payment for waiting. In view of the natural irksomeness of waiting, this payment is a necessary stimulus to the accumulation of an adequate stock of capital instruments in any community, nation, or region where such accumulation is entirely or largely voluntary.

Nature of the loan contract. In the case of many economic activities which involve waiting the interest element does not appear as a separate and distinguishable payment. When a capital instrument is used by the owner, the interest element, though present, may be lost from sight in the composite income stream. It is in connection with business transactions in which capital instruments or their value in money are loaned that interest appears in a form subject to measurement.

When we analyze the ordinary loan or investment contract we discover that it represents a transaction in which present and future incomes are exchanged against each other. At the moment the loan is made the borrower acquires an addition to his present income, while the lender suffers a reduction of present income. When the loan is repaid, the reverse effect is exerted upon the incomes of the two parties. Taking the transaction as a whole therefore it is seen that the lender has exchanged a part of his present income for income in the future and the borrower has made the opposite exchange. This is true of all loans regardless of their purpose. The interest element in the loan contract is a premium paid the lender to compensate for his waiting, paid by the borrower who willingly sacrifices a larger amount of future income to acquire the more highly prized present income.

The loan market and the rate of interest: Definition. The rate of interest is normally expressed as a percentage, the ratio of the amount of one year's interest in terms of

dollars to the amount of capital also in terms of dollars. Thus a borrower receiving $1,000 may contract to repay the same amount one year hence and in addition to pay $50 interest. The price of the loan — or the rate of interest — is $\frac{50}{1000}$ or 5 per cent per annum.

Like other purely competitive prices, the rate of interest is determined in the market by the forces of demand and supply. The modern world has evolved an organized market, in which exchanges of present goods for future goods are constantly taking place; *viz.*, the loan or investment market. Because of the vast number of transactions which take place in the loan market and the innumerable ramifications of this market through the business world, the rate of interest there determined becomes the governor of the terms on which all types of transactions involving an exchange of present for future incomes occur. Further inquiry into the problem of the rate of interest necessitates an analysis of the circumstances affecting the demand for and the supply of loan and investment funds. We begin with the supply side of the market.

The supply of loan funds: Reactions of lenders. Lenders (investors) react in general to the rate of interest as the suppliers of any other service or commodity react to its price. Amounts available for loans conform with the general law of supply. Under any given set of circumstances they vary directly with the rate of interest. The supply of loans however is subject to certain special peculiarities. Many people would postpone the enjoyment of some part of their wealth even though they received no interest. Men will make provision for old age, for the care of their children and other dependents, for unpredictable contingencies, such as accident, sickness, or premature death. Such savers accept interest on personal loans, savings bank deposits, mortgages, bonds, etc.; but even if these opportunities were not open to them, they would still save out of present income. Many well-to-do people habitually carry checking accounts at the banks in excess of their needs for the sake of the security and ease which a large

balance affords them. Many small depositors likewise continue to maintain their accounts even though service charges are levied against them. The balances carried in commercial banks by both these classes of depositors contribute to the basis of loans to business enterprise. With many people saving is a matter of habit determined by the customs of the social group rather than by cold calculation. It is evident therefore that some postponement of present consumption and some consequent formation of capital would occur, not only if no interest were paid, but even if a charge were made for the privilege of saving. In fact to some extent such waiting actually does take place.

Corporate saving. It is important to note also that certain developments in our economic organization are tending to compel men to save, either against their will or without their knowledge, and thus to create a supply of loan funds without much reference to the interest rate. One important tendency of this order is an outgrowth of the corporate form of organization, especially in the large-scale enterprises so common in our modern business world. Immediate control of the policy of a corporation lies in the hands of the directors, who are in turn responsible for their actions to the stockholders who elect them. When the corporation has grown to immense size and its stock has become widely scattered among a large number of small investors whose knowledge of the business is vague, the directors are virtually given a free hand in shaping the policy of the undertaking. One of the problems which they decide relates to the disposition of the corporation's net income. This of course constitutes interest and profits upon advances made by the stockholders. But it is seldom the practice of prudent directors to distribute the net income in the form of dividends as rapidly as it accrues. A part is usually set aside as a reserve against future contingencies which may imperil the corporation's ability to pay its debts; another part may be devoted to improving and expanding the material equipment of the business; still another to experimentation

with new processes and selling methods. In thus disposing of the corporation's income, the directors make savings which possibly would not be made by the stockholders if all corporate earnings were currently distributed in dividends. An examination of the capital accounts of large corporations over a period of years will bring to light immense annual additions to capital equipment resulting from this practice of "ploughing back into the business" the earnings of corporations.

The activities of the insurance companies afford similar examples of corporate saving. The "level premium policy," the customary form in which life insurance is sold, results in the accumulation of immense funds in the hands of the insurance companies as reserves against their liabilities to the policy holders. There is little reason to believe that the average policy holder adjusts the amount of his contribution to the reserves of insurance companies by scrutinizing the rate of interest. Furthermore the practice of making the carrying of insurance compulsory, deducting the premium from the income of the policy holder without his knowledge or consent, as for example in the case of legally enforced accident, sickness, unemployment, and other forms of social insurance, is spreading among industrial nations. This movement is tending to increase the proportion of society's investment fund whose accumulation bears no direct relation to the interest rate.

Other examples could be cited. Loanable funds arising from such sources as these are added at any given time to those resulting from voluntary choice of individuals and tend to increase the amount offered at various rates of interest. We conclude therefore that, although the total amount of funds available for loans at any given time is in some degree responsive to changes in the interest rate, nevertheless this total includes important inelastic elements, which are largely independent of the rate of interest. We take account of this peculiarity of the supply of loan funds in our graphical representation of the loan market on page 276.

The demand for loans: Borrowing for consumption. Two distinct types of borrowing create the demand for loans; *viz.*, *borrowing for consumption* and *borrowing for production.* Borrowing for consumption arises among those individuals whose time preference is so high in relation to the interest rate that they not only refrain from lending, but also pledge anticipated income for the furtherance of present consumption. This attitude among the poor brings them to pawnshops and among the middle classes stimulates purchases on the installment plan. Certain types of banks, for example the Morris plan banks, cater to this type of borrowing. Consumption loans are by no means limited to people with restricted means. Well-to-do persons, anticipating larger incomes in the future than at present, are inclined to borrow for present enjoyment. Anyone whose income is irregular may finance personal expenditures by occasional short loans. Government loans for productive investment are not unknown, but most of the borrowing of political bodies is for purposes of consumption.

Consumption borrowing determined by time preference. The determining factor in the case of the borrower for consumption is his time preference. A high rate of time preference leads to eagerness to borrow. The various conditions of individual character or of external circumstances which lead anyone to have a high rate of time preference may therefore be considered as influences impelling him to borrow. So long as the borrower's rate of time preference is higher than the rate of interest which he has to pay for loans there is gain to him from borrowing. But borrowing for consumption has the effect of enriching the present and impoverishing the future. In accordance with the general law of diminishing marginal utility, each unit of borrowing adds a progressively smaller quantum of satisfaction to the present income and entails a progressively larger future sacrifice. The marginal utility of the present gain decreases, while the marginal utility of the future income sacrificed increases. It follows that the individual's rate of time preference, based as it is on the com-

parison of these two marginal utilities, tends to decline as he borrows, until eventually it becomes equal to the rate of interest. Thereafter borrowing ceases to be advantageous. Hence the amount an individual is willing to borrow for consumption at any given time tends to vary inversely with the rate of interest.

Borrowing by governments. Although government loans are for the most part similar in nature to the consumption loans of private persons, their amount is not determined by the same forces which control the action of the individual with reference to the loan market. The amount of loan funds demanded by political bodies does not vary in any predictable way with the interest rate. A high rate of interest will sometimes cause the postponement of some government enterprise for which borrowing would be necessary; it is more likely to affect the tenure of the loan; *i.e.*, whether it runs for a long or a short term. In time of war, or other national emergency, the statesmen in charge of the nation's destinies simply borrow all that is needed to bridge the gap between tax receipts and expenditures, or all that the lenders can be induced to provide for this purpose, paying the lowest rate of interest that is practicable.

Borrowing for production: Purpose of the borrower. Unlike the spendthrift, the business man's loans are expected "to pay for themselves" and in addition to confer a gain upon the borrower. It is the purpose of the business man to exchange the borrowed purchasing power for certain forms of wealth and services to be productively employed in his enterprise. If we follow any one of the possible lines of expenditure to its conclusion, we discern that the loan has been employed by the enterpriser to divert a portion of the productive energy of society from the direct into the roundabout process of production. If additions to plant are constructed under the enterpriser's own management, a portion of the borrowed money is transferred to laborers in the form of wages on which to subsist while at work on a non-consumable product; another

portion, expended in the purchase of materials, goes to reimburse someone else for a similar transfer of purchasing power to laborers. If these capital instruments are bought ready-made, the expenditure of the borrowed money gives support to those industries in which labor and capital are devoted to improving the preliminary stages of the process of production. Essentially similar is the result of a loan when it is used to buy intangible services rather than concrete articles of wealth. In brief all funds borrowed for productive purposes are ultimately resolved into advances to laborers.

The productivity of capital. The reader will recall that labor applied in the roundabout way is generally more productive of future income than labor applied to direct production. This roundabout process, to be sure, may not actually prove more productive in every case; labor may be wasted in overelaboration of the intermediate stages of production. But the possibility of net gain from the capitalistic process exists, and this possibility is increased by the dynamic character of our economic order. The borrowing of enterprisers rests on this fact of the productivity of capital.

But the enterpriser will ask something more of his borrowed capital than that it increase the physical product of his enterprise. He is not consciously concerned with the broad social aspect of the use of capital; what he is after is a money income for himself, to be derived from the sale of the products of his enterprise. He will borrow only if he expects to sell the increased product of his enterprise at prices high enough to enable him to repay with interest the money expended in the purchase or construction of his capital equipment.

Now the productivity of capital conforms to a law of diminishing marginal productivity, whether viewed from the standpoint of society or from that of the individual business man. In taking the social point of view, we must define the productiveness of capital as its power to increase the output of goods per unit of labor expended. If only a small quantity

of capital is available for the use of society, it will be employed to equip men with the most essential labor-saving devices; as the quantity of capital increases, other things remaining constant, it will be applied to uses which grow less and less vital to the material welfare of mankind, until the last installments of the capital supply of a rich society take the form of trivial improvements on a highly complex machine equipment.

But the business man is more immediately concerned with a parallel tendency in the returns from his own enterprise. Whatever the form of capital instrument or productive service to which he devotes his own or borrowed purchasing power, he finds each one, as its quantity increases, becomes in time subject to a law of diminishing productivity in terms of money income. Though at each step in the expansion of his enterprise he hits upon the most efficient combination of productive elements, nevertheless beyond a certain point each successive increment of expenditure yields a diminishing monetary return. The ordinary business man, subject to conditions of imperfect competition, must, in order to dispose of more product, eventually lower his price. This lowering of price may at first increase his money income by increasing the physical volume of his operations and may thus compensate him for greater outlay in the use of capital, but the continued lowering of price necessary for further increases in volume of operations and further utilization of capital will bring smaller and smaller increases in money income. The business man borrows in the first instance because of his conviction that the net monetary return will exceed the interest rate. Diminishing marginal productivity of capital tends to remove this incentive as his borrowing increases. Accordingly the individual demand of the enterpriser-borrower, like that of the borrower for consumption, varies inversely with the interest rate. The business man tends to continue his demand for loans until the marginal productivity of the capital thus obtained is equal to the interest rate paid.

Total demand for loans. The demand of enterpriser-borrowers and individual consumer-borrowers may all be thrown into the form of schedules and combined to form a total demand. Government borrowing is not easily reduced to a demand schedule because non-economic motives weigh heavily in all political activity. The government comes into the market as occasion requires with a predetermined demand for loans and offers the lowest possible rate of interest. Though there are times when the necessitous borrowing of governments is controlling, as a rule the demand of enterprisers so vastly outweighs all other types of borrowing that it governs the market. Borrowers for consumption and governments in ordinary peace times must adjust their demands to a loan market dominated by business men, whose borrowing is based on the productivity of capital. Strictly speaking however all three types of demand must be added together to give a measure of the forces operating at any given time on the demand side of the market.

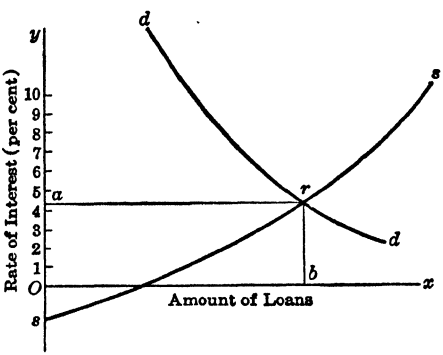


FIG. 22

Determination of the market rate of interest. As already noted, the rate of interest is a price — the price of waiting — determined in the loan market. Inasmuch as the elements constituting the total demand conform with the general law of demand, so does the total demand itself. It may be represented therefore by a curve such as *dd* in Figure 22, sloping downward to the right. That part of the supply curve *ss* in this diagram which lies below the base line illustrates the fact that a few lenders would, if they had to, lend without interest, some of them even if a charge were made for the safe-keeping of their savings. Figure 22 as a whole represents the

coworking of the forces of demand and supply in the determination of a price — in this case the rate of interest — in harmony with general principles set forth in earlier chapters of this book.

Adjustment of the individual to the loan market. It may serve to clarify our discussion of interest if we pause now to summarize the forces lying behind the demand and supply of loan funds. The rate of interest, like any other price, is the effect of the activities of a multitude of individuals demanding or supplying loan funds. But again like other prices, this rate is also one of the causes of the amount offered or taken by any given person. The individual looks upon the existing rate of interest (as he does upon any other price where pure competition prevails) as fixed by forces beyond his control; he is concerned with adjusting his activities so as to make the best of conditions as he finds them. In theory, certain principles may be laid down which control these individual adjustments.

(1) *The individual lender.* A man will lend when his rate of time preference is lower than the rate of interest. He will continue to lend until his marginal rate of time preference is equal to the interest rate. Theoretically he gains from all units of his loan except the last and, by lending up to this point, makes his gain a maximum. In a perfect loan market, the result of the activities of the individuals will be to give them marginal rates of time preference equal to each other. As noted in the preceding discussion some supplies of loan funds are drawn from individuals who save without reference to the interest rate and from others who have no free choice in the matter. This portion of the total supply is not controlled by this principle.

(2) *Consumption loans.* An individual will borrow for consumption when his time preference rate is higher than the rate of interest, and he will continue to borrow until this discrepancy is removed. Like the buyer of goods, he gains utility from all units of his loan except the last one. In a per-

fect loan market the tendency is for the rates of time preference of all such borrowers to reach equality. As stated above however the borrowing of governments is not generally controlled by this principle.

(3) *The enterpriser-borrower.* This person borrows when the anticipated marginal productivity of capital under his management exceeds the interest rate, and he continues to borrow until this discrepancy is removed. Since it is the income from the marginal unit which equates with the interest rate, there is a gain to be derived from the use of borrowed funds for productive purposes. This surplus is not all retained by the enterpriser-borrower. It represents that part of the income of the enterprise which is not absorbed by the interest cost, and it is distributed among the other productive factors which comprise the business unit. The tendency for each enterpriser to borrow until the marginal income from the loan equals the interest rate results, in a perfect loan market, in distributing the capital supply of society among the different business units and the different forms of capital instruments so that its marginal productivity is everywhere the same. This tendency works toward the end that society receives the maximum product from a given capital equipment.

(4) *The alternative of lending or borrowing.* Many people are in position to be either borrowers or lenders, the choice depending upon the relation of their own rates of time preference to the market rate of interest. Let us suppose that a certain person's rate of time preference is six per cent. If he finds that he can borrow at five per cent, it will be to his advantage to do so. But if the market rate of interest is seven per cent, he will have no inducement to borrow. On the contrary, if he is the possessor of any considerable capital, he will find it to his advantage to lend some part of his present resources, up to the point at which his own rate of time preference becomes seven per cent. In particular the business man who is himself an owner of capital will enlarge his enterprise with borrowed funds if the market rate of interest is lower

than the marginal productivity of his enterprise, whereas he will contract his operations and lend his free capital to others when the opposite relation holds. The possibility of shifting from one side of the market to the other contributes much to the sensitiveness of the loan market.

Lack of uniformity in interest rates. Thus far we have spoken of the interest rate as if there were a single uniform price applying to all types of transactions in the loan market. But when we turn to the business world for illustration of our theory we are impressed at once by the absence of any such uniform price. Different rates of interest appear in the market at the same time, each rate applying to a particular type of loan transaction. For example the following schedule of rates was published on three different dates for transactions in the New York money market:

Type of loan	*Per cent*		
	July 26, 1929	Feb. 16, 1935	July 15, 1938
Call loans	10	1	1
60–90 day loans	8	¾ and 1	1½
Commercial paper			
Best names	6	¾	1
Other names	6	1½	1¼

While these variations existed in rates of explicit interest, business ventures financed by their owners were returning rates of implicit interest which also varied widely.

Some of the variations appearing in the tables are due to conditions peculiar to the banking business. A large part of the banker's liability consists in deposits payable on demand against which he must maintain assets in liquid form. Bankers in a money centre like New York receive from banks in inland cities large cash deposits subject to continuous withdrawal and replenishment. For obvious reasons it is not profitable for these bankers to hold such deposits in the form of idle cash. The call loan is the best form of investment for funds of this kind; it is resolvable into cash on demand and secured against loss by a deposit of collateral. For this reason

the call loan in a large money centre usually commands a peculiarly low rate of interest. This rate also fluctuates sharply in response to variations in the amount of money seeking this peculiar form of investment. The student will realize that the interest rates listed in the second and third columns of the table indicate an abnormal condition of the money market. Prolonged business depression had caused the banks to carry an unusually large portion of their assets in cash. This abnormal supply of short-term loan funds, coupled with the decline of demand in the business world, resulted on the dates given in extremely low rates of interest in the New York money market.

True and nominal interest rates. On loan transactions not subject to these exigencies of the banking business, variations in the interest rate are often more apparent than real. We must view interest as a payment for a single kind of service; *i.e.*, the service of postponing income by a voluntary exchange of present goods for future. If the man who renders this service performs at the same time a number of other functions and receives a single payment as a recompense for all of them, only a part of the payment can strictly speaking be called interest; the remaining part is composed of as many different elements as there are different functions to be rewarded by the payment. The easiest of these elements to recognize is the insurance premium which generally accompanies every interest charge. When dealing with true interest we must assume that the lender is assured of his future return; but as a matter of fact no loan is absolutely secure. When repayment of a loan is uncertain to a marked degree, the fact evidences itself in an apparently higher interest rate. This interest rate consists of at least two elements; *i.e.*, interest itself and a premium for risk and uncertainty.

There are also other elements commonly included in commercial interest rates which do not correspond to the true interest concept. The rate which a bank charges its customers for example is not merely a charge for waiting; it is also a

charge for consultation and other forms of service. Interest paid a stock or investment broker also carries a concealed service charge of similar nature. The most common occasion for confusion of interest with other forms of remuneration is the use of an enterpriser's own funds in his own undertaking. In any such case it is exceedingly difficult to segregate true interest from wages which the enterpriser has earned by his own personal efforts; *i.e.*, the wages of management. Moreover what is in reality merely an "eating up of principal" is sometimes falsely regarded as interest. No true interest ever accrues from the use of a capital instrument unless provision has been made to replace it as it gradually wears out, and yet interest is often confused with depreciation.

In summary of these points we may say that nominal and true rates of interest never coincide unless the investment is one in which the investor bears no risk, exercises no managerial powers, and encounters no depreciation in connection with his income. Very probably no investment can be found which measures up exactly to this ideal; but it is approximated in the long-term bonds of a stable government, returns on which may be accepted as a fair exemplar of true interest.

Changes in the rate of interest. The behavior of the interest rate may involve important social consequences. Over very long-run periods, a remarkable stability in the rate is to be observed. When in the seventeenth century the interest rate in Holland fell to three per cent, it stood at a point only slightly lower than during the flourishing days of the Roman Empire. In the following century the rate rose slightly, only to fall during the nineteenth century to the same low point. Just prior to the World War the rate for first-grade securities stood again at about three per cent. But this stability over long ranges of time contrasts with fairly sharp fluctuations in short-run periods. The rate for current loans is a market price which tends to respond to momentary changes in the factors of demand and supply. There is also a tendency for the rate to rise in the wake of rising prices during the "prosperity"

phase of the business cycle and to fall to a low point during the subsequent reaction. But these responses of the interest rate are at best sluggish and retarded in action, a fact which has unfortunate economic consequences. The rates of pre-existing loans, being determined by legal contract, do not change at all in response to fluctuations of the business cycle or change in a disjointed manner at the end of the respective loan contracts. Moreover the rates on current loans are so influenced by customary business practice and other forces interfering with perfect competition in the money market that they too fail to move promptly and smoothly in harmony with changing business conditions. If the price of loans were to rise promptly during the early stages of rising prices, it would act as a brake on unhealthy credit expansion and thus tend to prevent a subsequent crisis, but because of these market frictions it fails to keep pace with the situation.

Similar lagging of the interest rate may be observed during the early stages of the business depression which follows a crisis. The general collapse of prices at such a time reduces the value of business assets and current business income, while the sluggish behavior of the interest rate prevents a corresponding reduction of business indebtedness and business costs. This tends to undermine the solvency of enterprises which are in debt as well as to make it more difficult to place current business transactions on a profitable basis. If the depression is prolonged, the rate may fall to abnormally low levels, as at the present time (1940), because of the continued lack of demand from the business world. This lack of substantial return from loan funds exerts a disturbing influence on the economic system through its effects on the banking structure and on all enterprises and individuals dependent on income from invested funds.

The capitalist class. We have now concluded our study of a second functional type of income, the income accruing to the owners of capital. That this class of income receivers possesses a right to share in the distribution of the product of in-

dustry is due to certain essential characteristics of our type of economic order. Our civilization is founded on the use of great masses of capital; progress in human well-being requires that we set aside each year an enormous amount of present income for the repair and replacement of existing capital instruments and the addition of new ones. The industrial mode of production calls for two distinct kinds of service: working and waiting. Neither of these services by itself can maintain the flow of consumable income at a volume sufficient to satisfy the wants of men on the present plane of living. Man's labor of hand or brain, unless reinforced by his willingness to wait and to endure the costs of waiting, must be applied to direct production without the help of those product-multiplying aids which we call capital instruments.

Since society relies on voluntary saving for the provision of capital, the numbers of those who draw interest incomes must increase, and the total amount of the incomes of this type which accrues to the capitalists must expand as the productive process grows more capitalistic. Indeed it appears to be the fact that the proportion of the total national income accruing to capital owners tends to increase as improvements are made in the technological efficiency of industry.

The income of the capitalist differs from that of the landowner in that the former is a payment for a productive service while the latter is derived merely from the rights of ownership in limited natural resources. Nevertheless at any given time the social position of the capitalist bears a striking resemblance to that of the landowner. Although the service of waiting is an indispensable productive function, this service is not rendered at the time the interest income is received but has been performed in the past, when the saving was done. Indeed in many cases the service of waiting is not even performed by the same man who receives the interest but by some predecessor from whom he has inherited his wealth. One consequence therefore of the creation of a distinct capitalist class is the existence of a group of people who, like the landowners,

appear to enjoy the right of sharing in the income stream of society while making no concurrent contribution to the stream.

The spectacle of the receipt of incomes by individuals, some of whom appear to be idlers, affronts the sense of justice of many a man who does not reason clearly with regard to the capitalist's function in society. Other grievances associated with the existence of interest as a private income add to his bitterness and give impulse to his attack upon the capitalist class. By inheritance ownership of capital falls to men and women who live in opulence as drones in the midst of a busy world. Powers other than the mere right to receive interest are gathered into the hands of the capitalists, power of command over the lives and welfare of propertyless men and women, power to influence for good or ill the functioning of society's industrial structure, preponderance of power in the political world, all of which may give the capitalist a sinister aspect in the view of the common man. Reaction against the institution of interest as a private income leads to organized movements which propose a thoroughgoing change in the present economic structure of society. Before finally concluding our study of economics we shall take occasion to examine these programs of revolutionary change.

CHAPTER XVII

Wages: The Income of Labor

Economic stratification of labor and of wages. The term *labor*, as used in economics, may be defined as an effort of man, made wholly or partly for the purpose of production. It includes every form of human exertion, from the purely muscular movements of the ditch-digger to the mental activity of the professional man or the business executive. The term *wages* is likewise broadly defined as *remuneration received in return for any form of labor.* In the vocabulary of economics it includes not only wages in the commonly accepted sense but also salaries, fees, commissions, and sometimes, as we have seen, certain parts of what are popularly called interest and profits.

At the outset of our search for an explanation of the share of social income accruing to the recipients of wages, we encounter apparently insuperable complexities. Looking out into the markets, shops, factories, and farms of the world, into great avenues of transportation, and into quieter by-paths, we immediately perceive numberless varieties of labor and of wage rates. Our first step must be to simplify the problem by attempting to throw these variations into some sort of order; and upon closer examination we discover that the gainfully employed people of any nation may be more or less simply classified. One classification, sufficiently exact for our purposes, divides all labor into five fairly self-contained groups.[1] At the bottom of the scale are the unskilled workers in all occupations — the manual laborers of the farms, the railroads, the industries of the country. In the second group are the manual workers of some little skill whose occupations require a certain amount of responsibility and mental alert-

[1] F. W. Taussig, *Principles of Economics*, fourth edition, 1939, Vol. II, pp. 234–238.

ness — the machine operatives of the factories, the pick miners, the motormen of the street railways, and the like. The third group comprises the skilled craftsmen, the aristocracy of the manual workers, such as the craftsmen of the building trades, the skilled workers of the steam railroads, the die and pattern makers of the metal trades. In the fourth group are found the "white collar" workers who provide the office and clerical staffs of business enterprise, the sales forces of retail establishments, teachers of the lowest salary grades. The fifth and highest group contains those superior workers who are either self-employed or occupy salaried positions of heavy responsibility. The learned professions, the managers of enterprises, the higher public officials, the independent enterprisers are examples.

For our immediate purpose the important point in this stratification of the labor market is the existence of competition within each major labor group and the relative lack of competition among the several groups. Competition of course implies ability to substitute one worker for another. This substitution can and does occur on the plane of any one of the economic strata described above; but there is comparatively little transference of labor from a lower to a higher class. As a result the wage rates tend toward uniformity within each broad labor class, and toward diversity between classes. The uniformities, of course, are not complete. Individuals even in the lowest group of manual labor receive somewhat different wages; diversities are even more marked in the earnings of the different types of skilled craftsmen; in the highest group quite striking differences may be found. We shall take account of these individual variations in due course, but in the discussion immediately following we shall stress the uniformity of the wage rate within a given labor group.

The demand for labor: Why employers hire laborers. Assuming therefore that we are dealing with a single homogeneous labor group — *i.e.*, a group within which one individual can be substituted at will for another — how is the

wage rate determined? A brief answer is "by demand and supply"; but full understanding requires us to consider the nature of the demand for labor, the peculiar characteristics of labor supply, and the joint action of demand and supply in the employment market.

Wages are paid by profit-seeking employers, who voluntarily undertake to give the workers a money income in exchange for their services. Each employer assumes responsibility for paying wages for the same reason which impels him to burden his future income with interest charges when borrowing capital. That is to say, he expects the workers to create the income out of which wages will be paid, just as when borrowing capital he expects the employment of the capital to increase his income by an amount at least equal to the interest charge. In both cases he has but one source from which to recover his costs; *i.e.*, through the sale of the product of capital and labor. The employer's demand for labor then arises out of his expectation of a money receipt for labor's product, in other words, out of the anticipated monetary productivity of labor.

Marginal productivity of labor. In any homogeneous labor group such as we are considering it is not the total but the marginal product of the group which governs the employers' demand for labor. Employers are not confronted with the choice of hiring the entire labor force as a single unit or hiring none at all. They employ laborers as individuals and are concerned with the specific contribution of the individual worker to the income of the enterprise. The amount of this contribution is measured for each member of a homogeneous labor group by the *marginal productivity* of the group; that is, *the amount per laborer by which a small change in the number of laborers changes the total product.* Were it practically possible to calculate the effect of adding a single laborer we might say "the amount added to the total product by employing an additional laborer."

The productivity of labor conforms to a law of diminishing

marginal productivity. There are two reasons for this. We have discovered as a general rule that when one productive factor increases while the coöperating factors remain fixed in quantity, the physical product of the combination tends to fall as measured in proportion to the number of units of the increasing factor. This is true when labor is the increasing factor. The physical product per unit of labor will eventually decline as the number of laborers is increased. In the second place, under conditions of imperfect competition, increasing quantities of output must generally be sold at lower prices. It follows that the contribution which each unit of labor makes to the income of a business enterprise tends eventually to fall as the employer expands his labor force, other factors remaining constant. The principle of diminishing returns, as applied to labor, means that the employer cannot go on expanding his labor force, land and capital being constant, without noting a decline in the amount of income attributable to one unit of labor.

Under these conditions the employer, confronted with a given rate of wages, will add to his labor force until the contribution of the marginal man in each group to the income of the enterprise is no greater than his wages. In simple language this means that the employer will continue to hire men as long as he gains by so doing, and he gains as long as the marginal man's product is in excess of the wage rate. Evidently the lower the wage rate the greater the number that can be employed before this limit is reached; hence *the number of laborers demanded by a given employer or by all employers under any given set of circumstances varies inversely with the rate of wages.*

Productivity discounted. One important technical feature remains to be considered. The workers will want their pay in advance of their employer's sale of their product. Generally at the end of each week the laborers will ask to be paid for the services performed up to that time. But in the majority of cases the products upon which they have been working are not in a completed form and ready for the market at

the time the wages are paid. This phase of the matter was touched in our study of interest, where the capitalistic process was described as a succession of advances to workers. But the employer will not thus make advances to laborers unless he is paid for his waiting; that is, unless he receives interest upon the advance payments. The wages which the employer can afford to pay the marginal worker therefore will fall short of the value of his contribution to the product of the business by the amount of the interest charge. Thus the employer is compelled to discount the marginal product of labor in deciding what wage he can offer.

The supply of labor. Other things being equal, the quantity of any good which will be offered for sale varies directly with the price. In general this is true of the supply of labor within any single occupational group and within the wage class as a whole; but it should be noted that the circumstances back of the supply of labor are in many respects peculiar. The amount of labor available at any given time is a comparatively fixed quantity, not susceptible to great variation in relation to the rate of wages. In other words, the total supply of labor is comparatively inelastic. Even if we consider a single homogeneous labor group, the inelasticity of supply is only slightly modified, since, as we have already seen, laborers are not easily drawn from one group into another, even by the inducement of considerably higher wages. Moreover changes in the number of laborers take place slowly, and there is only in a remote sense such a thing as a cost of production of labor. Therefore though the supply of labor is similar to the supply of things in general, in that the quantity offered varies directly with the price, nevertheless we must conclude that in the case of labor this variation is only moderate.

The general law of wages. These forces of demand and supply would, if unimpeded, determine a market rate of wages in a manner similar to the determination of other prices in a purely competitive market. The demand for each

type of labor is a schedule of rates representing the discounted marginal productivities of labor groups of differing sizes. Because of its inelasticity, the supply of labor is at any given time virtually a fixed number of workers in each homogeneous labor group. The conditions of demand will fix the highest rate at which this number of workers can be absorbed in employment under the free play of economic forces. The general law of money wages may be stated as follows: *The normal rate of money wages for any uniform group of workers is equal to the discounted marginal productivity of that group.*

This is the basic law of wages. It holds true on the assumption that each employer must pay the prevailing rate of wages as determined by the general forces of demand and supply. The individual employer assumes that the wage rate is fixed by market forces extraneous to himself, but he plays a part in causing this rate to be what it is. Were he to discontinue his enterprise and withdraw from the market, the workers whom he released would have to seek employment elsewhere, thereby (in theory) lowering the marginal productivity in other enterprises and reducing the wage rate. On the other hand, if this employer is enabled by some development in his business to increase his labor force without reducing their marginal worth, his increased demand for labor can be satisfied only by diverting wage earners from other industries, thus raising marginal productivity in general and increasing the wage rate.

Practical qualifications: Theoretical precision lacking. This is the theoretical view. But our general law of wages is obviously an abstract statement of a tendency, rather than an explanation of all the diversities which may be observed in the business world. We must now add qualifications in order to take account of some of the more important of these practical considerations. It is obvious, in the first place, that the appearance of precision given in our illustration to the employers' calculation of the worker's marginal productivity is artificial. In the complexities of modern industry with its elaborate division of labor no one can say exactly what one

worker contributes to the corporate income. At best these calculations are intelligent guesses based on experiment and experience.

Interdependence of different types of labor. In the second place, it never works out that the different enterprisers enlarge their several labor forces to the point where the marginal productivities of similar grades of labor are precisely equal. The number of workers of a certain type which can be employed profitably in a given industry is affected, not only by the marginal productivity of these workers, but also by the necessity of maintaining an effective proportion between the different types of labor in the enterprise. Complex division of labor creates a situation in which no single type of labor can be employed without necessitating the employment of other types. For example an industry employing one bookkeeper might not have need for the services of another at any rate of wages, so long as its present scale of operations remained unchanged. To make it possible for a second bookkeeper to produce anything might require the employment of a dozen machinists, a foreman, a shipping clerk, and two or three other types of labor, none of which could be employed profitably at their prevailing wage rates. In a certain large factory producing tennis shoes one inspector is employed in conjunction with twenty-five workers in other lines; one cutter works with half a dozen stitchers, etc.

These examples illustrate the fact that the demand of any employer for a given type of labor is limited by his force of coöperating types, which are limited in turn by the wage rates of these other labor groups. He tends to increase the number of workers in each group up to the point where their marginal productivity equates with the wage rate, but he may stop short of this point because of his inability profitably to expand the scale of his industry. Obviously the small business is much more affected by this consideration than is the large one, which, with adequate supplies of capital, can take more complete advantage of the gains inherent in complex

division of labor. When however attention is fixed on a single occupational group, it is probably true that in no individual enterprise could there be found that precise equivalence of marginal productivity with the wage rate which our law of wages implies.

The sphere of bargaining. Another qualification of much practical importance applies to the implication of our law of wages that the rate of wages tends to settle automatically and without friction at a point which measures the marginal productivity of the workers. This is because conditions of perfect competition have thus far been assumed. But in practice the rate for any group of wage earners in any given industry is settled by bargain between employer and worker and fluctuates according to the relative bargaining power of these two parties to the wage contract. The actual rate in a chosen industry or for a chosen section of the laboring population may for a period be lower than that determined by economic law; on the other hand, it may rise above this point for short-run periods in isolated enterprises or localities. We cannot undertake to follow out in detail these complex variations, but it will be profitable to ascertain as nearly as can be the limits of this bargaining area within which actual wage rates fluctuate.

Wages above marginal product. Circumstances may arise in which the employer will voluntarily pay some of his workmen more than they are worth, thus leaving himself worse off than if he had refused to hire them at all. Employers are not devoid of feelings of compassion and do at times mix charity with business. In other cases it may be to the positive advantage of an employer to accept a loss on the wage contract for a short time in order to avoid a greater loss involved in dismissing his employees. During periods of depression the enterpriser will look forward to a recovery of the market which will set his industry in full operation again. By keeping an experienced labor force together he will avoid the considerable loss involved in hiring new workers, breaking them in to the routine of the plant, and training them in his methods of

management. It may be worth his while to do this even at the expense of paying his workers for doing nothing. These are but indications of the motives which may impel employers voluntarily to pay wages above the rate set by economic law.

It is also possible for the workers, under certain conditions, to force an employer against his will to pay them more than their marginal worth. By threat of strike a group of workers which holds a temporary monopoly on the supply of labor may compel the employer to make a choice between paying the members of this group more than the value of their marginal product or of suspending operations. If an employer could always close his plant without loss, it is not probable that he would ever accede to such demands. But this he cannot do. To suspend operations will at least wipe out his own wage income along with the possibility of profits. If compelled to do so, he may pay wages above the marginal worth of the workers, surrendering to them a part of his own income up to the point where his share in the revenues of the enterprise has disappeared. Even here we do not have a final limit to the advance of wages, for it is conceivable that aggressive bargaining by the workers may raise the rate so high that the enterpriser is not only not receiving payment for his own labor but is actually operating at a loss. Most employers have fixed costs which will run on unchanged even if they close down their industries. It is the lesser of two evils to run at a current loss, provided the loss is not as large as the total of the current fixed charges. This condition is of course self-limiting; its ultimate limit will be reached when the employer's equity in the business has been wiped out. But it is quite possible that an employer can be made to bear a loss of this kind, and there are plenty of cases of employers who have been forced into bankruptcy as the result of a series of losses on the wage contract.

We see then that there are three resistance points to a rise of wages. The first and most important is set by marginal productivity. It is rarely "good business" to pay a higher rate

than this, and no employer will do so intentionally unless it be as a temporary expedient in abnormal circumstances. It is also rarely good business for laborers to demand more, since their employment is determined ultimately by profitable enterprise. The second check on rising wages is exerted at the point where the employer's profit and wages of management disappear. This is the maximum limit to the range of bargaining except in industries materially affected by fixed costs. The highest conceivable rate under any conditions is set at the point which causes a current operating loss equal to the loss involved in closing down the industry. Wages will rise above the first of these points, that determined by marginal productivity, only when the workers are possessed of some unusual bargaining advantage. Even then such a rise of wages can appear only in isolated cases and for short-run periods. When his wage cost rises to the second of our resistance points, the employer will soon give up the struggle. No employer will be content to work indefinitely without reward. The duration of a wage rate at a level between the second and third resistance points is automatically self-limiting by inevitable insolvency. We may safely conclude that advances of the wage rate above the point of the discounted marginal product will appear only as exceptional and short-lived occurrences.

Perhaps notice should be given to a type of situation sometimes encountered in the business world which is often cited in refutation of this statement as to the temporary nature of wage rates above the marginal productivity level. Tightly formed and aggressive unions do sometimes raise their wage rates to extravagant heights and hold them there over considerable periods of time. But they do this by monopolizing and restricting the supply of labor. This raises its marginal productivity and drives up the price of the product on the market. The employer who pays these high rates does not necessarily pay more than the discounted value of the marginal product in the monopolistic situation which exists.

Wages below marginal product. We may now consider the reverse condition. The assumption that employers will be required to pay wages equal to the full economic worth of the hired worker rests upon the belief that competition works as freely on one side of the labor market as on the other. But this is untrue in actual life. For wage earners obviously are not as a rule in position to take full advantage of competition among employers. To do so they must be able to play off one employer against his competitors, to move about at will from one locality to another, to discard service in one industry and undertake employment in a different one whenever it is to their immediate economic interest to do so. All of this requires a degree of mobility in the ranks of labor which the average workman does not possess and which he is prevented from acquiring by (to mention but one factor among many) the mere cost of transportation for himself, his family, and his household effects. It must be remembered that the labor market ordinarily requires the workman to carry his labor to the employer. Perfect competition in the labor market then implies physical mobility among laborers, to attain which they would have to overcome obstacles of financial expense, ties of home and community, custom, tradition, sentiment.

There seems to be no doubt that the bargaining weakness of the workers has often resulted in a wage rate temporarily or in particular situations lower than their marginal economic worth. Local situations may arise in which the play of ordinary competitive forces is interrupted. This condition of imperfect competition may enable employers to obtain labor for less than it is worth; *i.e.*, at wages less than marginal productivity. Such a state of affairs, in particular situations, may be long continued. Conditions of this kind have sometimes been disclosed when a compulsory increase in wages has not resulted in either unemployment or a higher price for the product.

It would aid us to limit the range through which the wage may fall if we could ascertain with precision lower limits to

the sphere of bargaining comparable to the resistance points which check a rise of wages. In a general way two such resistance points can be marked out. The first of these is determined by the wage earners' traditional standard of living. Among the skilled labor groups especially, self-respecting workers will reject offers of employment at rates which spell a degradation of their standard of life and a loss of social status. A second and ultimate lower limit is reached when the rate falls to the minimum of subsistence. If starvation is to be their fate in any case, free men will prefer to starve at leisure rather than while working for another man's profit. Even if they were content to work for less than subsistence wages, the duration of such a wage rate would be limited by its inability to maintain the supply of labor.

Differences in wage rates within occupational groups. For purposes of exposition, we have thus far assumed that the workers in each specific line of occupation are so homogeneous that one can be substituted for another at will. Under this assumption all members of a given occupational group would receive the same rate of wages; all carpenters would be paid alike, all stenographers, all machinists, all school teachers, etc. Everyone knows that this is not strictly true in fact. The wage rate varies for different members within an occupational group and also for different localities. Of course we mean not merely the money payment, but the real wage. It is necessary to be certain that apparent differences in wages are real differences. If there are wide discrepancies in living costs between different localities, variations in money wages may not measure the comparative real incomes of the workers. If employment is erratic, as in seasonal industries, a high daily rate may not mean large annual earnings.

Considering now the lack of uniformity in the real earnings of different members of the same occupational group, it is obvious that this fact may sometimes be due to relative difference in bargaining power such as we have discussed above.

Women are habitually paid less for the same service than men. Negro labor in the United States is similarly underpaid in comparison with white labor, and the same is true of recent immigrants in comparison with workers who have acquired our culture and standard of living. These are examples of low wages due to bargaining weakness.

But of chief importance in explaining lack of uniformity in the earnings of different members of the same labor group is the fact that no labor group is quite homogeneous in productivity. If we select any occupation at random, we will observe that all its members do not have the same productive efficiency. Noticeable differences appear in the number of tons of coal credited to the account of different miners at the end of an eight-hour day. The tenders of machines in a cotton mill will differ somewhat in the speed with which they perform their operations and in their ability to sustain a given rate of production through the hours of the working day. Closer observation will show that these differences in productivity are greater in the case of some labor groups than others. In the crudest forms of labor the difference in the output of one member of the working force as compared with another is usually slight. So also with the great army of workers in the factories; there is a marked uniformity in productivity among the millions of machine operatives in the great industries. The differences are great among skilled craftsmen, where native ability and training play a relatively large part in the mastering of the trade. And differences are greatest of all in the highest labor groups, the learned professions, and the superior grades of managers and superintendents.

Local variations in wage rates. Wage rates not only differ for different members of a given occupational group in the same place, but they vary between localities. To some extent these differences may be attributed to the effect of the bargaining factor which we have discussed. For various reasons the bargain which fixes the actual wage rate within the limits

described on a preceding page terminates variously in different localities. Enterprisers are not alike in their efficiency as managers, and consequently different business concerns vary in profitableness. Some units of a given branch of industry are equipped with better machinery or possessed of more favorable markets or otherwise favored as compared with competing units. These differences affect the ability of an enterprise to pay wages, and they also affect the aggressiveness with which the employer bargains for a saving of cost in the wage contract. Employers with the smallest profit and those operating without profit or at a loss are proverbially the most aggressive bargainers. On the other side of the labor market, differences in the character of the working people, in their standards of life, in the degree to which they are unionized, and in other local conditions materially affect the bargaining strength of different groups of laborers. Such differences in local conditions modify the resistance points in the fluctuation of wages and cause local variations from the norm.

The basic cause of local variations in wage rates however is the fact that the labor market is broken up into small fragments between which the forces of competition either do not operate at all or operate haltingly and with much friction. While prices of commodities in general in different places cannot differ by more than the cost of transportation between these places, labor is an exception to this rule on account of its extreme immobility. If all the workers in a given occupation could and would move at will over the whole industrial field, local wage rates would tend toward much greater uniformity. But in reality each business centre forms a labor market of its own, where the relation of supply to demand may be quite different from that of other centres and where consequently the marginal productivity of the workers will also be different. Local wage rates tend to vary in rough proportion to the cost of living; they are lower in small towns than in the metropolis, in southern than in northern cities, in the Middle West than on the seaboard. This fact lends

support to an erroneous theory that the wage rate is determined by living costs. But local differences in the cost of living can affect real wages only through their influence on the distribution of the labor supply and on the bargaining power of the workers.

Discounted marginal productivity is fundamental. Because of the complexity of the labor market it has been necessary to introduce various qualifications into our law of wages in order to bring it into rough conformity with the facts of industry. We have recognized the fact that the actual wage rate is determined by bargaining and have marked out the limits of the sphere within which the incidence of the bargain will fall. We have taken account of the existing lack of uniformity in the wage rate for members of a given occupational group and have accounted for this by referring to differences in bargaining power and especially to differences in individual productivity. We have also noted local variations in wage rates for a given type of labor and have shown that these result from local differences in relative bargaining power and especially from maladjustment in the labor supply.

But this recognition of complicating factors must not be permitted to obscure the essential truth embodied in the economic law of wages. The wage bargain revolves around the point of discounted marginal productivity as a norm; actual wage rates cannot long persist above this point, and all the developmental tendencies in our society work against their falling far or remaining for long below this point. Homogeneity of labor groups is promoted by various tendencies; other forces are broadening the scope of the labor market and promoting a more even adjustment of the local supply of labor to the demand for it. We may conclude that no device of wage payment and no alterations in the method and procedure of the labor market can materially affect the fact that the laboring class in a society sanctioning private property and individual enterprise will normally receive wages approximating the value of their specific contributions

to society's income. If they are members of groups whose individuals have much the same rate of productivity, the value of each man's contribution will be equated, through the principle of substitution, to that of the marginal man in the group. If they must receive their share of the social income before the sale of their product, their money incomes will be reduced by the discount process through which present and future values are equated. Upon these valid principles rests the economic law of general money wages, the law of the discounted marginal product of labor.

Other types of labor income. The foregoing law of wages was developed with reference to the hired workers; but its conclusions will serve to explain the money incomes of those who are their own employers. Here the typical case is complicated by the fact that the independent worker must almost of necessity be an owner of capital and to some extent a manager of enterprise. The physician, the lawyer, the barber — examples could be multiplied indefinitely — possess productive capital of one sort or another and often land as well; their money incomes are a mixture of rent, interest, and wages, no single element in the income being clearly defined in contrast with the rest. It is not impossible however to arrive at a rough measurement of that part of their income which is attributable to labor, by deducting allowances for capital and land at the market rates of interest and rent.

In the case of independent workers, no intermediary stands between them and the market. The fact that these groups are less homogeneous than are groups of hired workers, that there are, for example, relatively large differences in the productive efficiency of physicians, attorneys, and consulting engineers, tends to bring about marked variations in the earnings of these groups. But apart from these variations in money wages, these classes of the laboring population constitute no real exception to the law of wages. The normal income of the doctor will be high or low in harmony with the relative scarcity or plenty of the services of this profession or, to use the terms of

our law of wages, in harmony with the marginal productivity of the group. So it is with other classes of independent workers.

The more apparent exceptions to the law are exhibited by those groups of workers whose incomes are determined in some material degree by non-economic considerations. Such are those connected with charitable enterprises, members of religious orders, to a less degree teachers and government officials. There can be no direct relation between the earnings of these workers and the market prices of their products, for the sufficient reason that the prices are not determined by market considerations. But even here the economic law of wages is by no means without effect. These occupations must compete for workers with opportunities in the general world of business and professions. The rewards they offer cannot be set too far below those afforded by competing occupations to men and women of the same ability, else they will be unable to obtain the requisite supplies of labor. The periodic complaint over the dearth of teachers, ministers, soldiers in the standing army, and "good" men in all branches of the civil service shows what happens when the wage policy of these non-commercial enterprises is out of line with the general economic conditions of the times. Making allowance for the non-material rewards of such types of labor as we have been considering, such as the prestige which attaches to them, security of tenure, relief from strain of competition, congenial surroundings and associates, and the like, we shall find that wages in these occupations show rough similarities to the ruling rates fixed by economic forces in comparable lines in the business world.

Wages and the contrast of social classes. The greatest differences in wages which we observe in actual life mark the contrast in the economic positions of the several non-competing social classes; *e.g.*, the class of unskilled manual laborers and that of the executives, the managers, the self-employed enterprisers. These great inequalities do not measure the irksomeness of the task performed, for often the man

who does the most exhausting and unpleasant work is more poorly paid than any other. Nor do they indicate the relative social importance of the type of work as such. It is quite impossible to say whether the managerial function as such is more important than manual labor; both are indispensable. What the inequalities do measure is the relative marginal productivity of the social groups in question, and this in turn is a function of the relative plenty or scarcity of numbers in the group. The underlying reason is the fact that men cannot transfer at will from the overcrowded to the less crowded social classes; or to state it otherwise, that the members of these diverse classes do not compete across class lines.

Speaking broadly there are two contrasting explanations offered for the persistence of social stratification through the ages. One group of writers holds the view that conditions of birth and the heritage from his ancestors determine the rank which a man will take among his fellow men. Those in the humblest station are predestined at birth to occupy this position in the lowest income group. Those who rise to higher levels do so by virtue of inborn characteristics which can neither be produced intentionally nor balked of their reward when present. With these writers heredity is the controlling factor in the destiny of men; the present stratification of society is a natural order determined by nature's laws.

In contrast with this view stand the environmentalists, who believe that inequality is due fundamentally to the institutional structure of our social order. This group asserts that man's social destiny is determined after birth by the forces which play upon his plastic consciousness during his early years. In their opinion the children of the lowest group of wage earners fail to rise above their fathers' station in life because of the stunting of mind and body and the impoverishment of will which they suffer from the unwholesome environment of their childhood. Able men are born in about equal proportions in all ranks of society, but those in the lowest ranks are the victims of physical, moral, and intellectual mal-

nutrition. They are denied opportunities for self-development, opportunities which are showered in abundance upon the children of the rich, regardless of intrinsic merit.

The chief significance of this search for the basic cause of social stratification is its bearing on the possibility of reform. It may be granted that the stability of the social order, to say nothing of the welfare and happiness of the average human being, would be improved if life conditions within the social system were more nearly uniform. Striking inequalities between individuals, when not attributable either to natural law or to some obvious principle of justice but rather to the arbitrary workings of traditional social institutions, arouse in many men a determination to reconstruct the entire social order, if necessary, in order to promote greater equality.

Labor's share of the social income: The law of proportionality. Assuming that men continue to live and work in a society like our own, what prospect is there that labor, either in its entirety or within a given labor class, can increase its share of the social income? For society as a whole, improvement in real income can be due only to some general development affecting the physical productiveness of the economic system. But it does not necessarily follow that an increase in the total income of society will be evenly spread over the several shares in distribution. It is quite possible that the forces which bring about increased general prosperity may so operate as either to increase or decrease the fraction of the whole which goes to labor, as well as the average real wage rate.

Our analysis of diminishing returns, applicable alike to rent, interest, and wages, brings to light a principle of universal truth with regard to the relative economic prosperity of the different income receivers in society. This principle may be formulated as follows: *Any social change which reduces the ratio which one factor or combination of factors bears to those coöperating with it will tend to raise the average income of that factor or combination of factors as compared with the others.* If population in-

creases faster than land and capital, real wage rates tend to fall in proportion to rates of interest and rent. If capital accumulates faster than the increase in population and natural resources, wage rates and rent tend to rise in proportion to interest. If the available supply of land and natural agents expands out of proportion to labor and capital, the unit rent income tends to shrink as compared with wages and interest. This generalization is called the *law of proportionality.*

Applying this principle to the economic welfare of the laboring population, it will be seen that average wages per worker, as a share of the whole income, will be increased by any development which expands the social income by increasing the quantity of productive factors other than labor. The principle is applicable not only to changes in a factor of production taken as a whole in relation to the other factors, but also to changes of balance among the sub-groups of a given factor. If within the labor supply, for example, managerial labor increases in quantity while other types of labor do not keep pace with it, the average real income of the one will tend to fall proportionately to that of the others. This feature of our exchange economy has far-reaching effects which are not always favorable to general social welfare. It means that it is often to the immediate economic advantage of each separate labor group to restrict its membership and raise the value of its marginal product. When the workers in a given occupation are organized into trade unions and are able to prevent those who are not union members from engaging in the occupation, they are tempted to adopt restrictive practices which artificially limit their numbers and, in an equally artificial manner, increase the numbers in other lines of activity. This will raise their marginal productivity, make their product relatively scarce as compared with other products, and increase their average real wages. Much the same result is obtained if the group in question adopts "make work" policies which set a limit to the output of the average man in a given time. In other words, the arrangements of

our social order make it possible for specific labor groups to benefit themselves by working positive injury on society as a whole.

EXERCISES

1. The following is a statement of the cost conditions in an automobile repair shop, in relation to the number of mechanics employed: (1) wages are at the rate of $150 per month per man; (2) other costs (except wages) are $12,000 when 4 men are employed and increase in proportion to the number employed; (3) interest is $6,000 when 4 men are employed and increases 10 per cent whenever an additional employee is added (*i.e.*, 5 men require 10 per cent more capital and therefore 10 per cent more interest payment than 4 men; 6 men require 10 per cent more than 5 men, etc.), (4) rent is constant at $2,000 a year. The annual income realized with different numbers of mechanics would be as follows:

Mechanics	*Income*
4	$25,000
5	32,000
6	39,000
7	46,000
8	51,000
9	55,000

How many mechanics would be employed?

2. Men's shoes are sold in a retail store at a gross profit or "mark up" of $1.00 per pair. The salaries of salesmen are $160 per month. Assume that the following relationship exists between number of salesmen employed and monthly volume of sales:

Number of salesmen	*Volume of sales*
1	480 (pairs)
2	880
3	1,200
4	1,440
5	1,600
6	1,680

(*a*) Determine the marginal cost per unit of sales when different numbers of salesmen are employed.

(*b*) How many salesmen would be employed and what would be the volume of sales?

CHAPTER XVIII

Profits: The Income of the Risk Bearers

The word which stands as the title of this chapter is one of the commonplace expressions of everyday speech, yet it is rarely given a fixed or definite meaning. Like many other technical terms of economics, it requires careful definition in order to avoid confusion of thought. As used in economics, the term denotes a type of income distinct from the other types which have been considered in the preceding study of functional distribution; *i.e.*, a share of the social income separate from rent, wages, and interest. It is our purpose now to identify and define this particular category of income and to consider the forces which account for its existence and determine its amount.

The accountant's profit statement. As a point of departure let us place ourselves in the position of a typical business man recording the results of his year's activities. Let us suppose that he is a retail coal dealer, that he leases the land and borrows all the capital employed in his business, and that, instead of managing the business himself, he employs a manager. The following account, agreeing with the form with which the reader is familiar, records in round numbers the income and outgo of the business for a certain year.

PROFIT AND LOSS STATEMENT, 1939

Net sales		$100,000
Cost of goods sold	$75,000	
Selling and administrative expense	20,000	
Total costs of operation		95,000
Net operating profit		$ 5,000
Other income (none)		
Other expense		
Rent	$ 1,000	
Interest	2,000	$ 3,000
Net profit		$ 2,000

This account will suffice to indicate the essential nature of profits. It will be noted that the accountant uses the term "profit," with appropriate qualifying terms, in at least two senses. The net results of the business as an operating entity are shown by the item "net operating profit." But against this must be charged the fact that this business required that certain land and capital be devoted exclusively to its uses, thereby being withdrawn from other possible remunerative employment. The use of these factors could not have been obtained except by agreeing to pay their owners economic rent and interest respectively. And it is only after deducting these payments that we arrive at the true net earnings which go to the owner of the business as such. The accountant calls this "net profit," and this is true *profit* in the economic sense.

Implicit rent, interest, and wages of management. Let us now assume another retail coal business, precisely like the one we have considered except that the land and capital belong to the owner and the owner manages the business himself, adding the further assumption that the manager's salary in the first enterprise was $5,000. In our second case, — there being no item for manager's salary — the "selling and administrative expense" will be $15,000 instead of $20,000, and the "net operating profit" will be $10,000. This will apparently be the final "net profit," since there are no further deductions for rent and interest paid.

In harmony with popular usage, which assumes that profits are what is left to the owner out of his business after deduction of payments for costs, the ordinary man might regard this $10,000 as the "profits" of the year's business. But in our search for the true profit, we need to remember that the land site owned by this coal dealer could have been rented, had he so chosen, to another enterpriser for $1,000 a year, and that the capital funds, which he had temporarily tied up in buildings and other capital instruments (trucks, shovels, coal chutes, etc.) and a stock of coal, might have been invested so as to yield him interest to the amount of $2,000 per year. Let us

also recognize that the coal dealer himself, instead of devoting his time to this particular enterprise, might have hired himself out to another coal company at an annual salary, let us say, of $5,000. In view of these considerations the economist insists that this particular enterprise has not really brought in a profit of $10,000.

From this "net operating profit" of $10,000 there must be deducted the rent, interest, and owner's salary, all three of which could have been obtained by the owner had he not gone into the coal business for himself. Only the remainder, $2,000, can be said to have come to him on account of the business. This is the true profit. We thus reach the general conclusion that in economic analysis and in all accurate accounting, before true profits emerge due allowance must be made not only for explicit costs (payments to non-owners) but also for the implicit rent, interest, and wages of management where these costs are not represented by payments to others.

Definition of profits. We have now arrived at a clear and exact concept of profits and are in position to formulate a definition. *Profits are the net income of a business, or the difference between the income and the costs, the costs including rent, interest, and wages of management, whether explicit or implicit.* We must note that profits are essentially variable and uncertain. In any given enterprise they may be large one year and small the next. Costs may exactly equal income, in which case profits are zero, or may even exceed income, resulting in losses instead of profits. We shall find it an aid to our analysis to regard losses as negative profits, which makes it possible to say that profits are always obtained by subtracting costs from income, even though the former be the greater and the result therefore a negative quantity.

Actual profits and losses. Of the actual existence of profits, as a type of social income, there is evidence aplenty. During the twenty-year period between 1901 and 1922 the United States Steel Corporation accumulated surpluses of over three-

quarters of a billion dollars in addition to dividends paid to stockholders. In one ten-year period the Standard Oil Company of New Jersey paid annual dividends of 20 per cent to its common stockholders and accumulated several hundred millions out of earnings. Henry Ford began his manufacture of motor cars with an investment of $49,000 and within twenty years had reinvested earnings amounting to nearly half a billion dollars. These earnings were by no means all profits, since they must be assumed to include interest on the capital invested. But even though no precise segregation is possible, there can be no doubt that very large profits were present in these cases.

Such spectacular examples however should not be allowed to stand alone. The other side of the picture discloses an impressive record of losses or negative profits. Statistics show that at the end of each year more than one hundred thousand corporations have nothing left for dividends after payment of business costs. In 1917, a year of exceptional business prosperity, 119,000 corporations showed an aggregate deficit of 629 million dollars. The entire business world is subject to startling reversals of fortune. In 1919 only about two-thirds, and in 1921 less than half, of American corporations reported net earnings to the United States Treasury. In 1921, out of a total of 356,897 corporations reporting to the Treasury, 185,158 showed a combined loss of four billion dollars, an amount only a little less than the combined net earnings of the successful corporations. In 1932 the combined statements of 23,846 manufacturing corporations reporting to the Treasury showed that losses exceeded profits by one and a half billion dollars. These were bad years, to be sure, but they illustrate the fluctuation of earnings resulting from business risk.

Our statistical data are not sufficient to permit of anything approaching an exact balancing of the total profit income against total losses. But there can be no doubt that business losses are on the whole much more nearly equal to business profits than is popularly assumed. It is at least a reasonable

inference from available evidence that the amount of positive profit for the entrepreneur group as a whole is only a small share of the total social income.

The risk-bearing function: Business risks. What distinguishes profits as a functional share in distribution is that they arise out of the uncertainties of modern business. Business risks may be roughly classified in two main groups: (1) those which are incidents in routine business operations, and (2) those which arise out of dynamic changes in the general structure of industry and society. The former concern individual enterprises within an industrial field; the latter exert their effects upon industries in their entirety. Ordinary business risks, in so far as they become foreknown, lose their uncertainty and can be handled as costs. The development of the practice of insurance makes this clear. When it is discovered that a given risk is insurable, the losses which it entails are apportioned among all affected by it. Payment for insurance becomes one of the normal and calculable costs of doing business and plays its part in determining market prices.

But not all of the routine risks of business are thus amenable to insurance, and in addition there are the more important uncertainties, generally uninsurable, which are attendant upon growth, change, or decay in industry or society at large — changes in industrial processes over short and long periods, changes in the nature and range of human wants, changes in the availability of natural productive agents.

Risk bearing and its reward. The vast majority of us take no thought about these matters. The work of anticipating future needs, discounting future hazards, making timely adjustments to forthcoming alterations in the conditions of economic life is left to a self-constituted body of men who possess the requisite foresight, courage, and energy. We may call these men the risk bearers, if we are careful to keep in mind the broad implications of the term; but their major social function is that of guiding and directing the course of economic development so that future conditions will find us

prepared in advance to grapple with them. Those who draw profits from the anticipation of future costs and income are doing something of which the term risk bearing as ordinarily used is scarcely descriptive. It would be more illuminating of their real function in society to say that *they capitalize their prevision of the future.*

Consider for example the function of underwriting new ventures discharged by the investment banks. When they believe that timely provision of capital equipment along novel lines promises lucrative returns in the future, these banks guarantee to provide the necessary supplies of present funds, partly out of their own resources, partly by inducing others to take a share in the venture. Through their efforts capital instruments are constructed and assembled, managerial and manual labor is organized, the different parts of a productive establishment are brought into coöperation with each other. If the forecast is correct, this productive unit will be equipped and ready for service at the time it is needed. The bankers may then withdraw and surrender the actual ownership and management of the new enterprise to others. But they will have capitalized the opportunity for profit by retaining for themselves shares of stock in the enterprise, whose dividends absorb a large part of the pure profit which accrues.

The process we have just described is known in the business world as *promoting* new industries. Professional promoters, called into existence by the dynamic character of modern industry, devote their whole energy to forecasting the currents of change at work and to placing themselves in position to profit by an excess of future income from the sale of goods and services over the expenditures entailed in costs of production. A new railroad is built across a stretch of sparsely settled prairie at a cost of scores of millions. Its promoters are predicting a thickening of population in the region, a shifting of the geographical location of industry, a new relation between producing and consuming markets. If these changes occur, the promoters will have rendered society a

service by providing capital instruments of immense worth in the new setting, and they will profit accordingly. Another foresees that the need of a great nation for meat products cannot long be satisfied piecemeal from local sources of supply. He constructs a plant for the curing and preserving of meats, perfects the technical processes, elaborates the buying and selling mechanism, and profits with the coming of the conditions which he has foreseen. Illustrations of the promoter's activity could be multiplied indefinitely, for they lie ready at hand.

But we must not allow the glamour of "big business" or the spectacular risks borne by the great promotors to obscure the fact that the service of which we have been speaking and the profit which is its recompense are involved in the most commonplace and picayune of business operations. The little corner store which appears in the midst of a straggling suburban settlement signifies someone's faith in his foresight; it may be wiped out by a shifting of industries, or a change in transportation; it may flourish on an increasing need for the services of a middleman in the little community. Every independent enterpriser who embarks on a new line of productive service or adds to the existing facilities in the older lines is to some extent relying on his prevision of future events. Success will bring him excess of income over and above the necessary payments for his labor of management; failure will deny him this excess of income and may sweep away both the reward for his labor and his entire investment of capital.

Risk bearing not always a productive service. This brief discussion of the source of pure profit is by no means intended as an explanation, much less as a justification, of all income to which the word profit is ordinarily applied. The forecasting of future events and the bearing of the risks involved in them are indispensable and productive services. But this is not to say that all risk is an inevitable outcome of social forces and that risk bearing as such is always productive. Risk is sometimes deliberately created so that men may profit by

it at the expense of others. The gambler, including the ignorant player of the stock market, certainly takes a risk, but he does not thereby serve society. The man who "rigs" the market or spreads false reports about economic conditions multiplies the risks of other producers and may draw an income from their losses. This income he will call profit, but it is clearly a different case from the income of the man whose foresight prepares him to serve the future wants of society or to reduce its future costs. Mere foresight and the willingness to take a chance are not always a social service. The monopolist uses foresight for the purpose of exploiting the necessities of men by contributing to the scarcity of goods and services. The man who preëmpts a choice section of land is exercising foresight, but he has done nothing to ease our struggle for existence. Investment bankers have on occasion engaged in promotion activities for the purpose merely of creating an issue of securities for sale to the public without regard to the economic soundness or the social usefulness of the enterprise whose name the securities bear. In doing this they may display shrewdness above the average, but their own enrichment is no evidence that they have rendered a productive social service.

If we appear to disregard these cases of unproductive income parading under the name of profit and of true profit resulting from anti-social activities, it is not through ignorance of their existence and of the evil they may cause. Our purpose is to emphasize the normal and the essential. It must be evident that the need of prevision of the future and of risk bearing is inherent in the very nature of our economic order and that profits are the income resulting, to those who are successful, from these services.

Profits tend to be short-lived. Pure profit, in each individual case, tends to be a temporary income. The forces of competition work incessantly toward a readjustment of prices which will wipe out the margin of income above costs. Chance gains resulting from lucky ventures or the caprice of fortune

are by nature limited in duration to the continuance of the special events which have produced them. Special devices or methods of management, by which individual business men avoid loss from routine business hazards, become known to the trade in general and lead to an adjustment of prices. Even the larger flows of profits resulting from far-reaching changes in industry are of limited duration. A new invention will bestow profits on its first possessor by reducing his costs below those of his competitors. But when it becomes the common property of the trade, all costs will fall and the maladjustment will disappear. The long-run effect of improvements in machine equipment and in processes of manufacture is not to expand the profit income but to increase the real incomes of all groups in society. Even the forerunners in new fields of industry, who are the chief profit makers in the business world, are not long exempt from competition. Those first comers may withdraw with large amounts of capitalized profits before the levelling tendency of competition has brought prices and costs into adjustment, but the industry as a whole soon ceases to be a source of profit to those who exploit it.

Profits not a price determinant. Pure profit is not one of the price-determining costs of production. This income may not exist as a positive quantity for industry as a whole over a term of years. The service of the profit maker is induced by the chance of receiving the reward, not by the certainty of it. In any hazardous undertaking it is the anticipation of the reward which induces the effort of the competitors. In athletic contests, a winner's prize need be given to but one of the contestants. It is doubtless true that the service for which pure profit is the reward would not be rendered if none of the enterprisers were able to make a profit. But by holding forth chances of gain to individuals, society is supplied with the risk-bearing service of the enterprisers as a class without guaranteeing a reward for this service in any particular case. Recalling again that we speak only of competitive industry, the following statement, analogous to the case of economic rent, can be

made: "Prices are not high because profits exist; but profits exist because for other reasons sale prices are sometimes higher than costs."

Criticism of the profit system. Yet it is probably true that no feature of the capitalistic order of society has been subjected to more aggressive criticism from the advocates of social reconstruction than the system of profit industry. The more superficial of these criticisms are aimed at the size of the profit income. It is carelessly assumed that the profit income, both as a whole and in individual cases, is so large as to create great and unjust inequalities in the economic conditions of the different social classes. That this income is large in certain individual cases we have had occasion to note; and it may be true that the social consequences of even a few such cases present a serious problem. But the customary indictment of the profit income on the score of its size fails to take adequate account of the burden of losses borne by the same social class which is the recipient of profit.

The more weighty criticisms of the profit system however are concerned not with the size of the income but with the prerogatives and powers of the profit-getting class. The receipt of profits is associated with ownership and control. This allocation of power is inevitable in the present setting. The risk bearer must be the commanding officer of business enterprise, else he is powerless to perform his function. Foresight divorced from power can have no practical influence over the adjustment of the industrial structure to changing conditions.

Nevertheless the arrangement is open to attack. The profit seeker in control of an enterprise is tempted to take a distorted view of its social purpose, looking on it merely as an agency of self-enrichment. Such control of the productive system may operate to the injury of the workers and consumers whose economic welfare is dependent on its functioning. Moreover the owner is not the only risk bearer in a typical industrial establishment. The wage earners in particular are subjected to severe hazards from the uncertainties of business enter-

prise; they stand in the case of failure to lose their income through the loss of their jobs. The association of control with ownership draws a line of division between the profit getter and the worker which places the latter under the direction of the former and, without giving him a share in control, renders him liable to injury through mistakes in judgment on the part of the owner or his representatives.

In connection with our study of the other shares of distribution — rent, interest, and wages — we found that certain features of these institutions of our society gave rise to organized movements for change. Criticisms of the profit system provide additional stimulus to these movements. In a later chapter we shall examine in detail the programs of reconstruction to which these criticisms have given rise.

EXERCISES

1. Mr. X formerly was employed as department head of a manufacturing enterprise at a salary of $6,000 per year. He had $20,000 of savings invested in bonds which yielded him 5 per cent interest. He resigned his position and set up an enterprise of his own, investing his savings in the new enterprise. At the end of the first year his net income from the business amounted to $6,200.

   (*a*) What were his pure profits? Explain.

   (*b*) If his net income continued at this rate for ten years, would it be worth while for him to continue in this enterprise? Explain.

   (*c*) If this rate of return were typical of earnings of the industry as a whole, what would be the trend of future development in the industry? Why?

2. Suppose that Mr. X in the following year receives a net income from the enterprise of $10,000.

   (*a*) What now are his pure profits? Explain.

   (*b*) If this rate of return were typical of earnings of the industry as a whole, what would be the trend of future development in the industry? Why?

3. A new type of machine costs $1,000 and yields an annual net return of $120 on the investment. If the prevailing rate of interest is 5 per cent, what amount of pure profits will result from the use of the machine? Under conditions of competition, will the income from the new machine tend ultimately to be equalized with the income from other machinery? How?

CHAPTER XIX

Personal Distribution

Functional shares in distribution. In the preceding chapters the problem of distribution has been regarded from the functional point of view. This has involved a classification of the total income fund of society into distinct categories or classes, each category corresponding to some peculiar productive function performed by the recipient of the income or by the wealth which the recipient owns. We may now summarize the result of this analysis. The broadest classification of income from the functional viewpoint presents two major types of income: income from *ownership* and income from the *labor* or *service* of the recipient. Each of these chief groups contains a number of sub-groups as set forth in the following summary:

FUNCTIONAL SHARES IN DISTRIBUTION

Sharers	*Name of income*
I. Owners of productive instruments	
1. Landowners	Economic rent
2. Owners of capital	Interest
II. Performers of human service	
1. Laborers (in the broad sense)	Wages (in the broad sense)
(1) Employees	Wages and salaries
(2) Owner-managers (as workers)	Wages of management
(3) Professional workers	Fees
2. Enterprisers (as such)	Profits

Total and average national incomes. It is obvious that the study of *personal distribution* is largely a statistical matter. Lacking an accurate census of individual ownership, the first step in the problem is to devise means of getting the statistics together, using such sources of information as are available. One crude method of obtaining this information is to estimate the total national income and strike an average by dividing this figure by the number of people in a country. The following table presents statistics of distribution in this form:

TOTAL AND PER CAPITA INCOME OF THE UNITED STATES (1799–1938) [1]

Year	*Total income (millions)*	*Total income at 1926 prices*	*Per capita income*	*Per capita income at 1926 prices*
1799	$ 677	$ 1,092	$131	$211
1809	915	$ 1,423	130	202
1819	876	1,576	93	168
1829	975	2,083	78	166
1839	1,631	3,282	98	197
1849	2,420	5,450	107	241
1859	4,311	9,212	140	300
1869	6,827	8,843	180	233
1879	7,227	15,442	147	315
1889	10,701	23,780	173	385
1899	15,364	34,142	205	456
1909	26,456	48,102	292	530
1910	28,166	49,675	305	538
1911	28,104	50,096	300	535
1912	29,422	50,294	309	529
1913	31,450	53,761	326	557
1914	31,213	53,356	319	545
1915	32,533	54,042	327	544
1916	38,739	56,636	384	562
1917	46,376	57,043	454	558
1918	56,956	62,044	550	599
1919	62,945	62,199	599	592
1920	68,434	60,615	642	569
1921	56,689	59,485	524	550
1922	57,171	61,873	520	563
1923	65,662	68,044	589	610
1924	67,003	69,004	592	610
1925	70,051	70,474	610	614
1926	73,523	73,523	631	631
1927	73,966	73,966	626	626
1928	75,904	73,765	633	615
1929	79,498	75,929	654	625
1930	72,398	73,725	588	599
1931	60,203	68,647	485	553
1932	46,708	60,503	374	484
1933	44,713	59,301	356	472
1934	51,560	64,370	407	508
1935	56,254	66,337	441	520
1936	65,246	72,415	508	564
1937	69,419	73,693	537	570
1938 *	62,286	69,130	478	531

* Preliminary estimate.

[1] R. F. Martin, *National Income in the United States 1799–1938*, National Industrial Conference Board, 1939, pp. 6 and 7.

The figures given in the second column of the table show a wide variation of aggregate national money income over the course of years. Since the beginning of the present century there has been a steady increase up to 1929, interrupted only by a slight recession in 1921–23, then a striking decline until 1933, and an increase after that date through 1937. Per capita incomes were subject to similar fluctuations.

Since these incomes are stated in terms of money, it is necessary to take the changing price level into account when comparing the prosperity of one year with that of another. This has been done in the third and fifth columns of the table, which translate the yearly figures into dollars having the same purchasing power as in 1926. When thus translated into the terms of a stable price level, the changes in national and per capita income over the quarter century do not appear so striking, although substantial fluctuations still remain.

Needless to say, the figures for per capita money income do not represent the money receipts of any real individual. They are crude averages indicating only estimates of what each individual member of the population would have received in a given year if the national income of that year had been divided equally. Nevertheless these abstract estimates are of real importance in our study of personal distribution. They dispose of many false notions regarding the level of individual welfare which could be obtained by merely altering the existing pattern of distribution. In 1937 for example each person would have received only $537 if all had fared alike, and in other years absolute uniformity would have resulted in individual incomes of surprisingly small size. We may well bear this fact in mind when it is proposed to make everybody rich by forcible redistribution of the nation's total resources.

Income distribution in America. The year 1926 is generally taken as a norm in studies of income distribution because of the relatively stable conditions of that time as compared with preceding and subsequent periods. The following

table is based on income statistics for that year. The information is taken chiefly from income tax returns showing the actual money receipts of individuals in the United States who made declarations of their incomes to the federal government.

DISTRIBUTION OF MONEY INCOMES IN THE UNITED STATES IN THE YEAR 1926 [1]

Income class	*Cumulative percentage of persons*	*Cumulative percentage of incomes*
Under $1,000	2.89	0.31
$1,000–$2,000	28.15	8.27
$2,000–$3,000	48.40	17.57
$3,000–$5,000	78.37	39.56
$5,000–$10,000	91.92	57.24
$10,000–$25,000	97.88	73.91
$25,000–$50,000	99.27	82.81
$50,000–$100,000	99.776	89.14
$100,000–$150,000	99.881	91.74
$150,000–$300,000	99.960	94.75
$300,000–$500,000	99.982	96.30
$500,000–$1,000,000	99.994	97.75
$1,000,000–and over	100.00	100.00

This table does not really tell us how much money any specific individual or group of individuals received in 1926. It would appear from the figures given that approximately half of the people (48.4 per cent) received incomes of $3,000 or less. We must remember however that the "percentage of persons" listed in the second column refers to individuals who acted as the declarers of income. The income unit in modern society however is the family, not the individual. In many cases one income tax declarer represents an entire family among whom the reported income must be divided. On the other hand, many families possess more than one income receiver. Other studies have attempted to depict the distribution of income among the families of this country with a result as shown in the following table.

[1] United States Treasury Department, Bureau of Internal Revenue, *Statistics of Income from Returns of Net Income for 1926*, p. 4.

DISTRIBUTION OF INCOME BY FAMILIES [1]
(The United States 1910 and 1929)

Income Class	*Per cent of families*		*Per cent of total national income*	
	1910	1929	1910	1929
Richest Families	2	2	20	27
Middle-class Families	33	33	41	44
Upper Middle-class	18	18	27	29
Lower Middle-class	15	15	14	15
Poorest Families	65	65	39	29

The most striking feature of the pattern of distribution in America is the extreme inequality which it portrays. Under conditions of complete equality any 10 per cent of the people would receive 10 per cent of the national income, any 50 per cent of the people would receive half of the income, and so on. A glance at the table of individual incomes will show how far from this line of uniformity is the actual pattern. As things were in 1926, the lowest income class included 2.89 per cent of the people and received only 0.31 per cent of the total income. A group of about the same size at the top of the scale (the 2.12 per cent receiving over $25,000 a year) got 26.09 per cent of the national income. The richest 1 per cent received 18 per cent of the total income, while at the bottom of the scale a section of income similarly large was shared by nearly half the people; in other words, the poorest 48.40 per cent of the population had a share a little smaller than that of the upper 1 per cent. In whatever manner the table is analyzed, this condition of inequality is thrown into sharp relief.

The figures of family distribution not only call attention to extreme inequality but appear also to indicate that during the period under review this inequality was increasing. The richest families, 2 per cent of the total number of families, received 20 per cent of the whole national income in 1910

[1] Taken from C. R. Daugherty, *Labor Problems in American Industry*, Revised Edition, 1938, p. 155.

and 27 per cent in 1929. The middle-class families, approximately a third of the total, showed less variation, their share of the whole income increasing from 41 per cent in 1910 to 44 per cent in 1929. The 65 per cent of the families standing at the bottom of the income scale received 39 per cent in the former year, 29 per cent in the latter. Too great reliance should not be placed in these figures as evidence that the rich are growing richer and the poor poorer over long-run periods; not only is the time covered by the table too brief to show a permanent trend, but it was disturbed by war and its aftermath and phenomenal technological changes. However we are not concerned here with the developmental tendencies in modern society but rather with its general pattern of distribution at a given time.

Distribution of wealth in America. If we turn from a study of the distribution of income to that of wealth we shall discover a similar, but even more extreme, picture of inequality. Wage earners as a class receive a large fraction of the total national income, but most of them own only trifling amounts of wealth, and their total of property rights represents but a small portion of the nation's wealth resources. In seeking statistics of the distribution of wealth we encounter even greater difficulties than in the case of income. There is of course no census of ownership to be drawn upon for information regarding this problem. The most useful sources of data are the records of the probate courts, through which estates pass upon the death of their owners. The statistics assembled in the following paragraphs are taken from these records.

It would be well before examining the data to point out certain inaccuracies necessarily resulting from this method of estimating the distribution of ownership. In the first place, an individual's accumulation of wealth is not always transmitted as a whole upon his death; many family estates are divided up in whole or in part during the owner's lifetime. This practice will cause the estates passing through the pro-

bate courts to be smaller on the average than are the property holdings of individuals still living. Secondly, it must be remembered that the average age of those who die in a given year is higher than the average of the living population, and the decedents have therefore had a longer time in which to accumulate wealth. This factor will have an effect opposite to that just mentioned; it will cause the record of the probate courts to exaggerate the size of individual estates. Because of such uncertainties these records cannot be used with safety to estimate the actual amounts of wealth owned by different social classes at any given time. Our real interest in the problem however is to depict the character of distribution rather than to discover the absolute size of different individual estates, and the data we are using are significant for this purpose.

The following table gives a simple analysis of wealth distribution as shown by the probate records of two American states, Massachusetts and Wisconsin, on the dates given. The figures listed in the fourth column are of some interest as indicating the average amount of property held by the respective categories of the population: the poorest, lower middle, upper middle, and richest classes. However for the

SECTIONS OF THE POPULATION CLASSIFIED ACCORDING TO WEALTH OWNERSHIP[1]

Section of population	*State*	*Per cent of total wealth*	*Average value of wealth*
Poorest 65%	Mass. 1891	4.5	$ 399
	Wis. 1900	5.2	381
Lower Middle	Mass. 1891	3.9	1,499
Next 15%	Wis. 1900	4.8	1,524
Upper Middle	Mass. 1891	32.8	10,509
Next 18%	Wis. 1900	33.0	8,730
Richest Class	Mass. 1891	58.8	169,550
Next 2%	Wis. 1900	57.0	135,715

[1] W. I. King, *Wealth and Income of the People of the United States*, 1915, p. 79.

reasons given above they are not entirely reliable. What is more important for our purposes is a comparison of the percentages given in the first and third columns of the table, for it is this that discloses the extent of inequality in wealth ownership.

If wealth had been divided evenly in these states on the dates given each fraction of the population would have owned an equal fraction of the wealth which passed through the probate courts. Instead of this situation we find that the poorest two-thirds of the people had a combined ownership of 4 or 5 per cent of the wealth. The lower middle class, 15 per cent of all the people, owned a somewhat smaller proportion. The richest 2 per cent held nearly three-fifths of all the wealth. Or to throw the figures into somewhat different arrangement, the distribution can be given very simply as follows: 80 per cent of the people appear to own 10 per cent of the wealth; the remaining fifth of the people have all the rest — 90 per cent of the total.

For the reasons given the evidence of probated wills probably exaggerates the extent of inequality in the distribution of wealth. Nevertheless the most careful studies of the problem using other data show that most of the wealth of this country is owned by a minor fraction of the people. Thus in 1927 it was found that wealth was distributed substantially as shown in the following table, whose figures have been grouped so as to make them comparable with the figures of income distribution presented earlier in this chapter.

ESTIMATED DISTRIBUTION OF WEALTH, UNITED STATES, 1927 [1]

Property class	*Per cent of population*	*Per cent of total national wealth*
Richest Class	2	40
Middle Class	33	45
Poorest Class	65	15

[1] W. I. King: Wealth Distribution in the Continental United States, *Journal American Statistical Association*, June, 1927.

Interrelation of wealth and income distribution. In the foregoing pages we have considered the distribution of income and that of wealth as separate and distinct phenomena. It must be obvious to all however that the two things are interrelated in many ways. In the first place, unequal distribution of wealth is one of the causes of the inequality of personal incomes. Most of the large aggregations of wealth are productive of income in the sense in which we have been using the term; *i.e.*, money receipts. The small property holdings of the poor are almost invariably non-productive in this sense; they consist of household and personal effects. The distribution of income in any society would be much more nearly uniform if income-bearing property were shared equally. This point is so self-evident that it requires no elaboration. Still it may be of interest to present some definite information upon it with reference to the situation in the United States. The Bureau of Internal Revenue has published an elaborate analysis of the percentages of their total incomes drawn from property and labor respectively by various classes of the population in 1930. These facts are summarized in the following table:

PROPORTIONS OF INCOME FROM EARNINGS AND PROPERTY RESPECTIVELY, 1930

Income group	*Per cent of total income provided by earnings*	*Per cent of total income provided by property*
Over $1,000,000 a year	2.2	97.8
$25,000–$50,000 a year	29.5	70.5
$ 5,000–$10,000 a year	62.2	37.7
$ 1,000–$ 2,000 a year	82.2	17.8

These figures seem to show that incomes would still be very unequal, though much less so, if income-bearing property were not privately owned. It should be borne in mind however that the portion of the rich man's income which he attributes to labor is often very arbitrarily determined. The owner of a large enterprise can assign himself a large salary,

leaving proportionately less to accrue to him in the form of dividends.

Unequal division of wealth also affects differently the security or stability of income at various levels in society. The propertyless man has all his eggs in one basket; when he loses his job his income ceases entirely. Incomes drawn from property are by no means free of uncertainties, but the possession of such an income gives the individual an offset against the loss of his wages or salary, and the larger property holdings are usually distributed over different types of investment, thus reducing the risk of fluctuation. The stability of income therefore tends to vary inversely with the size of income from top to bottom of the social scale. As we shall see in another place, insecurity and the fear it arouses have a powerful psychological effect on the wage-earning classes, causing them as individuals and as organized groups to adopt practices which are often injurious to the general social welfare.

Effect of inheritance. Unequal distribution of wealth tends to perpetuate inequality of income during successive generations. Under the free play of economic forces individual incomes would differ in size even if all were drawn from labor. They would be determined by the wage rates for different types of economic service and would vary in accordance with the importance of the individuals as producers of economic goods and services. But in this situation the individuals of each generation would start afresh in the competition for income, and inequalities between income classes would not be handed on from generation to generation. As things are, the transmission of property by inheritance prevents the adjustment of incomes to the productive services of the members of society; some are born with rights to income. It is of course in the higher income classes that income-bearing property is transferred from father to son. The portion of income which can thus be transferred at different social levels is indicated in a preceding paragraph by the varying percentages of income from property in these different classes. The lines of

division between classes tend to remain fixed over long periods of time. Not only is each generation classified in unequal income groups, but these groups — particularly at the higher levels — tend to consist principally of the same families. It is sometimes said that the legal institution of inheritance is responsible for this perpetuation of inequality; but this is not the basic cause. The right to transmit property by bequest is available to all social classes. It is because the distribution of wealth at any given time is unequal that inheritance prolongs inequality of income from age to age.

The problem of poverty. We have been discussing the problem of inequality, which, as has been pointed out, is not the same thing as the problem of poverty. The question of the actual welfare of the people is however an important aspect of our subject. What for example is the grand total of income available to the people of the United States? How well off would each person be if all had an equal share? How well off is the mass of people under the existing scheme of division?

The first two questions are answered by the table given on page 318. To answer the third question adequately we need three types of information: (1) the average size of families; (2) a standard of living, representing the minimum necessary for health and decency for American families, against which to measure the conditions which actually prevail; (3) reliable statistics of the actual money receipts of these families. A word of explanation should be given regarding these three points before proceeding with our inquiry.

The family unit. Division of income in our type of society is carried out by families, while statistics of income, such as wage rates, are given in terms of individuals. To appraise the adequacy of a wage earner's income we need to know how large is the family which relies on it for support. The size of the family is also an important element in our computation of the standard of living. Now in such studies as this it is customary to assume that the average family consists of

five persons: mother, father, and three children. This is justified on the ground that three children per family are required to maintain the level of the population, and of course social income could hardly be considered adequate if it failed to keep the population as a whole alive from generation to generation. It is also assumed that each family is supported on the earnings of one member — the breadwinner — and has no income aside from his labor. These assumptions are made in the statements which follow. They are, as a matter of fact, rather wide of the truth. Investigations have shown that 27 per cent of all adult males gainfully employed are unmarried; that two-thirds of our wage-earning families have less than five members; that half the families obtain supplementary earnings from others than the chief breadwinner; and that wage earners as a whole draw between 5 and 8 per cent of their income from property.[1] These are important qualifications, tending to mitigate considerably the condition of poverty pictured by the figures that will be given.

The "standard of living." There is no absolute "standard of living." When we say that a certain minimum of money income is needed to support a family we always reflect our preconceptions of what decent living should be; and these preconceptions vary from time to time and from country to country as conditions and customs change. Nevertheless it is customary to measure the actual economic condition of families against an objective standard, determined by carefully estimating the kinds and quantities of food, clothing, housing, and other items of consumption and of saving required, and then resolving these estimates into money terms at prevailing prices. As a matter of fact, several different standards are usually recognized at a given time, the most important being the *minimum-health-and-decency* standard and the *minimum-comfort* standard. The former provides respectable

[1] Should this statement appear in conflict with the last line of the table on page 325, the reader need only recall that the earlier table relates to those who filed United States income tax returns, a group which represents only a very small fraction of the wage-earning class, and the wealthiest part at that.

housing, nourishing but inexpensive food, adequate and decent clothing, and a small surplus for such items as cheap recreation, a newspaper, and the like, but no saving for emergencies. The latter provides these things somewhat more generously, particularly in the case of housing, and leaves a margin for secondary education, insurance, and medical care. Incomes slightly below the health-and-decency standard are interpreted as bare subsistence; anything less than this is recognized as outright poverty. The minimum-comfort standard corresponds roughly to what is meant by the expression, the "American" standard of living.

The costs of these standards vary from year to year with price fluctuations; they differ at any one time in different localities, and needless to say they will differ too for families of different sizes. The costs of the different standards, estimated for a family of five in the years given, are shown in the table below. The subsistence level omitted from this table may be assumed to lie in the area between the poverty and the health-and-decency standards.

ESTIMATED COSTS OF DIFFERENT STANDARDS OF LIVING FOR A FAMILY OF FIVE

Year	*Poverty level*	*Health and decency*	*Minimum comfort*
1910	$ 570	$1,130	$1,700
1918	930	1,860	2,780
1929	1,000	2,000	3,000
1932	810	1,610	2,420
1935	850	1,700	2,550

Family incomes. Actual family incomes are difficult to discover. Wage rates do not give us this information in accurate form, since total earnings are affected by unrecorded factors, such as loss of work through illness and unemployment, overtime earnings, and supplementary income from other sources. The figures we shall use are based primarily on wage rates, checked wherever possible by direct inquiries among wage-earning families regarding the actual size of

their incomes. Statistics relating to this problem take acccunt of price fluctuations over periods of time, but otherwise they necessarily deal in averages which conceal local differences in living costs and actual variations in the size of families. Accordingly the data presented in the following paragraphs [1] relating to living costs and average earnings of American working-class families are merely suggestive of the conditions which obtain. As averages they are subject to the errors just mentioned, and they also fail to take account of many supplementary incomes referred to in a preceding paragraph. They do not tell us precisely how many actual human beings in this country were living on these planes in the years given.

Taking but two dates in recent history, 1914 and 1920, we discover the following facts regarding the economic well-being of the mass of people in the United States. In 1914, the minimum standard of living for laborers' families in New York City, according to an estimate of the city government, cost $845 a year. For the state as a whole another official estimate placed the minimum cost at $876. For the country as a whole the cost, according to an independent estimate of a careful student of the problem, was $825. We have three estimates of the actual family earnings of industrial workers at this time, one for New York State, a second for Pennsylvania, a third for the entire country. In New York State industrial and office workers could earn $649 if fully employed. In Pennsylvania the actual earnings of wage earners averaged $720 a year. A nationwide survey of actual incomes showed that four-fifths of the heads of families in the wage-earning population earned less than $800 a year.

By 1920 the cost of living had risen. Estimates of incomes needed to support a minimum standard of life varied from $1,700 for an irreducible standard of health and decency to $2,000 for a standard of reasonable comfort. In this year it was found that factory workers in New York State could earn about $1,500 a year if employed full-time. Throughout the

[1] From C. R. Daugherty, *op. cit.*, pp. 138 and 144.

country as a whole workers of all grades of skill were receiving family incomes which averaged $1,455. The average family incomes of skilled factory workers were approximately $1,700.[1]

The dates we have taken for purposes of illustration were times of fairly normal industrial activity, although it is to be noted that estimates of living standards for 1920 are abnormally high on account of the very high price level in that year. Between 1920 and 1929 earnings rose somewhat more rapidly than living costs, and the working population was accordingly more prosperous. Since 1929 poverty has of course increased; but the times have been too disrupted by the business depression to serve as an index of the general welfare.

The table which follows records the distribution of all families, regardless of size, among the different planes of living, as estimated by one student of the problem.[2]

CLASSIFICATION OF FAMILY INCOMES
(Percentage of All Families regardless of Size or Class)

Year	*On poverty planes*	*On subsistence planes*	*On health and decency planes*	*On comfort planes*
1910	15	55	20	10
1918	15	45	25	15
1929	20	35	20	25
1932	40	35	15	10

The figures given suffice to show that even in a country as rich as the United States, certainly the most prosperous of the modern world and probably more opulent than any other society in human history, poverty is the lot of the lower income classes.

Causes of poverty and the remedies. Poverty, if confined to isolated individuals or families, would not be recognized as a social problem. It would be set down to the indolence or

[1] For a comprehensive study of incomes and living costs, see A. Epstein, "American Wages, 1890–1920, and the Standard of Living," *Annals of the American Academy of Social and Political Science*, September, 1921.

[2] C. R. Daugherty, *Labor Problems in American Industry*, Revised Edition, 1938, p. 144.

thriftlessness of the sufferer, or else attributed to hard luck. Remedies, if any were offered, would take the form of philanthropy and charitable relief, coupled perhaps with exhortations to do better in the future. But widespread and persistent poverty is acknowledged to be a reproach to our civilization, repugnant to the prevailing standards of an age which prides itself on the efficiency of its productive system. It is poverty in this form with which we are dealing here. In seeking remedies for the problem of poverty the distinction between occasional distress and the inadequate income of a whole social class should be carefully observed. To apply to this latter problem the correction — charitable relief — pertinent to the former will not remove the cause of the evil but merely tend to make a bad condition worse. Indiscriminate charity administered on a broad scale by public officials or others equally remote from the individual case demoralizes the beneficiaries and perpetuates the problem. No remedy will be truly effective in the social sense unless it is made a part of the productive system, operating regularly to remove the causes of poverty.

These causes may be grouped in three major categories: (1) the relative inefficiency of the productive system itself; (2) the hazards to workers involved in the modern process of production; (3) the obstacles which impede the rise of individuals from the lower to the higher wage groups. The first of these requires little explanation. Our inquiry has shown us that there is a surprisingly small per capita income even in the most prosperous countries. Any improvement of technique or organization which increases this average income will tend to relieve the condition of the poorest class even if incomes continue to be unequally divided. Conversely any program of reform, even if intended for the immediate benefit of the poor, if it reduces the efficiency of the productive system as a whole most ultimately aggravate the problem of poverty. These are truisms; unfortunately they are frequently forgotten by men of philanthropic or reforming spirit.

By the "hazards of industry" we do not mean sporadic cases of bad luck which bring hardship to individuals here and there, but rather the widespread risks of the worker's life inherent in the very nature of our machine age and bearing generally on a whole social class. Occupational disease, industrial accident, premature death, and unemployment are commonplace incidents in the lives of wage earners. We shall study these problems in detail in another place and there examine practical measures for dealing with them. Here it will suffice to state that in the aggregate they place an enormous financial burden on the lowest income classes of society, tending to reduce their average earnings below the poverty line. Certain of these hazards are insurable risks against which provision could be made by methods well established in business practice. Moreover they must be viewed as costs of production properly to be borne by society as a whole. A system of compulsory insurance is the logical remedy for this cause of poverty.

To remove the obstacles which pen the mass of people within the lowest income classes we must rely upon education and vocational training. When a large section of the population is underpaid it is not because the economic function they perform is of little value in itself. On the contrary, the labor of society's machine-tenders, clerical workers, and even unskilled laborers is indispensable in the sense that the productive system would collapse if these functions were not rendered. The fault lies not in the vocation but in the number of people who occupy it. Any policy whose effect is to reduce the numbers in the low-wage groups by transferring some of them to less crowded occupations will raise the lower incomes and reduce the higher, at the same time increasing the average. Opportunities for training provided at public expense and available to all will tend to do this and to leave in the low-wage occupations only those people who are actually incapable of acquiring greater skill. Americans are prone to believe that they are already making ample provision for

education. But this is far from the case. There is lack of a sound and comprehensive system of vocational guidance for the young. There is no thoroughgoing policy of adult education or the retraining of individuals already lodged in overcrowded occupations. The general program of public education is largely a rule-of-thumb affair, too little adjusted to the needs and capacities of future workers. Much might be done along these lines to combat the poverty of the many.

The problem of inequality. Extreme inequality of material welfare within a nation gives rise to serious social problems. These are not necessarily problems of poverty. Even if the poorest sections of the people should escape actual privation, the startling contrast between their lot in life and that of their more fortunate fellow citizens would introduce an element of instability into the relationships of social classes within the country. This is especially true in societies whose governmental system is founded on the principle of individual equality in political matters.

Especially does extreme inequality in the distribution of wealth tend toward instability of the social order. Our legal system is definitely concerned with the protection of property rights; the fundamental law of our government provides that each owner shall be protected in the possession of his wealth. What this really means is that all the people are called upon to defend the rights of those who own the wealth. A political and legal system based on these principles is secure only as long as the mass of people retain a direct, personal interest in the institution of private property. It is difficult to keep this regard for the rights of property alive in a social group unless the majority of its members have property of their own or can hope to acquire some during a lifetime of labor. When this ceases to be true, proposals of politicians or social reformers that wealth be redistributed by authority may win a ready hearing from the rank and file of the people.

It is not only the radicals who condemn extreme inequality and hope to do something to reduce it. To many people who

approve and wish to retain the basic institutions of capitalism the present pattern of distribution seems essentially unjust. Others not moved by such sentiments but concerned to maintain the stability and smooth functioning of the present order recognize inequality as a menace to their own welfare and that of society as a whole. It is now generally admitted by thoughtful people that arrangements which tend to accentuate inequalities between social classes are likely to be impermanent. They must either be modified by timely conservative reform, or the resentment they engender among the people may ultimately destroy them by force. The impulses toward conservative reform of inequality are stimulated by the growing conviction that the condition is not being improved by the automatic development of economic forces. Careful studies of the trend of distribution in the capitalistic order over periods of time have shown that the best that can be said of the matter is that inequality is not increasing.

The reform programs which call for complete transformation of the social order will be reserved for later discussion, but as evidence that reform of inequality may take place within the framework of capitalist society we should note that conservative governments have begun to adopt measures looking toward a greater diffusion of income and wealth. Individual liberty and free enterprise are no longer uncontrolled and unlimited. Many business practices which were permissible a generation ago are now forbidden because they enrich the strong or the unscrupulous at the expense of other members of society. In certain cases governmental bodies have taken over the ownership and operation of certain forms of enterprise which were formerly a source of private profit. The powers of the monopolist are more and more circumscribed by public regulation of his prices. The weakest bargainers in the laboring classes are protected in many countries by minimum wage laws which prevent their exploitation. This many-sided movement toward restriction of individual freedom in economic affairs is tending to prevent the enrich-

ment of individuals by methods which are condemned by prevailing standards of justice and fair play.

Similarly property rights have been modified by the movement toward greater equality of income and opportunity. Graduated income taxes,[1] in conjunction with the expansion of governmental expenditure along lines of general public service, result in an indirect transference of income from the rich to the poor. Inheritances are subject to graduated taxes, which tend substantially to reduce the largest fortunes. All of these methods of modifying the operation of the industrial system in the direction of greater equality of income and opportunity may be illustrated from the practices of any modern government.

EXERCISES

1. Refer to the table on page 318.
   (*a*) Draw, on the same diagram, a curve to indicate changes in per capita money income and a curve to indicate changes in per capita real income (*i.e.*, money income corrected for changes in the price level).
   (*b*) Which shows the greater fluctuation? What was the per cent decline in per capita money income from 1929 to 1932 and what was the corresponding per cent decline in per capita real income?

[1] Income taxes whose rates become higher as the taxable income increases. Cf. Chapter XXV.

CHAPTER XX

Principles of Money

Some historical examples. Today as in the past money takes a variety of forms. There are coins of gold, silver, nickel, bronze, and other metals. There are pieces of paper inscribed with promises and other statements. Cattle and sheep have been used as money, as well as wheat, corn, tobacco, musket balls, beaver skins, fish, meat, shell beads, gold dust, and so on. Monetary history discloses that each community adopts as its money some commodity or commodities of common use among the people, something that is desired by most of the people most of the time.

One of the earliest forms of money of which we have knowledge is the sheep and cattle of peoples on the pastoral stage of economic development. Most of the people derived their living from their flocks, and these represented their principal form of wealth. One would always welcome an addition to his flock, and thus it became the natural thing to use the domestic animals as money. The word "pecuniary," derived from the Latin word *pecus* for cattle, is an interesting heritage of the early stage in which cattle and money were in a sense synonymous.

The common characteristic. All these forms of money have one feature in common—general acceptability in exchange for other things. The things used for money acquired this characteristic originally from the fact that they were desirable for their own sake; *i.e.*, they served to satisfy human wants. In certain cases, it is true, the thing used as money could not be applied directly to the satisfaction of wants. An example is the wampum used by the New England colonists. Though directly useful to the Indians, it satisfied directly no want of the colonists. Yet it became useful as money as soon

as the habit of accepting it in exchange became established generally throughout the community.

For the individual, it is the fact that money is generally acceptable which makes him willing to accept it when offered in exchange for his goods or services. If a person who has valuable goods to sell will take in exchange a coin of gold, or a piece of paper engraved with the promissory note of a bank, or a bale of tobacco leaves, it is not ordinarily because he has any need of these particular articles. He accepts them simply because he knows that at any time he can find other people willing to take them in exchange for whatever they may have to sell to him. This is the fundamental characteristic of money; so long as this qualification is met there may be the utmost diversity as to other features.

The general acceptability of money is always more or less local. Most American merchants would refuse to accept a Bank of England note in payment for goods sold, and a Canadian quarter is not generally acceptable in America except close to the border. In general each nation has its own monetary system, and the money of one nation is usually not acceptable within the borders of other nations. By general acceptability we mean acceptability within the particular community.

Definition of money. We have now all the essentials of the scientific definition of money, which was stated but not explained in a previous chapter. *Money is anything which is generally accepted in a certain community in exchange for other things.*

Functions of money. The principal function of money is to facilitate exchange, in other words to serve as a *medium of exchange.* Without such a medium, exchange can operate only with great difficulty. The maker of shoes, for example, who wished to exchange some of his product for bread could succeed only when he had discovered the double coincidence of a baker who wanted to exchange bread for shoes. A similar problem would face the shoemaker in connection with every other good which he sought to obtain in exchange for his own

product, shoes. To exchange so conducted the name of *barter* is applied. The modern economic order, based on division of labor and exchange of products, could never have become established had not the introduction of money released exchange from the restrictions of barter.

A second essential function of money is to furnish a *standard of value*. This function follows necessarily from the first. It is inevitable that the unit of the good which serves as the medium of exchange should be the unit for measuring the values of all exchangeable goods. Thus money furnishes the universal unit by which alone all kinds of wealth and services may be measured and compared and totaled. This function of money is as essential to the existence of the modern economic system as is the presence of a medium of exchange.

Money performs a third function in serving as a *store of value*. In so doing it enables us to separate our sales from our purchases. One can sell what he produces for money and later according to his convenience buy the various things he needs; one can even, by borrowing money, buy first and sell later. The consumer "holds" a certain portion of his wealth in the form of money, as does also the producer.

Government control of money: In general. Money originated in each community naturally and unconsciously as the result of slowly established custom. People simply fell into the habit of accepting some one commodity in preference to others in exchange for goods and services which they had to sell, much as boys at a certain stage in their development fall into the habit of using marbles as a medium of exchange. Thus the monetary system became firmly established in the mores of the community, and eventually the government took cognizance of it and proceeded to define and enforce it. For centuries the control of the monetary system has been jealously guarded by kings and rulers as one of the most precious attributes of sovereignty. Today the system in all its details is determined by law, and it is this modern legal monetary system with which we are concerned. Government (1) speci-

fies exactly what kinds of wealth and property shall be money, fixes the monetary unit, defines the different kinds of money and their relations to each other, (2) coins the metallic money and engraves and prints the paper money, (3) in certain cases regulates the quantity of money, and (4) makes and enforces rules regarding legal tender.

Legal tender. While the general acceptability which is the essential characteristic of money grew up naturally through custom, modern governments have commonly found it desirable to enforce the acceptability of money by law. We thus have *legal tender*, which is defined as *any kind of money which according to law must be accepted when offered in payment of any obligation expressed in terms of the country's monetary unit.* When, as is usual, the monetary system consists of several kinds of money, the legal tender quality is often not the same with respect to the several kinds of money. The government may bestow legal tender on certain kinds of money and not on other kinds equally serviceable as media of exchange. In the United States prior to 1933, for example, gold coins and United States notes were legal tender, while national bank notes and silver certificates were not. The legal tender distinction is purely a matter of legal enactment and does not necessarily indicate superior intrinsic worth of the kind of money thus designated.

It should be made clear that the legal tender law is not concerned with contracts or other obligations not payable in money. Neither does it apply to contracts which by their terms are payable in a particular kind of money. There is nothing to prevent people making contracts payable in any kind of money they may agree upon. Only when the agreement is in terms of dollars (or other monetary unit), no particular kind of dollars being specified, does the legal tender law apply. It then serves, in case of dispute, to settle just what is meant by the word "dollar." Since practically all money agreements are thus generally stated or understood, legal tender laws are actually of wide application. Their normal purpose is (1) to protect the creditor from being forced

to accept anything but the best kind of money when different kinds are in circulation and (2) tc protect the debtor against a capricious creditor who might refuse to accept a perfectly good kind of money or insist upon receiving some form which the debtor could not conveniently provide. It may be observed in passing that the government has the power, by altering the monetary unit or amending the law as to legal tender, to introduce unforeseen changes in the fulfillment of contracts, both public and private, previously made in terms of money.

Money and currency. Although, as we have noted, it is the principal function of money to serve as a medium of exchange, the fact is that many exchanges are consummated by means of instruments which are not in the strict sense money. Bank deposits — more strictly, individual deposits subject to check — are a medium of exchange, and indeed it appears that in the modern industrial nations bank deposits bear the larger part of the burden of this fundamental monetary function.[1] It has been estimated for example that in the United States some eighty or ninety per cent of the total volume of exchanges is conducted by means of checks or similar instruments which transfer bank deposits. Such instruments do not qualify as money under our definition because they are not "generally acceptable." This makes it convenient to have another term, broader than money, which will include everything that serves as a medium of exchange. The term *currency* has been generally adopted. The complete picture of a *currency system* must include bank deposits and certain other forms of currency, such as postal money orders, express checks, etc. The subject of bank deposits will be treated at length in a later chapter.

In noting the great importance of bank deposits and other forms of currency which are not money, the reader will of course not forget that all forms of currency (except money it-

[1] This statement is particularly applicable to the United States, Great Britain, and other Anglo-Saxon communities, where the deposit system of transfer by check is very highly developed. In the countries of Continental Europe there is relatively less reliance upon deposits and more upon money.

self) are themselves rights to receive money; in other words, that the whole currency system is based upon money. This point may be expressed in the statement that, although all currency performs the functions of a medium of exchange and a store of value, only money serves as a standard of value.

Varieties of money and monetary systems. Monetary experience has developed several kinds of money. The term *standard money* is used to designate those forms of money which are commodities or articles of wealth, having value in their own right apart from their use as money; in other words, standard money has "intrinsic value" equal to its full value. In modern times standard money has usually consisted of gold or silver coins or gold or silver bullion.

Standard money is usually accompanied in circulation by certain other kinds of money, whose value is not based upon the materials from which they are made but arises from the right of the holder to redeem such kinds of money in the standard money. To all kinds of money which are freely redeemable in standard money we give the title of *representative money.*

Money which is neither standard money nor representative money is called *irredeemable paper money.* Its value, not imparted by the material from which it is made, arises from the fact that it circulates and is generally acceptable in the community, reinforced sometimes by the hope of ultimate redemption in standard money. In practice such money is generally representative money in form, being government notes or bank notes whose promise to pay is not being honored.

A monetary system which consists of standard money or of both standard and representative money is known as a *standard money system.* This is the system which generally prevailed in most of the civilized world for a century or two preceding the World War. In the United States a standard money system prevailed, with occasional interruptions, from 1792 until the abandonment of the gold standard in 1933.

Any money system not based on standard money may be

classed as an *irredeemable money system.* Such a system contains neither standard nor representative money, and no form of money is redeemable in anything of intrinsic value. Historically irredeemable paper money systems have usually developed out of previously existing standard money systems. During the American Revolution an irredeemable paper money system prevailed, based upon the Continental bills of credit. The United States was again on an irredeemable money basis during the greenback régime, from 1862 to January 1, 1879. Many European nations went over to irredeemable money during the World War, and most of the nations of the world are, in one form or another, on that basis today.

Standard money systems: Principal kinds of money. In our further description and analysis of money, it will be convenient to consider these two types of monetary system in order. The discussion immediately following will deal with the standard money system.

Experience has developed several kinds of representative money, and the following classification of the component parts of a typical standard money system may prove useful: [1]

1. Standard money
2. Representative money
    (1) Standard money certificates
    (2) Credit money
        (*a*) Government notes
        (*b*) Bank notes
        (*c*) Token money.

A standard money certificate testifies to the fact that the government holds standard money which will be delivered to the bearer of the certificate on demand. It is thus of the nature of the warehouse receipt, and the reserve held by the govern-

[1] Classification must not be taken too seriously. It is never to be regarded as an end in itself; its purpose is merely to facilitate investigation. Classification therefore must always be more or less arbitrary, and it is not contended that the classification of money here presented is the only one that might be set up and defended. It does commend itself to us as on the whole the most useful for the present purpose.

ment may never be less than the total amount of the certificates outstanding. The function is solely one of convenience; *i.e.*, to permit the use of the standard money without the necessity of handling the actual coins. The United States gold certificates, prior to 1933, were an example of standard money certificates.

The term *credit money* covers those forms which are of the nature of the promissory note. A *government note* is the government's promise to pay standard money or some other kind of money to the bearer on demand. A *bank note* is a similar promise of a bank. The United States notes (greenbacks) and the federal reserve notes respectively are examples in the United States.

Token money is credit money consisting of coins containing less than their nominal value of metal and redeemable in standard money. The token coins are usually made of silver or of other less precious metals, as copper, nickel, or bronze. When made of the same metal as the standard coins, the token coins are of proportionally less weight. Thus there is not as much silver in two half dollars, four quarters, or ten dimes, as in a silver dollar (the latter having formerly been a standard coin). Token coins are usually of smaller denominations than the monetary unit. Their function is to permit the "making of change"; *i.e.*, payments involving fractional parts of the monetary unit. Examples are the fractional coins of silver, nickel, and bronze in use in the United States. In the case of token money, the promise to pay is not stated upon the coin but is contained in the law requiring the public treasury to redeem the coin on demand of the bearer. Conversely the government will always give out the token coins on request in exchange for standard money.

Credit money (unlike standard money certificates) does not require the holding by the government of standard money equal in amount to the credit money outstanding; all that is required is a reserve sufficient to meet actual demands for redemption.

The ideal money commodity. Standard money is a commodity. As we have seen, many different commodities have at various times and places been employed as money. Yet as time has gone on most of these monetary expedients have been gradually eliminated, till in modern times the civilized world quite generally settled down to use of the precious metals, gold and silver, as the standard money. The explanation of this important development will appear if we inquire what qualities a commodity should have in order to serve successfully as a standard money and note how well gold and silver meet the specifications.

First of all, the ideal standard money must have *general desirability.* The precious metals have always been desired for the fabrication of ornaments, and in modern times they have also important industrial uses, as in dentistry. In the second place, gold and silver have *great value in small bulk.* This is an important consideration, since a medium of exchange should be susceptible of convenient use and easy transportation. Consider for example the inconveniences which would accompany the use as money of such otherwise useful commodities as wheat, coal, or salt.

Thirdly gold and silver have the advantage of *durability.* Age and exposure do not materially affect them, and they may thus be kept indefinitely without danger of loss from physical deterioration. This is of course an important matter, if money is to perform its function as a "store of value." The Virginia colonists for example frequently suffered loss through the deterioration of their stores of tobacco money. Another desideratum of a convenient money commodity is *uniform quality.* All the possible advantages of monetary exchange are not obtained when each piece of money must be examined and tested as to its quality. Wheat would make a poor kind of money for this reason; so would coal or cotton or indeed almost any other possible commodity. The Virginia legislature had to take cognizance of the distinction between "good tobacco" and "bad tobacco" and provide officials to examine

and grade the tobacco. Even so, the colonist had always to be on his guard against receiving payment in inferior or spoiled money. The precious metals, on the other hand, have perfect uniformity. When to this is added the quality of durability, which we have just noted, it results that an ounce of fine gold is exactly like any other ounce of fine gold, whether it came from a California mine in 1849 or was mined last year in South Africa or the Klondike.

The convenience of a monetary commodity is greatly enhanced by *divisibility*. Thus gold and silver may be divided and made into coins of any desired size without loss of value. Gold and silver have also the obvious advantage of being easily recognized, of *cognizability*. They are not easily confused with other substances, and counterfeiting is difficult.

Types of standard money systems. There are many forms which a standard money system may take, but for our purposes it will be sufficient to distinguish four types for discussion in detail. They are the following:

1. The gold standard, or gold monometallism
2. The bimetallic standard, or bimetallism
3. The gold bullion standard
4. The gold exchange standard.

The gold standard. Under a gold standard the standard money consists of gold coins; all other kinds of money which may circulate as a part of the system (representative money) derive their value from the fact that they are redeemable in terms of gold coins. In such a system there is (1) a monetary unit, prescribed by law, (2) free coinage of gold, (3) legal tender of gold coins, (4) free circulation of gold coins, including the right to hoard, and (5) a free market in gold, treated as a commodity. The first two of these features require some further discussion.

The monetary unit. In a standard money system the law prescribes the monetary unit in which the various kinds of money are stated. When the United States was on the gold

standard (from 1879 to 1933) the unit was the dollar, consisting of 25.8 grains of standard gold, standard gold being a mixture of nine-tenths pure gold and one-tenth an alloy of copper. The dollar thus contained 23.22 grains of pure gold and 2.58 grains of copper. Similarly the British monetary unit before September, 1931, was the sovereign, or pound, containing 123.273 grains of standard gold; British standard gold being eleven-twelfths fine, the sovereign contained 113.001 grains of pure gold and 10.273 grains of alloy. Note that the unit in a standard money system is always a certain quantity of metal whose weight and composition are exactly prescribed by law. The unit may or may not be represented by a coin. Very often the monetary unit is too small to make a convenient coin. This is of course the case in the United States.

Free coinage. The government obtains the metal for coinage in one of two ways; either it buys the metal, with coin or paper money, in the open market, or it allows *free coinage.* The former is simple enough, but the latter requires some explanation. Free coinage may be defined as *that system by which the government is legally required to coin for any person any amount of a particular standard metal which he may bring to the mint.* Free coinage is a necessary attribute of standard coin. For example there was in the United States until 1933 free coinage of gold.[1] Any person might bring to the mint any amount of gold of the standard prescribed by the coinage laws, and the mint would give back to him the same weight of gold coin. Governments sometimes perform this service for nothing; otherwise they make a charge sufficient to cover the cost of the coinage. In the United States the practice has varied. At first there was no charge; from 1853 to 1873 the charge was $\frac{1}{2}$ of one per cent; from 1873 to 1875 it was fixed at $\frac{1}{5}$ of one per cent; after 1875 there was no charge. When free coinage is granted without a charge for the cost of the opera-

[1] Under the bimetallic standard or irredeemable paper standard to 1879 and the gold standard thereafter till 1933.

tion, the term *gratuitous coinage* is employed. Note that there is free coinage, as defined above, whether the coinage is gratuitous or not.

The metal brought to the mint for coinage is often not of the standard composition; that is, not the exact mixture of pure gold and alloy prescribed by the law. In that case the mint or the government assay office will assay and refine it, charging the cost to the one who brought the gold or ore. The gold may then be returned to its owner or made into coin for him, and in the latter case there is a further charge for the alloy. In practice in the United States, the owner usually did not either take the gold or wait for its coinage, but took payment in lawful money; that is, sold the gold to the mint. The mint then held the gold either in "fine bars" for commercial use or in "mint bars" for coinage at its discretion.

The value relation between gold and money. Since the standard money under the gold standard consists of gold coins of an established weight and fineness, which the holder is free to hoard or melt down into bullion, the value of the standard money can never fall below that of an equivalent weight of gold in the form of bullion. On the other hand, if the value of gold coin were ever greater than the value of the corresponding weight of gold bullion, dealers, taking advantage of the right of free coinage, would buy gold in the commodity market and have it turned into coin until there was no longer profit in the operation because the values had become the same. Because gold coin and gold bullion can be freely interchanged without loss, the value of standard money tends always to be equal to the value of the substance of which it is made; *i.e.*, to its intrinsic value.

Under this system the various forms of representative money will have values equal to their nominal values because of the fact that they are all redeemable in terms of standard money, either by the government or by the agency which has issued them.

The bimetallic standard. *Bimetallism*, though no longer in effect in any nation, has played an important rôle in monetary history and involves economic principles of great interest. In a bimetallic system both gold and silver are concurrently used as materials for the standard money, and the monetary unit is defined as either a certain weight of standard gold or a certain weight of standard silver, the law prescribing also the exact composition of each standard metal. The ratio between the weight of pure silver in the silver unit and the weight of pure gold in the gold unit is called the *mint ratio*, coinage ratio, or legal ratio. There is free coinage of both gold and silver, and the standard coins of either metal are unlimited legal tender.

The gold bullion standard. A modification of the gold standard was employed by France and England for a time after the World War. For example England, having departed from the gold standard during the War, returned in 1925, not to her former standard, but to what is known as the *gold bullion standard*. The right of free coinage of gold was given up (except in the case of gold bullion brought to the mint by the Bank of England), and the Bank of England notes and other representative money were denied the right of redemption in gold coin. Instead there was substituted redemption by the Bank of England of its notes in gold bullion at the price of £3 17*s.* 10½*d.* per standard ounce, but only in large bars containing approximately 400 ounces of fine gold. The Bank was also required to purchase gold bullion whenever offered at the price of £3 17*s.* 9*d.* per standard ounce.

We have here illustrated the essentials of the gold bullion standard. There is not free coinage and little or no circulation of standard coins. Representative money is redeemable, not in gold coin, but in gold bullion. Redemption is usually limited to such large quantities of gold that the right is in practice available only for gold export or for very large domestic transactions. The fact remains however that there is redemption

in the standard metal, and this fact is sufficient to keep all forms of representative money at par with gold. Free trading in gold bullion is permitted, but the absence of standard coin and the limitation of redemption to such large quantities of money virtually prevents hoarding.

The gold exchange standard. During the period following the World War, another variation of the gold standard was extensively adopted in Europe, especially by small countries. In order to avoid the necessity of holding a large supply of gold as a basis for the currency, these nations made their representative money redeemable, not in gold, but in exchange upon some country having the gold standard. When an individual presented a sum of representative money for redemption he would receive foreign exchange, which could be used to obtain gold from abroad. The government itself treated bank deposits in such financial centres as New York, Paris, and London as parts of its own monetary reserve or standard money. This is the *gold exchange standard.*

It will be noted that under this standard the currency of a foreign country becomes the standard money. The government is obligated to redeem representative money either in gold or in foreign exchange at its own option. It is usual for the government to keep part of its reserves in the form of gold and part in the form of foreign exchange. There is of course no free coinage of gold, and little or no circulation of gold coin, but redeemability in gold exchange keeps the various forms of representative money at parity with gold. When a number of countries employ the gold exchange standard, the gold reserve of one country may serve at the same time as the gold reserve for several others. Should all of the countries on the gold exchange standard experience a drain of gold at the same time, their withdrawals of foreign exchange from the financial centres would impose a grave strain. Therefore the wide employment of the gold exchange standard introduces an element of precariousness into the international monetary system.

Irredeemable money systems: Origin and essential features. The distinguishing characteristic of an irredeemable paper money system is the absence of any standard money whatever. And by definition there can of course be no representative money. The monetary authority, whether it be the government or the central bank, no longer redeems money either in a commodity of intrinsic value or in the money of any other nation.

While it is perhaps conceivable that an irredeemable paper money system might be originally established without reference to a previously existing standard money system, this has as a matter of history never occurred. Actually the origin is to be found in the suspension of specie redemption of the representative elements (generally government notes or bank notes) in a standard money system. This action calls into operation a principle known as "Gresham's law," which is fundamental to all monetary theory and may be stated as follows: *When two or more kinds of money of unequal value are in concurrent circulation, each being available for payments, the inferior tends to drive the better out of circulation.* Anyone who has a payment to make naturally selects, from the several kinds of money in his possession, the least valuable ones, and makes his payments with them. The better forms thus tend to be held back out of circulation. The operation of Gresham's law forces out of circulation the standard money and any other forms of money which may still be redeemable, and the irredeemable paper money thereafter becomes the basis of the system.

The world's monetary history is replete with examples of irredeemable money. To mention early experience in America only, irredeemable money was a commonplace among the colonists; the chief monetary medium of the Revolution, the Continental bills of credit, were irredeemable money; and the greenbacks of the Civil War put the nation on an irredeemable money standard from 1861 to 1879.

Up to the time of the great depression (beginning in 1929), nations had rarely given up their standard money systems

except under the pressure of war or of some other crisis. Departure from the standard money system was a symptom of financial weakness. In recent years however many nations have abandoned the gold standard in favor of irredeemable money systems, not because they were actually forced to do so or because their metallic reserves were in danger of exhaustion, but for reasons of deliberate policy.

The irredeemable money system, like the standard money system, may comprise several different kinds of money. The basic money will generally be either government notes or bank notes. Frequently both of these forms of credit money will be in use, the bank notes being redeemable in government notes. Token money is generally present, performing the same function as in a standard money system, of course not redeemable in standard money, though it may be redeemable in other forms of money of larger denomination. The token money may be paper notes instead of coin. Bank deposits of course play the same rôle as in a standard money system, deposits being payable in the irredeemable paper money.

It may at first sight be hard to understand how an irredeemable money system can have any value at all, or why anyone should consent to receive such money. Yet that this does occur is demonstrated over and over in the annals of the world's monetary history. Irredeemable money, as we have seen, has almost always been representative money which has ceased to be redeemable. The people having become used to the credit money, force of habit continues its use even when redemption has ceased and the chance of future redemption has become virtually nil. Even when it declines in value, as evidenced by rising prices, it continues to circulate. Each individual feels that he can accept the money so long as he can pass it on to others, and so the one essential characteristic of money, general acceptability, is maintained. It is remarkable how far habit will thus lead a people. The Continental bills of credit of the American Revolution circulated till they were worth less than one cent on the dollar. The common people

of Germany after the World War were still using their paper mark when its value in terms of gold coin was impossible to calculate and when it was finally stabilized by the government giving one gold mark for one trillion marks of the paper money!

General abandonment of the gold standard. Upon the outbreak of the World War all European countries, neutrals as well as belligerents, abandoned the gold standard in order to safeguard their metallic reserves, and most other countries followed suit. Irredeemable money systems, consisting generally of the notes of central banks, prevailed throughout most of the world. The United States was one of the few nations still adhering to the gold standard. Even in the United States an embargo on gold exports was enforced from late in 1917 to the summer of 1919, and while redemption of other forms of money in gold was not refused, it was discouraged. During this period therefore the United States was not completely on the gold standard.

The period of recovery between 1923 and 1929 witnessed a concerted return to the gold standard. Certain small countries — for example, Turkey and Spain — failed to take part in this restoration movement, but some thirty other countries did so. England returned to gold in 1925 at the pre-war valuation of the pound. Many other nations, such as France, Italy, and Belgium, reduced the gold content of their monetary units upon returning to the gold standard.

The depression overthrew the gold standard again. In 1929 and the years following, one country after another — more than thirty countries by March, 1934 — abandoned gold, most of them for the second time. England stood out for many months, but was forced off gold in September, 1931. The United States abandoned the gold standard in 1933. The British dominions, the Scandinavian countries, and certain others subsequently regulated their currencies so as to maintain a determined relationship to the pound sterling and thus came to be known collectively as the sterling area. Only a

small group of nations — France, Belgium, Holland, Switzerland, Italy, and Poland — remained to constitute the gold bloc,[1] resolved to defend the gold standard against the forces threatening its complete overthrow.

Recent American monetary experience: Abandonment of the gold standard. The crisis of 1929 did not develop weaknesses in the American money system as such, but it did accentuate certain defects of long standing in the banking mechanism, which progressed to a state of almost complete collapse late in 1932. Runs upon the banks, withdrawal of money for hoarding by frightened depositors, and wholesale bank failures led the governors of several states to order all their banks closed in February and March of 1933. On March 5, only one day after his inauguration, President Roosevelt by executive proclamation ordered all the banks in the United States to be closed on March 6.

The proclamation of March 5, 1933, went further and forbade, except with special government license, the export of gold or silver or the paying of gold or silver by any bank. Simultaneously the United States Treasury ceased redemption of gold certificates and all gold payments. Four days later the emergency banking act of March 9 confirmed the President's proclamation and further authorized the President to regulate or prohibit hoarding of gold or silver and to require all persons to surrender to the United States Treasury all gold coin and bullion and gold certificates in exchange for other kinds of money. These provisions as to gold were immediately put into effect. It was generally expected that at the end of the banking crisis the nation would return promptly to the gold standard. Acceptance of this understanding is demonstrated by the fact that foreign exchange rates registered no immediate depreciation of the dollar.

Actually, while the acute banking crisis was virtually over in a week, no move was made to restore the gold standard,

[1] Germany and a number of central and eastern European nations remained nominally on gold but imposed restrictions on foreign exchange.

and on April 19, 1933, the President proclaimed an embargo of indefinite duration on gold exports, having previously extended the regulations against hoarding of gold by individuals. This was formal notice of indefinite abandonment of the gold standard. The proclamation was immediately reflected by depreciation of the dollar on the foreign exchange markets of the world.

The concluding act came on June 5, 1933, when the President approved a joint resolution of Congress which canceled the gold clause in all federal and private obligations, making them as well as all other contracts and debts payable in any form of legal tender. This action led to lawsuits ultimately involving some 100 billion dollars in gold-clause obligations. On February 18, 1935, the Supreme Court rendered a 5 to 4 verdict in favor of the government, holding that creditors must accept in depreciated currency, dollar for dollar, the sums named in their contracts, and that, while the government's action in repealing the gold clause in its own contracts was unconstitutional, plaintiff had no case against the government because no actual damage had been shown. The government subsequently enacted legislation designed to guard itself against further suits in this connection.

Devaluation of the dollar. The monetary policy of the Administration was designed to raise the price level, especially the prices of staple commodities dealt with in international trade, such as wheat and cotton. To accomplish these ends it was considered necessary first to cut the monetary system loose from gold and then to depreciate the value of the dollar in terms of gold. Depreciation of the dollar was sought by means of an extensive program of gold purchases, begun late in October, 1933. The offer by the government to purchase gold with dollars was of course equivalent to an offer to sell dollars for gold; and any buying price set above the former gold standard rate was equivalent to depreciating the gold price of the dollar. The government began to buy newly mined domestic gold at $31.36 an ounce, far above the old

gold standard price of $20.67 an ounce. Shortly thereafter this program was extended to include buying gold abroad, and the buying price was successively raised until it reached $34.45 an ounce on January 15, 1934.

The intended effect of this policy, operating through the foreign exchange markets, was to raise the dollar value of all American commodities whose prices were determined by sales abroad. The program was based on the following general principles: reduction of the gold content of the dollar and government purchases of gold result in a change in the foreign exchange rate whereby a pound sterling or other foreign monetary unit will exchange for more dollars than before; this decreases the prices of American export goods when expressed in foreign currencies and, other things being equal, increases the foreign purchases of American exports; increased sales abroad decrease the supply of the goods at home and so tend to raise domestic prices. The government's policy, as a matter of fact, did have some effect in raising American commodity prices, though the result fell short of what was intended.

Gold reserve act and proclamation of 1934. The gold-purchase program was followed on January 30, 1934, by the far-reaching gold reserve act and the accompanying proclamation of January 31, which devalued the dollar and nationalized gold. These measures provide that there shall be no gold coinage, no circulation within the United States of gold or gold certificates, no free redemption of currency in gold, no dealing in or holding of gold except under Treasury license; they empower the Treasury to deal in gold so as to maintain the parity of all kinds of currency and provide that gold bullion shall be made available under Treasury license for the settlement of international balances. The Treasury buys all gold presented, except such as may be known to have been hoarded in violation of the orders established in 1933, subject to the usual mint and handling charges; hoarded gold remains subject to confiscation with a 100 per cent pen-

alty charge, although in practice the Treasury has been purchasing such gold at the former mint price of $20.67.

Title to all gold coin and bullion held by the federal reserve banks was transferred to the Treasury in return for a payment in gold certificates at a rate determined by the old gold content of the dollar. These new gold certificates are held by the banks as final reserves and are redeemable at the Treasury only when gold bullion is needed to settle foreign balances. Coinage of gold has been stopped, and all gold held by the Treasury has been converted into bars. The notes of the federal reserve banks are now redeemable in "lawful money," which means other forms of paper money.

The new monetary unit. Legal tender. The gold reserve act also prescribed that the weight of gold defined as a dollar should not be more than 60 per cent of the former weight and granted the President continuing power to make successive revaluations within specified limits for a period of three years. This grant of power was subsequently extended and is still operative. It was directed that any increase in the value of gold (*i.e.*, the number of gold dollars equivalent to an ounce of gold) then held by the Treasury resulting from a reduction in the weight of the gold dollar should be carried in the Treasury as a miscellaneous receipt and that, in the event of any increase in the weight of the gold dollar, the reserve for notes and the security for gold certificates should be correspondingly augmented by a transfer of gold from the general fund of the Treasury.

The President immediately fixed the gold content of the dollar at $15\frac{5}{21}$ grains, $\frac{9}{10}$ fine, the equivalent of an even $35.00 an ounce as the dollar price of gold. The gold dollar since 1837 had contained a pure gold weight of 23.22 grains and a standard weight (*i.e.*, $\frac{9}{10}$ gold, $\frac{1}{10}$ alloy) of 25.8 grains. As now fixed, the dollar contains $13\frac{15}{21}$ (approximately 13.7) grains of pure gold and has a standard weight of $15\frac{5}{21}$ (about 15.24) grains. This is a reduction of the dollar to 59.06 per cent of its former weight.

The act of June 5, 1933, which abrogated the gold clause in public and private contracts, enacted that "all coins and currencies of the United States (including federal reserve notes and circulating notes of federal reserve banks and national banking associations) heretofore or hereafter coined or issued, shall be legal tender for all debts, public and private . . ." This makes all forms of money, apparently even the small-change coins, unlimited legal tender.

"Gold profit" and stabilization fund. The dollar value of the Treasury's gold was by this action increased by an amount slightly in excess of two and three-quarters billion dollars. The federal reserve banks were left with a claim to as many dollars' worth of gold as before devaluation. But since the new dollar contained only 59 per cent of the former weight of gold, the banks' share of the total gold in the country's monetary stock (measured either by weight or by dollar's worth) was reduced to 59 per cent of what it had been.

The gold reserve act gave the Secretary of the Treasury power for a period, originally of three years but later extended, to deal in gold and foreign exchange and any other necessary instruments for stabilizing the exchange value of the dollar. A fund of two billion dollars was provided for this purpose, to be taken from the increased value of the Treasury's gold which might result from devaluation, with the permission that any portion of the fund not being currently used in stabilization operations might be applied by the Secretary of the Treasury to the purchase of government obligations.

Silver policy. With the abandonment of the gold standard, the silver issue once more achieved prominence. Agitation for increased recognition of silver in the monetary system was insistent and led to a number of legislative enactments. The Thomas amendment to the emergency farm relief act in 1933 authorized the President to accept silver up to two hundred million dollars' worth on the basis of fifty cents per ounce in payment of the allied war debts due in 1933 and to issue silver certificates against the silver so received. A num-

ber of payments received in 1933 were thus made in depreciated silver. In the ill-starred world monetary and economic conference in London in July, 1933, an agreement was negotiated designed to raise and stabilize the price of silver. As its share of the plan the United States agreed to purchase twenty-four and a half million ounces of silver a year. In effect the agreement called upon this country to protect the price of silver against whatever damage might be done it by the limited disposal of some of India's surplus stock. In December, 1933, the President, acting under the provisions of the gold reserve act, announced that the United States mints were open, during the next four years, for the purchase of all the silver mined in the country, at the statutory price (unchanged since 1837) of approximately $1.29 per ounce, the producer however to receive only half of this amount, or approximately 64.5 cents per ounce, while the balance represented a 50 per cent *seigniorage charge* or profit to the government.[1] This was 21 cents above the current market price at that time.

Demand for further action on behalf of silver persisted and came to a head in June, 1934, when Congress passed the silver purchase act. The Secretary of the Treasury is directed to buy silver as he may deem it advisable at a price not above its monetary value ($1.29), except that in the case of silver in the United States prior to May, 1934, the price may not exceed $0.50. The Secretary may sell silver whenever its market price is above its monetary value. The purchases are to continue until silver constitutes 25 per cent of the monetary stock of the country. Silver certificates, now made full legal tender, must be issued at a face value not less than the cost of all the silver bought, and they may be issued in an amount not larger than the statutory valuation of the silver.

By the same act the President was given power to combat the hoarding of silver by nationalizing the commodity, which

[1] A charge made by a government for coining bullion is called *seigniorage* if it exceeds, and *brassage* if it does not exceed, the actual expense of coinage.

he did by executive order in August, 1934. The order required the delivery of silver bullion to the Treasury within 90 days at $0.50 an ounce. The Treasury promised to see that adequate supplies of silver would always be available at a fair price for legitimate commercial purposes.

In May, 1939, the price to be paid American silver producers was fixed by statute at 71.1 cents per ounce, over 22 cents above the world price.

So far as the American monetary system is concerned, the action taken with regard to silver has thus far been of little significance. The United States Treasury has acquired a considerable amount of silver bullion; it held 1,004 million ounces on December 31, 1939. There has been a moderate increase in the circulating medium through issue of silver certificates. On December 31, 1939, there were 1,777 million dollars of silver certificates outstanding, an increase of 1,295 millions since February 28, 1933. The stock of silver dollars (and bullion), above the requirements for backing the certificates, increased only 11 millions between these two dates. The total effect is an increase of 1,306 millions in the country's monetary stock on account of silver. During the same period the increase in the total monetary stock was 15,605 millions. Silver accounted for one-twelfth of the increase. Of the 1,298 million dollars in silver bullion held by the Treasury on December 31, 1939 the entire amount was pledged to redeem outstanding silver certificates.

The announced one-to-three ratio between silver and gold in the monetary stock is of no immediate practical significance. The present stock of silver dollars and bullion is far short of the amount required and is not likely to reach the announced ratio in the near future. And in any event the ratio means little in a monetary system which has largely lost connection with the Treasury's stock of gold. Only as presaging a future return to bimetallism is there real monetary significance in the present silver purchase policy. The essential result of the policy is to grant silver producers a government subsidy.

Summary. The present American monetary system. The gold reserve act and following executive order completed the process by which the United States definitely abandoned the gold standard and went onto a new money system. This is an irredeemable paper money system, though it is in certain respects different from such irredeemable money systems as prevailed during the Revolution and the greenback period. Subject to certain formalities and at the option of the Secretary of the Treasury, gold may be obtained in return for paper money for export and for domestic use commercially and in the arts. The restrictions which are now contained in the monetary system of the United States were designed primarily to make impossible the hoarding of gold, while not interfering unduly with its use for paying foreign balances or in the arts. This involves denial of free coinage and of the use of gold coin in circulation and withdraws the essentials of a standard money system; *i.e.*, free redeemability of all forms of paper money in gold at the pleasure of the holder and a free gold market, both domestic and for export.

The United States Treasury is obliged to purchase gold bullion whenever offered at the rate of $35.00 an ounce. This of course prevents the value of the paper money from ever rising above that ratio in terms of gold. On the other hand, the right of the citizen to receive gold bullion for export in exchange for paper money at $35.00 an ounce will keep the money from falling below that ratio in the international market so long as the Secretary of the Treasury, in the exercise of his discretion, permits such redemption without undue restriction. Exchange of paper money for gold for domestic use in the arts, also limited and at the discretion of the Secretary of the Treasury, is clearly not a sufficient redeemability to assure the gold value of the paper money internally, though it does maintain a certain limited relation between the paper money and gold. The present American monetary system is thus not turned entirely adrift from gold as in the extreme forms of irredeemable paper money systems.

CIRCULATION STATEMENT OF UNITED STATES MONEY, DECEMBER 31, 1939

Kind of money	*Held in U. S. Treasury*	*Held outside U. S. Treasury*			*Total stock*
		By Federal Reserve Banks and Agents	In Circulation		
			Amount	Per Capita	
1. Gold	$17,643,576,601	.....	.....	.....	$17,643,576,601
2. Standard silver dollars	498,973,344	$ 2,933,722	$ 45,171,523	$ 0.34	547,078,589
3. Silver bullion	1,298,212,193	.....	.....	.....	1,298,212,193
4. Token coins					
a. Subsidiary silver	5,340,761	13,234,104	380,885,556	2.89	399,460,421
b. Minor coin	1,613,498	2,989,778	164,339,244	1.25	168,942,520
Total coin and bullion	19,447,716,397	19,157,604	590,396,323	4.48	20,057,270,324
5. Gold certificates	12,393,580,447 *	2,815,444,500	69,482,369	0.53	15,278,507,316 *
6. Silver certificates	.....	223,078,939	1,554,314,016	11.80	1,777,392,955 *
Total certificates	12,393,580,447 *	3,038,523,439	1,623,796,385	12.33	17,055,900,271 *
7. U. S. notes	2,745,073	71,830,619	272,105,324	2.07	346,681,016
8. Treasury notes of 1890	.....	.....	1,164,572	0.01	1,164,572 *
9. National bank notes	2,135,623	979,750	175,106,914	1.33	178,222,287
10. Federal reserve bank notes	336,642	144,900	23,794,677	0.18	24,276,219
11. Federal reserve notes	13,312,152	349,432,330	4,911,778,868	37.31	5,274,523,350
Total notes	18,529,490	422,387,599	5,383,950,355	40.90	5,824,867,444
Total monetary stock	19,466,245,887	3,480,068,642	7,598,143,063	57.71	25,880,973,196

* Not included in the total of this column, in order to avoid double counting.

The table on page 362 shows the composition of the United States monetary system as it was on December 31, 1939. Comparison with corresponding figures for June 30, 1932, discloses certain important changes. The dollar value of the gold holdings of the Treasury is almost six times what it was in 1932 (partly due to reduction of the unit). This is all in the form of bullion. There is no longer any gold coin in circulation. The amount of gold certificates has increased greatly, but these are mostly held by the Treasury and the federal reserve banks and agents; only 69 million dollars are still in circulation. The principal part of the gold certificates in the Treasury is due and payable to the federal reserve banks. More than a billion dollars' worth of silver bullion has been acquired by the Treasury. No substantial change has taken place in the silver dollars, but the stock of silver certificates has more than trebled. Silver certificates are redeemable in silver dollars, and the Treasury is required to maintain the value of the silver dollars at a parity with other kinds of money, as before. No revolutionary changes have occurred in the other elements of the monetary system.

The currency system of course includes as always individual bank deposits subject to check. Deposits are payable in lawful money all forms of money being now legal tender.

EXERCISES

1. The silver dollar contains 371.25 grains of pure silver. There are 480 grains in an ounce.
   (*a*) The market price of silver today is about 43 cents per ounce. What is the intrinsic value of the silver dollar?
   (*b*) At what price of silver would the silver in a dollar actually be worth a dollar?
   (*c*) Suppose silver dollars were recoined into an equal number of dollars of lighter weight, and the surplus silver sold as bullion. How would this affect their intrinsic value? Their value as coins? How would it affect silver producers? Industrial consumers of silver? Explain, in each case.
2. The figures on page 364 indicate the weight of pure gold in the gold dollar at different periods in our monetary history:

1792–1834	24.75 grains of pure gold
1837–1933	23.22 grains of pure gold
1934–	13.71 grains of pure gold

(*a*) Determine the mint price per ounce for gold, during each of these periods.

(*b*) Daily quotations are to be found in the newspapers for silver, but no quotations for gold. Why?

(*c*) Prior to 1933, Congress had fixed the weight of the gold dollar at 23.22 grains of pure gold, and the weight of the silver dollar at 371.25 grains of pure silver. The price of gold was fixed, while the price of silver fluctuated. How was this possible?

3. After 1873 the market ratio rose as high as 80 to 1. At that ratio what was the market price of silver? Why did not silver coin drive out gold coin and leave the United States on a silver standard?

4. What would be the weight of the gold dollar if the mint price for gold were fixed at $20 an ounce? $24 an ounce? $30 an ounce? $35 an ounce?

CHAPTER XXI

Principles of Banking

Financing production. Let us for the moment assume a very simple type of productive organization. There are only manufacturers, wholesalers, retailers, and consumers. Goods pass in a direct line from the manufacturer to the wholesaler, to the retailer, and to the consumer. There are no banks or other financial houses to assist any one of them. When the manufacturer buys materials and pays wages he must pay in cash. When the wholesaler buys goods from the manufacturer he pays cash for them; the retailer and the consumer do likewise. The financial functions of these seem slight, but are they? The manufacturer is compelled to use his own funds while he is manufacturing the goods; the wholesaler ties up his resources in a stock of goods until he has disposed of them to the retailer; and the retailer is in the same position. Even the consumer may buy somewhat in excess of his immediate needs. As a matter of fact, each one is *financing* himself.

Some of the consumers, let us assume, find it impossible to pay for the goods immediately, and they induce the retailer to "trust" them for small amounts. But as a result of this situation the retailer may find himself unable to pay cash for the goods he buys from the wholesaler. He is confident that, if he were given three months in which to make payment, he could sell a certain amount of goods and secure payment from his customers, and so he suggests to the wholesaler that he be allowed to purchase the goods on the basis of deferred payments — that he be given three months' credit. If the wholesaler agrees to this, he assumes the burden of financing the retailer as well as himself. Obviously to do this he must have fairly large resources or be compelled to curtail the extent of his operations, provided he still pays the manufacturer

cash on delivery of goods. But he may make to the manufacturer the same plea which the retailer made to him, urging in extenuation his treatment of the retailer. If the manufacturer accedes to his request, he is financing himself, the wholesaler, the retailer, and the consumer. He is selling goods and receiving in exchange a written or verbal promise to pay for them some time in the future.

Credit. Here we see the essence of the credit system. Credit involves the purchase of something in the present — merchandise, money, or services — agreeing to pay for it in the future. In business the amount and time of the future payment is usually determined, and the payment is almost always made in money or some other form of currency. But this is not necessarily so. A farmer may borrow a few bushels of seed wheat promising to return a certain number of bushels of wheat when his crop is harvested.

The credit system is conditioned on the willingness and ability of individuals or of business organizations to advance present goods, whether money, merchandise, or services, in return for a promise to pay in the future. It therefore depends on the existence of a surplus, over and above present needs, which can be placed at the disposal of others, and the extent of the credit system is determined by the volume of savings.

In our present example credit has been entirely of the type we shall call *commercial credit;* that is, it has been extended by the manufacturer, the wholesaler, or the retail merchant. This held true in the main of the industry and commerce of former times. The financial houses devoted exclusively or primarily to financing business men are of fairly recent origin. For although it is true that banks have existed for many centuries — the Bank of Venice was founded in the twelfth century — and that we find money lenders and financial houses of importance in the medieval period, they were the exceptions to the general rule, and the producers were forced to rely on their own or their families' resources or on each other. It was not until the eighteenth and nineteenth centuries that

banks and other institutions so increased in number and resources as to become a vital part of the economic fabric and *bank credit* (a promise by a bank to pay) surpassed in importance what we have here called commercial credit (a promise by a manufacturer, wholesaler, or retailer to pay).

The commercial bank. Among the many types of banks that have developed during the course of history we need at the present day to distinguish at least three; namely, commercial banks, trust companies, and savings banks. Land banks might be added as a fourth class. Participating in the function of providing credit we find also note brokers, commercial paper houses, and private banks. It is however the commercial banks which play the predominant rôle in the modern economic organization, and it is to this type of bank that our attention will be directed.

The reader may be presumed now to know that a commercial bank's business consists primarily in making loans to and receiving deposits from its customers. Let us plunge at once into the middle of our subject by assuming a moderate-sized city bank and following through in detail from beginning to end some of its typical operations.

Lending: Example of borrowing. Let us assume that J. C. Miller, a dry goods merchant in the town, is engaged in buying his stock of goods for the early spring trade. As is usual, he does not have enough capital of his own for the purchase of his entire stock but depends regularly upon the financial assistance of his bank. He therefore goes to the bank and asks to borrow $50,000. The officers of the bank, after satisfying themselves as to his character and the state of his business, agree to accept his promissory note for $50,000

$50,000	New Haven, Conn., July 10, 1939
Sixty days after date I promise to pay to the order of the City Bank fifty thousand dollars. Value received.	
	(signed) *J. C. MILLER*

payable, let us say, sixty days from date. This note would be in some such form as the note shown on the preceding page.

Interest and discount. The compensation which the bank gets for supplying Mr. Miller with funds is called interest or discount, and the process is called loaning or discounting, depending upon the way the note is drawn. If the note is drawn *with interest;* that is, if Mr. Miller promises to repay at maturity the $50,000 plus interest at 6 per cent (the rate asked by the bank) for 60 days, he is said to be getting a *loan* of $50,000 and to be paying *interest* on $50,000. In this case, at the time of borrowing he receives $50,000 and at maturity will pay $50,500. The bank properly regards the promissory note as worth $50,000 at the time of borrowing and so enters it among its assets.

If however Mr. Miller writes his note for $50,000 *without interest,* the note is not worth $50,000 this day to the bank, and it must determine what value to allow him for it. The bank knows that 60 days from date it will bring $50,000; according to the principles of valuation with which the reader is now familiar,[1] it might be supposed that the bank would determine what sum put at interest for 60 days at 6 per cent would amount to $50,000; namely, $50,000 ÷ 1.01, which is $49,504.95. The $495.05 represents the *true discount,* or the amount by which the face value exceeds the present value of the note.

Banks are accustomed however, because of the extra profit and the easier computation, simply to deduct 60 days' interest at 6 per cent on the $50,000 face value, leaving $49,500. In banking parlance the $500 is called *bank discount* or simply *discount,* and the $49,500 is called the *proceeds.* Mr. Miller in this case really procures only $49,500 from the bank but pays interest on $50,000; and if he had needed exactly $50,000 it would have been necessary for him to draw his note for $50,505.05. According to custom, the bank enters this note among its assets at $50,000, which involves an immediate overvaluation of $500. This is corrected by adding $500 to

[1] See Chapter XIV

an item, "unearned interest," on the liability side of the bank's statement. There is transfer from this item to undivided profits as notes approach maturity.

The term *discount* is used both to refer to the interest deduction and to describe the whole process of evaluating and disposing of the note. In the financing of commerce the bank discount method is most common at the banks, but there are many types of financing in which the loan method prevails.[1]

Proceeds. The next question is as to the form in which Mr. Miller shall receive the funds borrowed. There are ordinarily three forms, (1) money (other than the bank's own notes), (2) the notes of the bank, (3) deposits. The first is simple enough and needs no explanation. In the second case the bank gives Mr. Miller, in exchange for his promissory note, its own promissory notes. It is a swapping of notes. But the notes of the bank are money; they are generally acceptable throughout the community, and Mr. Miller has received what he wanted, the means for making immediate payment for his goods. Indeed the ordinary business man does not distinguish between different kinds of money; to him there is practically no distinction between the first two cases. To the bank, as will appear later, there is a very real distinction. Both of these forms of payment however are the exceptional cases, at least in the United States. In the great majority of instances the borrower receives the proceeds of his note, not in money ("cash"), but in the form of a deposit credit.

When money is taken to a bank and left — "deposited" — the bank credits the customer's account on its books, and the latter obtains the right to withdraw the specified amount of money, or any part of it, at will. Now when a note is discounted the amount of the proceeds is (in the ordinary case) credited to the account of the borrower as a deposit, precisely as though he had brought money. If it is any aid to visualizing the process, it may be imagined that the customer received

[1] As is the case with most economic categories, the terms, *loan* and *discount*, *interest* and *discount*, are loosely and interchangeably used by the public.

the proceeds in money and immediately stepped over to the receiving teller's window and deposited the money. Actually however no money has been involved; this is indeed not a money transaction. It is once more a swapping of credit. The bank receives the customer's promissory note and gives him in exchange the right to receive money from it on demand, which is precisely what he would have if he had taken instead the bank's notes.

Principal types of trade paper. We must now retrace our steps in order to consider the several forms of paper which may be used in ordinary commercial borrowing at a bank. There is first the simple promissory note of the borrower. Such a note may be *unsecured*, or it may be *secured* by "collateral security." In the latter case the borrower delivers to the bank along with the note certain property of his, worth at least as much as the amount of the note, with an agreement that in case the note is not paid when due the bank may sell the security and so reimburse itself for the note. Stocks and bonds of corporations and government bonds are the property most used as collateral security, though warehouse receipts for certain commodities, bills of lading, etc., may be used.

In the example which we have used, Mr. Miller discounted at the bank a simple promissory note, either secured or unsecured, and obtained the $49,500 which he used to pay the wholesaler. The purchase of his stock in trade might however have been somewhat differently financed. It is common practice for the wholesalers to sell to their customers on credit; that is, to accept their promissory notes in place of cash. If Mr. Miller has such an arrangement with a wholesaler, R. B. Norton, in the same city, he will give Mr. Norton in exchange for the goods his promissory note for $50,000 payable in 60 days. The wholesaler, Norton, being himself actively engaged in business with all his capital, will not be willing to hold Miller's note till maturity; he needs the cash before that and will therefore avail himself of the services of the bank. Norton first *endorses* Miller's note; that is, signs

his own name, thereby making himself liable for its payment in case Miller should not pay. He then takes the note to the bank, which discounts it precisely as in the case when Miller borrowed on his own note. The proceeds are given to Norton, either in money or, more likely, as a credit on his deposit account. The note becomes the property of the bank, and when it is due Miller will pay, not Norton, to whom he gave the note originally, but the bank. A note thus endorsed is called "two-name paper" as compared with the simple promissory note of the maker, called "one-name paper."

The paper offered to the bank may take still a third form; namely, the *draft* or *acceptance*. The retailer, instead of giving the wholesaler his promissory note in return for goods received, may give him the authority to *draw* upon him. This, in the case we are using for illustration, would mean that when Norton sent the goods to Miller he would write out an order upon Miller directing him to pay the $50,000 in sixty days to Norton's bank. This draft would then be taken to Miller, who would *accept* it; that is, write the word "accepted" and his signature across its face, which is his acknowledgment of his obligation to pay at the maturity of the draft. This draft would be in form similar to the following:

New Haven, Conn., July 10, 1939

To J. C. Miller

On September 11, 1939, pay to the order of the City Bank fifty thousand dollars ($50,000).

The transaction which gives rise to this instrument is the purchase of goods by the acceptor from the drawer.

Accepted at New Haven, Conn.
July 10, 1939
Payable at the City Bank,
New Haven, Conn.

(signed) *J. C. MILLER* (signed) *R. B. NORTON*

A draft when thus accepted is called an *acceptance*. It is two-name paper since the drawer, Norton, is liable as well as

Miller, upon whom it is drawn. Mr. Norton then takes the acceptance to the bank and discounts it precisely as has been explained in the case of Miller's promissory note discounted by Norton. At the maturity of the draft, Miller will pay the bank. These various forms of business men's notes and drafts are all called *trade paper*.

A bank's balance sheet. These matters may be made clearer and the foundation laid for our further investigation, if we now give some attention to the balance sheet of the bank. A commercial bank is ordinarily a corporation, and to it may be applied what has already been learned about the organization and accounts of corporations. Avoiding details not here essential, we may say that a bank is organized about as any other business corporation.[1] The first balance sheet might read somewhat as follows:

STATEMENT OF THE CITY BANK, JULY 1, 1939

Assets		*Liabilities*	
Cash	$400,000	Capital stock	$400,000

The bank proceeds at once to buy land and a banking building, furniture and supplies, etc., for $60,000, and thereafter its balance sheet reads as follows:

STATEMENT OF THE CITY BANK, JULY 8, 1939

Assets		*Liabilities*	
Real estate	$ 45,000	Capital stock	$400,000
Furniture, etc.	15,000		
Cash	340,000		
	$400,000		$400,000

Discounting a note. Now comes Mr. Miller, the merchant, with his request for a loan. As has been already assumed, he offers his promissory note for $50,000 payable in sixty days, the bank's rate of discount is six per cent, the discount is $500, and the proceeds, $49,500. Mr. Miller elects to receive "cash" other than the bank's own notes. The effect of this

[1] See Chapter III.

transaction upon the bank's condition is shown in the following statement:

STATEMENT OF THE CITY BANK, JULY 10, 1939

Assets		*Liabilities*	
Loans and discounts	$ 50,000	Capital stock	$400,000
Real estate	45,000	Unearned interest	500
Furniture, etc.	15,000		
Cash	290,500		
	$400,500		$400,500

It will be noted that the $50,000 promissory note acquired by the bank is listed among its assets opposite the heading "loans and discounts," that the cash has been reduced by the payment of $49,500 to Mr. Miller, and that the difference between these amounts, the discount, appears as unearned interest. Obviously the effect on the bank's statement will be the same whether Mr. Miller borrows by means of an unsecured note or a note secured by collateral. Even if Miller finances his business by giving his promissory note to the wholesaler or permitting the wholesaler to draw upon him, the final result is the same, the only difference being that in these cases (that is, the endorsed note or the acceptance) Miller's obligation is discounted at the bank by Norton, the wholesaler, instead of by Miller himself. In any case the bank comes into possession of Miller's obligation, which he must pay to the bank at maturity.

The statement will however differ according to the form in which the proceeds are given to the customer. The statement above illustrates the case where the proceeds are taken in the form of money other than the bank's own notes. If the bank had given its notes the statement would have read thus:

STATEMENT OF THE CITY BANK, JULY 10, 1939

Assets		*Liabilities*	
Loans and discounts	$ 50,000	Capital stock	$400,000
Real estate	45,000	Unearned interest	500
Furniture, etc.	15,000	Notes	49,500
Cash	340,000		
	$450,000		$450,000

If finally the proceeds had been credited to the deposit account of the borrower, the statement would have been as follows:

STATEMENT OF THE CITY BANK, JULY 10, 1939

Assets		*Liabilities*	
Loans and discounts	$ 50,000	Capital stock	$400,000
Real estate	45,000	Unearned interest	500
Furniture, etc.	15,000	Deposits	49,500
Cash	340,000		
	$450,000		$450,000

A typical bank statement. To carry our illustration somewhat further, let us assume that during the next few months the bank has done a flourishing business, making loans, discounting paper, and correspondingly increasing its deposits and notes and profits, that a part of the latter has been transferred to surplus, and that the bank has made some investments in securities. The statement, corresponding more closely to a normal banking business, may be assumed to be as follows:

STATEMENT OF THE CITY BANK, DECEMBER 31, 1939

Assets		*Liabilities*	
Loans and discounts	$1,250,000	Capital stock	$ 400,000
Securities	450,000	Surplus	75,000
Real estate	45,000	Undivided profits	8,117
Furniture, etc.	15,000	Unearned interest	4,210
Reserve	182,460	Notes	50,000
		Deposits	1,405,133
	$1,942,460		$1,942,460

The reader will note the appearance in this statement of a new term, the *reserve*. This is an inclusive term indicating the sum of such assets as are available for payment of the bank's liabilities, especially its notes and deposits. The term is usually synonymous with cash or items equivalent to cash. American banks do not generally use this term, specifying rather the separate items which compose their legal reserves.

Bank deposits: Definition. Attention has thus far been directed principally to the function of discount. We must now inquire further into the subject of deposit. *A bank deposit is a*

right to receive money from a bank, evidenced by an entry on the bank's books and by the customer's passbook, duplicate deposit slip, or other device. Deposits are classified as demand or time deposits, the bank being obligated to pay the former on demand of the depositor but obligated to pay the latter only after a certain number of days' notice.[1] Deposits may result from the discount of trade paper as illustrated above, or from deposit of money, or from deposit of checks drawn by another depositor, either of this or another bank, or in certain other ways that need not concern us here.

To illustrate the deposit of money, let us now suppose that a customer brings in $10,000 of money for deposit. The bank's statement immediately thereafter will read thus:

STATEMENT OF THE CITY BANK, JANUARY 9, 1940

Assets		*Liabilities*	
Loans and discounts	$1,250,000	Capital stock	$ 400,000
Securities	450,000	Surplus	75,000
Real estate	45,000	Undivided profits	8,117
Furniture, etc.	15,000	Unearned interest	4,210
Reserve	192,460	Notes	50,000
		Deposits	1,415,133
	$1,952,460		$1,952,460

It will be noted that nothing on the bank's statement distinguishes between deposits obtained by paying money to the bank and deposits obtained by the process of discount. There is no difference. What the depositor has is exactly the same in either case, not money, but the right to receive money. Deposits are not, as is sometimes popularly supposed, "money in the bank." Banks never have on hand a stock of money equal to their deposits. That could hardly be expected since, as we have seen, deposits may originate without any money being given to the bank. Even when money is deposited, the money ceases from that moment to be the property of the depositor; it belongs to the bank and may be spent by the bank as it sees fit. The depositor has simply the right to receive a certain amount of money from the bank.

[1] Thirty days' notice is commonly required.

Discount and deposit. In fact both discount and deposit, contrary to the meanings popularly attached to the terms "loan" and "deposit," are actually exchanges, transfers of certain rights for other rights or for money. When the bank discounted Mr. Miller's promissory note it acquired a property right against Mr. Miller and in exchange gave him a property right against the bank; namely, a deposit. Quite properly the bank lists the note as an asset and the deposit as a liability. On the other hand, Mr. Miller enters the note on his accounts as a liability and the deposit at the bank as an asset. When Mr. Norton, having accepted Mr. Miller's promissory note in payment for goods, discounted the note, he sold it to the bank in exchange for a deposit, or perhaps for some form of money. Money deposited at the bank is, so to speak, "spent" — paid for a deposit.

The double function of discount and deposit is ordinarily a swapping of credit — the person's promissory note or draft or acceptance for the bank's credit in the form of notes or deposits. If it be asked wherein lies the advantage of merely exchanging one form of credit for another, the answer is that what the bank's customer wants is a medium for making payments. His own promissory note will not be generally accepted; the bank's note will. He has exchanged a form of credit generally unknown and unacceptable for the well-known and acceptable credit of the bank. The same is true of deposits, since practically all the customer's payments may readily be made by check. Herein is the fundamental characteristic of bank credit and of the service rendered by the commercial bank.

The check: Definition. The right to receive money which a bank deposit gives the depositor is exercised by means of the check. If Mr. Miller desires to withdraw fifty dollars from the bank in order to make a certain payment or for pocket money, he makes out an order upon the bank; *i.e.*, a check, in the following form:

51–100

No. *731*

New Haven, Conn., *Jan. 9*, 1940

THE CITY BANK

Pay to the order of.....*Myself*.....*$50.00*

.....*Fifty*.....Dollars.

(signed) *J. C. MILLER*

He presents this order to the bank, which thereupon gives him fifty dollars in money and deducts fifty dollars from the amount due him on its books. It will also deduct fifty dollars from its reserve item in the assets column.

If Mr. Miller wished to pay a doctor's bill of fifty dollars, he might obtain from the bank fifty dollars by means of a check like the above and then take the money to the physician. But a simpler method would be to order the bank to make the payment direct to the physician. For this purpose a check would be used in this form:

51–100

No. *732*

New Haven, Conn., *Jan. 9*, 1940

THE CITY BANK

Pay to the order of.....*Robert Brown*.....*$50.00*

.....*Fifty*.....Dollars.

(signed) *J. C. MILLER*

From these examples the nature of the check is apparent. *A check is a written order by a depositor directing his bank to pay money to some person.* The person to whom payment is to be made may be the depositor himself or the bank or any specified third party or the bearer.

Payment by check: Within the bank. Let us follow the history of this check. The check is handed or sent by mail to

Dr. Brown, who takes it to the bank and demands the money. Certain essential facts must be established before the paying teller of the bank will hand over the money. First he must be satisfied as to the genuineness of the check, which is usually sufficiently attested by the signature of the maker, with which the teller is familiar. Next he must be certain that Mr. Miller actually has as much as fifty dollars to his credit on the bank's books. The teller must also be certain that it is Dr. Brown himself who is asking for payment; if he is not personally known he will have to be identified. Before payment is made the check must be endorsed; that is, the payee, Robert Brown in this case, must sign his name on the back of the check. Endorsement is the payee's acknowledgment that he has received the money; it makes the check evidence of payment for the bank and for the maker. Having made the payment, the bank debits the amount to Mr. Miller's account, and eventually, when the account is balanced, the cancelled check will be returned to him to be kept as evidence of the payment. The effect of this operation will be to alter two items on the bank's balance sheet, making it read as compared with the last statement (on page 375) as follows:

STATEMENT OF THE CITY BANK, JANUARY 10, 1940

Assets		*Liabilities*	
Loans and discounts	$1,250,000	Capital stock	$ 400,000
Securities	450,000	Surplus	75,000
Real estate	45,000	Undivided profits	8,117
Furniture, etc.	15,000	Unearned interest	4,210
Reserve	192,410	Deposits	1,465,083
	$1,952,410		$1,952,410

Instead of demanding the money, the payee may "deposit" the check. He endorses the check and takes it or sends it by mail or messenger to the bank, with the request that the amount stated be credited to his deposit account. The bank must be certain as to all the facts except identification, just as when money is demanded, and must be satisfied as to the genuineness of the endorsement. It will then credit the account

of the payee (*e.g.*, Robert Brown) and debit the account of the maker (*e.g.*, J. C. Miller). The result of this transaction is exactly as though the payee had collected the money from the bank and then immediately deposited it. The balance sheet of the bank is not changed at all.

The check thus furnishes a means of making payments so convenient, safe, and generally advantageous that, in the Anglo-Saxon countries at least, it is availed of very generally, not only by business men but by others. Deposit accounts are kept at the bank. Money when received is regularly deposited at the bank, and the same is true of the far greater volume of receipts which come in the form of checks. Conversely fully four-fifths of all payments are made by means of checks rather than with money. The bank thus becomes an agent for making payments between the people by the transfer of deposits from one to another.

Collection between banks: In general. We have proceeded thus far upon the assumption of one bank with which all persons concerned have their dealings. As a matter of fact in every place of any size there will generally be found two or more banks. Moreover the useful device of payment by check has not been permitted to remain a merely local convenience but has absorbed the great mass of payments between persons in different cities, states, and even nations. This introduces certain complications, which must now be investigated.

When one receives a check, the obvious thing is to take it to one's own bank and either deposit it or ask for the money. This is done regardless of whether the check happens to be drawn on this same bank or upon another. In the first case, the matter is simply arranged, as has been explained. Even in the second case, the bank, though under no obligation to cash or credit the check, will ordinarily do so as an accommodation to its customer, sometimes making a small charge for performing this service. Having given its customer the cash or credited his account, the bank thereupon becomes possessed of the check with its claim upon the other bank. If

the other bank is in the same place, the check will ordinarily be presented and paid within twenty-four hours. A longer time may be required when the debtor bank is in another city. The other bank of course then debits the account of its depositor who drew the check.

Thus the check system for payment of money or transfer of deposits operates when there are two or more banks in the town. Daily each bank sends a messenger to each of the others and collects for all checks of each other bank which it has paid. But in a large city with many banks this business of collecting by messenger becomes troublesome and expensive. In a city with fifteen banks there might have to be 15×14, or 210, separate calls each day. Sooner or later the point is reached where it is worth while to avoid this great labor and expense by organizing a clearing house.

Clearing house. A clearing house is an association of the banks of a city through which all the claims and obligations between the banks for paying each other's checks and drafts are settled. While the fundamental nature of the clearing house and its business is everywhere the same, there are numerous variations of detail. The following account is based primarily upon the methods employed in the New York Clearing House. The business of clearing takes place daily at certain hours. Before each clearing there occurs in each bank the same preparatory operations. The clerks gather together all checks of the other banks which have been received since the last clearing and arrange them in bundles, one for each bank. A list is made showing the amount thus due from each of the other banks, with the total due from all the banks. Sometimes there is also made out a ticket for each of the other banks showing the amount due from that particular bank. Then two clerks, gathering together all the bundles, the tickets, and the list, proceed to the headquarters of the clearing house. Here there is a room containing a raised desk at one end for the governor and his assistants and on the floor a row or rows of desks, one for each member bank.

Of the two clerks one takes his place behind his own bank's desk while the other, carrying the bundles of checks and the tickets, stands in front of the desk. The list showing the amount due this bank from all others has been sent up to the governor's desk. In like manner, each of the other banks is represented by two clerks, one behind and one in front of its own desk.

At the exact moment for beginning the operation of clearing a signal is sounded on the governor's desk. The clerks in front then walk along the row of desks, depositing at each the bundle of checks and the ticket against that particular bank. In a minute or two the checks and tickets have thus all been distributed, and it takes only a few minutes more for each desk clerk to add up the tickets or the sums on the outside of the envelopes and so ascertain the amount his bank owes to all of the other banks. These results are reported to the governor, whose assistants are able quickly to discover whether any error has been made. As soon as the figures, corrected if necessary, are compiled, the governor announces the result. Each bank now knows exactly how much it owes all the other banks and how much all the other banks owe it. These two sums are regarded respectively as a debt to the clearing house and a credit due from the clearing house. The balance represents an amount which the bank must pay to the clearing house or receive from the clearing house as the case may be. Thus by one payment either to or from the clearing house, each bank settles its accounts with all the other banks. Of course the amount payable by all the debtor banks to the clearing house is exactly equal to the amount payable by the clearing house to all the creditor banks, and the clearing house itself comes out even each day.

Clearing for outside banks. All the banks of a city do not necessarily belong to the clearing house. An outside bank will either make its collections by messenger and make its payments to messengers sent by the other banks in the primitive way, or it may arrange to clear through another bank

which belongs to the clearing house. In this case the agent bank presents as its own the claims of the outside bank and accepts as obligations against itself claims of the other clearing house banks against the outside bank.

The volume of clearings. The volume of the business thus simply and smoothly effected by a large city clearing house is enormous. In the New York City Clearing House there were exchanged, during the month of December, 1938, checks and similar obligations to the amount of $17,950,623,781. For the year 1938, the New York City clearings were 163 billion dollars. The New York City Clearing House stands far ahead of any other in volume of clearings. For the whole of the United States the clearings during the month of December, 1938, were about 30½ billions, and for the whole year 1938, 307 billions. Since our available sources of such statistics do not include reports from all the cities having clearing houses, the figures here given for the whole country are materially less than the actual totals of all clearings. It should also be noted that the year 1938 was not one of active business; at the year's close the country was definitely in a state of industrial depression. In 1929, for example, clearings were about two and a half times those of 1938. One can imagine what a gigantic task would be imposed upon the banks were they still compelled to rely upon the primitive method of exchanging their obligations.

The reader should be reminded that even these figures do not give a true picture of the total volume of banking business. In fact the recent movement toward bank consolidation, while accompanied by increased business, has obviously tended to a diminished use of the services of the clearing houses. A more reliable indication of banking business is presented by the figures of debit check transactions, which for the year 1938 were 169 billion dollars for New York City and 406 billions for the principal cities of the country.

Out-of-town checks. Checks are regularly used for payments in other towns. When goods are purchased by mail

order, payment is ordinarily made by a check drawn by the purchaser upon his own bank and sent by mail to the seller, it may be a thousand miles away. The maker of the check gives himself no further concern over the matter. The receiver of the check deposits it in his own bank, receives credit for it, and concerns himself no further. His bank has the task of collecting from the other distant bank. Thousands of payments are thus made every day, and every day the mails carry great numbers of checks back and forth all over the country. There is thus imposed upon the banks a service of great importance and enormous volume. An elaborate system has been developed whereby this business is handled with swiftness and certainty and the minimum of expense and, by offsetting checks against each other, with remarkably little actual transfer of money.

Paying a loan. We must now return to the further study of the business of discount and deposit. We have used for illustration the discount of a note for $50,000 for sixty days at the City Bank by J. C. Miller. On September 11, 1939, the loan matures and must be paid at the bank. This is equally the case whether Mr. Miller borrowed directly from the bank on his own promissory note or gave his promissory note to the wholesaler, Norton, or accepted a draft drawn upon him by Norton, so long as the paper has been discounted at the bank.

Mr. Miller now presents himself and offers to pay his note. He may pay in any of five ways: (1) in money, other than the notes of this bank, (2) in notes of this bank, (3) by check upon a deposit account in this bank, (4) by a check upon an account in some other bank, or (5) by a bank draft; *i.e.*, a check drawn by a bank either on itself or on its account in some other bank. The effect upon the bank's condition is readily determined. In any case the loans and discounts will be reduced by $50,000, representing the note paid and returned to Mr. Miller. If the note is paid in money the bank's cash, or reserve, will rise by $50,000, and the same result will follow, within a few days

at most, if payment is made by a check or a bank draft on another bank, since this will be promptly presented to the other bank, through the clearing house or by mail or messenger, and collected. If the note is paid by giving notes of this bank, the note item on the liability side of the account will be reduced by $50,000. Finally if payment is by a check either on Mr. Miller's account or on the account of another depositor in this bank, the bank's deposits will be correspondingly reduced. Two changes of $50,000 each counteract each other, and the statement remains in balance.

Notes and deposits: Fundamental similarity. The bank deposit has already been defined as a right to receive money from a bank, evidenced by an entry on the bank's books and by other means. *A bank note is a bank's promise to pay money on demand.* The term "money" in both these definitions is of course exclusive of the bank's own notes. It will be noted that these definitions are almost identical. Deposits and notes are both forms of bank credit; both are liabilities of the bank; both are (except in the special case of time deposits) payable in money on demand. When a bank makes a loan it is immaterial to it whether the credit given the borrower is in the form of notes or deposits. To the borrower it is merely a question whether he shall receive the evidence of his right to receive money from the bank as an entry on the bank's ledger and in his own pass book or inscribed on little pieces of paper which he will take away with him. These differences are superficial, not fundamental.

Differences between notes and deposits. There are however real differences between notes and deposits. The right to receive money from a bank, if in the form of a deposit, may be transferred only by check. Now a check is of itself no obligation of the bank. The person who receives it must be sure (1) that the signature upon the check is genuine and (2) that the maker actually has at least that amount on deposit and that it will still be on deposit when he presents the check to the bank for payment. Otherwise the check is not good. Checks

are thus available for use only among people who are acquainted with one another or at least have confidence in one another. Their availability is also limited to those who are in the habit of doing business with banks. Checks are thus not generally convenient for payments between strangers, for paying traveling expenses, or for making ordinary wage payments. Where they are available, checks are usually a safer and more convenient means of payment than money. But after all they are not money, since they are not generally acceptable.

The bank note, on the other hand, is money. It is the definite obligation of a bank and, barring the risk of counterfeit or of the failure of the bank, may be accepted without any question as to the honesty of the one who tenders it. The bank note is generally acceptable; it may be used for all kinds of payments. Bank notes again are expressed in round sums, convenient for payments of all amounts. Although representing rights to receive money from the bank, bank notes may pass about for weeks or months or years before being presented to the bank for payment. They may travel far away from the bank of issue. The check, on the contrary, usually appears for the purpose of a particular payment; it is expressed in an odd amount of dollars and cents, and its life is ordinarily limited to that one transaction, after which it returns to the bank and is cancelled. That is, deposits require generally a new check for each separate transfer.

These distinctions between notes and deposits, though not so fundamental as their similarities, are yet real and important and justify the issue of bank credit in the two forms.

Legal restrictions. An important corollary of the differences between notes and deposits is that the opportunity for abuse and fraud on the part of an unsound or dishonest bank is far greater in the case of notes than of deposits. This is in part the explanation of the prevailing attitude of government toward the banking functions of deposit and note issue. In spite of their fundamental similarity and their original status as common law rights, the legislation of most countries has

drawn a sharp distinction between them. While the banks have generally been permitted to handle their deposits with little interference by the government, the function of note issue has almost everywhere been subject to strict regulation. In most countries the right of note issue is restricted to certain kinds of banks or even to one particular bank. Thus in the United States (since August 1, 1935) only the federal reserve banks issue notes. In most of the leading nations of Europe, such as Great Britain, France, Germany, etc., the business of note issue is virtually the monopoly of a single great central bank, more or less closely associated with the government. Legislation further puts a limit to the amount of notes which the banks may issue, prescribes certain forms of security and a certain amount of cash reserve, determines the form and denomination of notes, etc. As a result the issue of bank notes is in most countries a narrowly restricted and somewhat artificial function as compared with the business of deposit.

The normal business of banking. Acquaintance has now been made with the more important operations which make up the normal daily business of the ordinary commercial bank. The bank's "portfolio" of loans and discounts contains numerous notes and acceptances of individuals and corporations, to which new notes are being added daily and from which notes are being as constantly withdrawn when they mature and are paid. The bank's deposits are being constantly added to as new loans are made or as customers bring in for deposit either money or checks upon other banks, while, on the other hand, withdrawals by check are continually tending to reduce deposits, particularly in connection with the payment of loans. The bank's notes outstanding tend likewise to rise and fall in harmony with the loans and discounts. From time to time securities held by the bank are sold or new ones purchased, with corresponding changes in the "securities" item. In connection with many of these operations changes are continually occurring in the bank's reserve, as cash is either received or paid out. So long as the business is profitable, the "un-

divided profits" item gradually increases, reduced at intervals by the declaration of dividends or by transfer to surplus.

The reader cannot have failed to observe the close relation between loans and discounts and notes and deposits. Ordinarily the customers' borrowing increases both the loans and discounts item and notes and deposits. Even if cash is taken instead of a deposit credit, the cash tends to come back and increase deposits as it is deposited by those persons to whom the original borrower made payments. When loans are repaid, there is normally a decrease in loans and discounts and notes and deposits. These relationships are fairly obvious in the case of a bank which is alone in its community. For several banks in a community and for the whole system, including the banks of other places, the same results follow for all the banks as a whole, though they may be accompanied by considerable shifting of cash and deposits as between different banks and different regions.

The reserve: Its function. The deposits and notes of a bank represent claims against the bank held by outsiders and payable in cash on demand. For this purpose the bank must have always on hand a sufficient stock of cash, appropriately called the reserve. But no commercial bank ever holds a reserve equal to the total of its demand liabilities. In fact the reserve will seldom be as much as twenty-five per cent and may be very much less. In the last illustrative statement (page 378) the bank's liabilities payable on demand (notes and deposits) appear as $1,465,083, for which there is a reserve of $192,410—a *reserve ratio* of a little more than thirteen per cent. Evidently a simultaneous demand from all depositors and note holders would wreck the bank. But of course no such simultaneous desire for payment is to be expected, and the bankers learn by experience about what demands may be anticipated from day to day. They make provision for a corresponding reserve with a proper margin for emergencies, and it is thus that the reserve ratio is determined except when legal requirements interfere.

How the reserve ratio is controlled. The banker is impelled by two opposing motives — profits and safety. The bank derives its profits principally from the making of loans and discounts, and the larger its portfolio of loans and discounts the larger in general will its profits be. But as loans and discounts are made, cash is immediately withdrawn, or deposits are created, or the bank's note issues are increased. Thus the reserve ratio falls, and the bank's condition becomes proportionately less safe. To the banker's desire for profits is thus opposed the necessity of keeping a safe ratio between the reserve and the demand liabilities.

There are five principal means of controlling the reserve ratio: (1) by the purchase and sale of securities, (2) by rediscounting commercial paper, (3) by borrowing from another bank, (4) by calling in loans which are payable on demand, and (5) by making it more or less easy or advantageous to borrow. The latter may be accomplished by changing the rate of discount, by simple refusal to loan, by moral suasion inducing applicants not to apply for loans, by raising or lowering the collateral security requirements, or by requiring more or fewer names on the paper.

Buying and selling securities. As we have seen, a bank usually invests a part of its resources in corporate securities, the main purpose being to derive some income upon funds which if kept as idle cash in the reserve would be earning nothing, while at the same time having something which may be converted into cash whenever needed. The bank therefore chooses for investment securities which may be sold at short notice. In the statement on page 378 the bank has a reserve ratio of a little more than thirteen per cent. Let us assume that this is considered too small and that securities carried on the books at $100,000 are sold for that amount, reducing the stocks and bonds to $350,000 and increasing the reserve to $292,410. The reserve ratio is now $\frac{292,410}{1,465,083}$, or almost exactly twenty per cent. The bank has sacrificed the

future interest or dividends on $100,000 of securities, but it has made its position materially safer. Conversely a purchase of securities would reduce the reserve and consequently the reserve ratio.

Rediscounting and borrowing. Instead of selling securities, the bank may strengthen its reserve ratio by selling some of the paper in its portfolio. This is called *rediscounting*, which means to discount again. Some investor is found, either another bank or a broker, willing to take certain promissory notes or drafts at a price determined by discounting them at a certain rate for the time that must still elapse before their maturity. Suppose our bank selects from its portfolio a note for $60,000, originally acquired two months ago and payable three months from its date, having therefore still one month to run. If another bank is ready to rediscount it at eight per cent, the City Bank will receive for it $59,600; that is, the face value less one month's interest at eight per cent. The result of this operation, starting from the statement on page 378, will be reflected in a new statement, as follows:

STATEMENT OF THE CITY BANK, JANUARY 11, 1940

Assets		*Liabilities*	
Loans and discounts	$1,190,000	Capital stock	$ 400,000
Securities	450,000	Surplus	75,000
Real estate	45,000	Undivided profits	8,117
Furniture, etc.	15,000	Unearned interest	3,810
Reserve	252,010	Notes	50,000
		Deposits	1,415,083
	$1,952,010		$1,952,010

It will be noted that loans and discounts have been decreased by $60,000, the reserve is increased by $59,600, while the discount is reflected as a loss, decreasing the unearned interest by $400. The reserve ratio is now $\frac{252,010}{1,465,083}$, or something more than seventeen per cent, instead of thirteen per cent as it was before, showing that the desired result has been accomplished.

Similarly a bank may be able to obtain cash by borrowing from another bank on its secured or unsecured short-term promissory note. Referring to the statement on page 378, we see that if $50,000 in cash were secured in this way the reserve would be increased to $242,410, and the reserve ratio would be $\frac{242,410}{1,465,083}$, or something over sixteen per cent, as compared with thirteen per cent.

Calling loans. In contrast to the loans which have a definite date of maturity, there are loans which are made payable on demand. If a bank chooses to call for the immediate payment of such loans it can increase its reserve ratio. If payment is made in cash or in checks on another bank the reserve will increase, while payment in notes of this bank or in checks on this bank will decrease notes and deposits; in either case the reserve ratio rises.

Control by the rate of discount. Another method of regulation is by varying the rate of discount. As has been observed, in the course of a bank's daily business new loans are constantly being made and notes held are constantly maturing and being paid. There is thus an inflow and an outflow of paper, which under ordinary circumstances may about offset each other, keeping the amount of loans and discounts fairly constant though the individual notes held are always changing. Now the payment of loans on hand is already determined by their respective maturities, but anything that can be done to speed up or slow down the rate of making new loans will have power sooner or later to change the amount of loans on hand. For accomplishing this purpose one instrument which the banker has always at hand is the discount rate.

Reverting once more to the statement on page 378, let us suppose that the banker decides that his reserve ratio, thirteen per cent, is too small. He has been discounting, let us say, at six per cent, and for some time the new discounts have about offset those that were maturing. He therefore raises his discount rate to seven per cent. Since he is now charging

a higher price for the service of discounting, fewer people will avail themselves of his services in accordance with the law of demand, and the making of new discounts will fall off. The old loans and discounts continuing to mature as before, the net result will be a decline in the bank's loans and discounts. Furthermore note that when customers pay off their notes they must pay either (1) in cash or in checks on other banks which will promptly be converted into cash, or (2) in the notes of this bank, or (3) in checks on their deposit accounts in this bank. Since loans and discounts are being paid faster than new ones are being made, the bank thus finds its reserve increasing and its demand liabilities decreasing. The reserve ratio thus rises.

Conversely a reduction of the bank's discount rate may stimulate business and cause new loans and discounts to come in relatively faster than the old are being paid off.

Two qualifications should be noted in this connection. In the first place, while the bank has positive power to restrict loan expansion, its power to promote expansion is passive; it can make the borrowing terms more inviting to applicant borrowers, but the decision whether to borrow or not rests with the applicant. Oftentimes, especially in the later stages of business depression, low discount rates and other favoring acts of the banks do not avail to start revival. This principle was illustrated by the generally unsuccessful efforts to stimulate business by easy money and easy bank credit during the years from 1929 to 1939. In the second place, control of the discount rate is not, in real fact, an important method in case of regular commercial and industrial borrowers at the banks. The banks use other methods of discouraging or encouraging applicants. When money is tight, commercial and industrial borrowers are able to get funds, if at all, at practically the same rates as when money is easy. Interest rates, except in highly organized markets, are very lethargic. The central bank rate and the rate for certain types of paper broadly dealt in do vary, with the purpose of control; but in ordinary times

and in the ordinary bank the rate of discount tends to be customary and stable.

The problem of elasticity: Need of elastic currency. A peculiar and important function of bank credit is to furnish an element of elasticity to the currency system. No community's need of money is uniform at all times and seasons. In an agricultural community for example the volume of exchanges is especially great in the late summer and fall, on account of the business of harvesting and marketing the crops. Unusually large sums are being paid in wages to farm laborers, the farmers' receipts are at a maximum, and the local merchants find their sales swelling, while at the same time their credit customers are paying their bills. In a manufacturing town the weekly pay envelopes present a special need of money on Saturdays. The payment of salaries on the first of the month and the prevailing habit of settling accounts at the stores and elsewhere during the first few days of each month make a monthly variation in the need of currency. Corporations very commonly make their payments of dividends and interest quarterly, usually on the first of January, April, July, and October. There are also the cyclical fluctuations, the successive periods of business activity and business depression, of "good times and bad times." Perfectly to meet this need, the volume of currency should automatically increase and decrease in relation to these normal changes in business activity, so that no changes in the general price level will result.

Elasticity of bank credit. Recalling the classification of money set forth in an earlier chapter,[1] and the essential characteristics of the several kinds of money, the reader will realize that no item in the standard money system, except bank notes, is so related to the state of business as automatically to expand and contract with the fluctuating need of currency.

With bank credit (notes and deposits) the situation is different. The demand for loans at the bank arises directly out of business transactions, either recently completed or con-

[1] Chapter XX.

templated for the immediate future. When business is active the banks are called upon for many and large loans. And as the banks respond to the demand for loans they increase their issue of notes and their deposits. Currency thus expands to meet business needs. On the other hand, when business declines the customers of the banks have less need for currency. They pay off their obligations at the banks faster than new loans and discounts are made. In paying they either return to the banks the notes previously issued or they draw checks which reduce deposits. Currency contracts in response to declining need. Evidently there is here that close and automatic connection between volume of currency and business activity which is the mark of an elastic medium of exchange.

It is desirable that elasticity characterize both bank notes and deposits. Many people do not have bank accounts or do not understand the use of checks; in certain regions banks are few and scattered and checks are awkward to use, and for small retail purchases bank notes are more handy. Inasmuch as the volume of all such transactions fluctuates, it is desirable that bank notes vary with the needs of the people. Unfortunately in the effort of certain governments to make bank notes safe their natural elasticity has been destroyed.

The elasticity of bank credit is limited by the necessity which is upon the banks to maintain adequate reserves. As has been seen, expansion of the banks' notes and deposits reduces their reserve ratio and inclines them to slow down the rate of expansion by raising the rate of discount. On the other hand, contraction of demand liabilities raises the reserve ratio, and the bankers seek to stimulate business by lowering the rate of discount. Banking reserves thus act, not to destroy the normal elasticity of bank credit, but to put on the brakes and so to prevent extremes of either expansion or contraction.

The function of credit. The importance of bank credit has been emphasized because it is fundamental in the economic organization, but it would be an error to ignore the service of commercial credit; that is, credit extended by the

manufacturer to the wholesaler, or by the wholesaler to the retailer. In a great many transactions the banker is not called upon for credit. The manufacturer is able to take care of himself, or the producer from whom he gets his raw materials gives him time in which to work them up and pay for them.

We may in fact distinguish two points; namely, the importance of credit and the importance of a particular kind of credit. Credit, whether commercial or banking, is a sort of solvent factor in our organization, enabling us in a sense to make use of future goods at the present time. For society as a whole this cannot hold true of course, but it can for a given individual, for he gets present goods in return for a promise to pay in the future, and the transaction is conditioned on the belief of the lender in the borrower's ability to produce goods or services at some time in the future. Credit, which enables some to make present use of the surpluses of others, facilitates that smooth working of the productive apparatus which is essential to economy. It is one of the factors which promote efficiency in production and which therefore tend to keep the output of consumable goods at a high level.

Peculiar importance of bank credit. Bank credit performs a certain function in society better and more cheaply than other kinds of credit. The resources of any chain of producers are great, no doubt, but they are limited when compared with the resources of society as a whole, including the active leaders of business, those who are active in business as wage earners, and those who live on incomes derived from investments. The banks stand at the central point of the financial world and serve as reservoirs which attract the savings of all. They are in touch with more people and more lines of industry than the typical firm can possibly be. If there is for the time being a surplus of funds in one line of industry, it can be made available for other industries. Without the bank as an intermediary, how for example would the proprietor of a cotton mill in Massachusetts know that the owners of brass factories in Connecticut had funds which would be

idle for six months? And if he did discover this fact, how could such funds be made available to him? It might be done, but it would be difficult and costly. The bank stands ready to receive the funds of the Connecticut manufacturer and to pay him interest on a time deposit for six months; it can loan these directly to the cotton manufacturer in return for a promissory note of six months' duration. The thing is accomplished directly, simply, and at little expense. The accomplishment of this result is clearly an aid to production and so a service to society as a whole, not merely to the lenders and borrowers directly concerned.

Furthermore the bank serves as a selective factor in directing the investment of funds. If funds are needed more urgently in one industry than in another they are directed toward that industry, for the greater the need the higher the rate of interest which will be paid to secure a loan, and, other things being equal, a bank in making its loans will give preference to those willing to pay the higher rate of interest.

EXERCISES

1. Draw up a bank statement, using the following items: Stocks and bonds, $80,000; Undivided profits, $30,000; Reserve, $50,000; Notes, $75,000; Loans and discounts, $350,000; Real estate, etc., $100,000; Surplus, $50,000; Capital stock, $100,000; Deposits, $325,000; Other assets, $15,200; Unearned interest, $15,200.

2. Assume that a bank discounts a note for $5,000 for three months, the bank's rate of discount being 6 per cent. Payment is taken in cash. What changes would this transaction make in the bank's statement?

3. Suppose that the proceeds in Exercise 2 were taken as a credit on the customer's deposit account. What changes would the transaction make on the bank's statement?

4. (*a*) What is the reserve ratio of the bank in Exercise 1?
   (*b*) What is the bank's reserve ratio after the transaction of Exercise 2?
   (*c*) What is the bank's reserve ratio after the transaction of Exercise 3?

5. The bank of Exercise 1, after discounting the note as in Exercise 2, holds it one month and then rediscounts it with the federal reserve bank, the rate of the latter bank being 5 per cent, and the proceeds taken in notes of the federal reserve bank. Draw up a new statement showing the bank's condition after the operation. What is now the bank's reserve ratio?

6. The clearing house association in a certain city includes five banks. On a given day they present checks at the clearing house as follows:

Checks presented by:	*Drawn against:*				
	BANK A	BANK B	BANK C	BANK D	BANK E
Bank A	——	$2,100	$1,800	$2,400	$1,600
B	$1,900	——	1,500	2,100	1,800
C	2,300	1,600	——	1,800	1,500
D	2,600	1,800	2,000	——	1,700
E	1,400	2,000	1,400	1,900	——

Determine the amount to be paid or received by each bank.

CHAPTER XXII

The American Banking System

Historical background. The banking system of the United States is the result of continuous experimentation, by private individuals and the governments of colonies, states, and the nation, in furnishing money and credit to the people; it bears the earmarks of struggle and compromise among the powerful forces that have featured American national development. Chief among these forces have been insistence upon democratic ideals and local self-government, with corresponding reluctance to concede authority to the federal government, and the fear lest control of the institutions of money and credit be monopolized in the hands of a few. The general result emerging from this background is a loose decentralized system of thousands of small local unit banks, each being chartered and supervised by one of the forty-eight states or the federal government. At the head of the system are twelve central banks, federated under a scheme of shifting authority and responsibility though at best representing less than half of the unit banks. The skeptic is wont to question the propriety of calling such a conglomeration a "system," a term connoting order, articulation of parts, and central control. A true understanding of this system can best be acquired by the aid of a study of its background.

Banking in the United States before the Civil War. The so-called "banks" of the American colonial period were not real banks, but merely "batches of paper money" as one historian puts it; they did not engage in the business of discount and deposit and even in the issue of their notes did not conform to the principles of banking which our study has developed. But during the Revolution and the years immediately following a few real banks were founded, which in

addition to issuing notes made a start in the field of discount and deposit.

In 1791, two years after the formation of the national government, Congress established a central bank, the first Bank of the United States. This bank performed all the normal functions of a central bank, was the fiscal agent of the United States government, and did a regular banking business with the general public in competition with the growing company of local banks. Its notes became the principal circulating medium of the country. It was wisely and conservatively managed, performed its various functions well, and exercised also a salutary influence over the local banks, whose tendency to wild and unsound methods was apparent from the first. Indeed it was this check upon the ambitions of the local bankers and its competition with them that finally proved the bank's undoing. Opposition arose, and Congress eventually failed to renew its twenty-year charter, which expired by limitation in 1811.

For the next five years the nation's banking was wholly in the hands of local banks, chartered by the several states. These banks increased prodigiously in number. Relieved of the restraining influence of the great central bank, the majority ran wild in the paths of reckless and unsound banking. Finally in the financial panic of 1814 the whole banking structure went to pieces, a condition made doubly disastrous as it came in the midst of the War of 1812.

The people learned the lesson, their misgivings against a strong central bank were temporarily overcome, and in 1816 the second Bank of the United States was founded. This bank's record is in its essential features a repetition of that of the first Bank of the United States, including its notable services to the government and the community and its final engulfment in the bog of politics. Its charter was not renewed, and it closed its doors in 1836. There was now a repetition of the experiences of 1811–16, and the banking system collapsed once more in the panic of 1837.

Reëstablishment of a central bank proving politically impossible, the United States government cut itself loose from banking altogether by the establishment, first in 1841 and finally in 1846, of the Independent Treasury for the custody of all government funds and the management of the government's financial business. The public was left to rely for banking services upon the local banks. Gradually some order came out of the chaos, and in the older and more settled parts of the country, particularly New England, New York, and Indiana, many sound banks arose and became more or less closely associated in an efficient banking system. "Wildcat" banking continued to prevail in the frontier regions, and the country suffered grievously, especially from the multiplicity of note issues of all degrees of goodness and badness. Thus things continued down to the time of the Civil War.

The national banking system: Origin. The Civil War ushered in a new chapter in American banking history. Laws passed in 1863 and 1864 provided for a new class of banks chartered by the United States government, thus inaugurating the *national banking system.* The purpose of Congress in this legislation was primarily to aid the financing of the war and only secondarily to reform the banking system. In harmony with the first purpose, the law provided (as amended after the Civil War) that each of the new national banks with a capital in excess of $150,000 must, before commencing business, purchase and deposit with the Treasurer of the United States bonds of the United States to the amount of at least $50,000 and that each national bank with a capital of $150,000 or less must deposit an amount of bonds equal to at least one fourth of its capital. Upon the security of the United States bonds so deposited the bank was permitted to issue its notes in any amount not in excess of the par value [1] of the bonds or the capital of the bank. National banks also issued notes secured by the deposit of "lawful money."

[1] Originally and up to 1900 the limit was 90 per cent of the par value or the market value, whichever was smaller.

By an act of 1865 Congress imposed an annual tax of ten per cent upon the notes issued by state banks, thus making note issue unprofitable and reserving that function to the national banks. There was also relaxation of the Independent Treasury system to the extent of permitting deposit of government funds in selected national banks and the receipt and disbursement of national bank notes by the Treasury.

After the first few years the organization of the new national banks proceeded rapidly, and they soon assumed a position of importance in the American banking system, which they have since maintained. The national banks, the state-chartered banking institutions (state banks, trust companies, and savings banks), and the private banks have existed side by side, dividing among themselves the country's banking business.

National bank notes. It will be observed that the notes of the national banks were not issued freely upon the basis of their general assets, in accordance with the simple banking principle discussed in the preceding chapter, but were secured by a particular class of assets set aside for the purpose with a trustee. The United States law further provided for the safety of the national bank notes by imposing upon the United States Treasury the obligation to redeem them on demand, with of course the Treasury's consequent right to collect the amount from the bank of issue. For this purpose each bank was required to maintain a cash fund with the United States Treasury equal to five per cent of its outstanding notes.

While the device of bond security made the national bank notes eminently safe, this was at the expense of elasticity. Since government bonds had to be bought and deposited before additional notes might be issued, the banks could not quickly expand their note issues in time of need. This was especially true in later years, for as the national banking system expanded and the national debt was reduced all the bonds with the circulation privilege appearing on the market were

snapped up by the banks, and their prices were driven above par. On the other hand, the artificially valued bonds could not be sold in large volume to purchasers except other national banks save at a serious discount, and the banks were for the most part unwilling to hold such low-rate bonds without having the full amount of notes outstanding. One can readily see that there was no likelihood of large contraction or expansion of the volume of these notes in relation to business activity. The reader will not fail to recall however that during the last quarter of the nineteenth century the nation was passing from the use of note currency to deposit currency and that the importance of an elastic system of bank notes was gradually decreasing.

The problem of reserves. The United States has been *par excellence* the country of small independent banks; by 1913 the nation's banking business was in the hands of some 26,000 local banking institutions, with an average capital of about $80,000 and average resources of about $775,000. Branch banking was until recently forbidden by law to the national banks and in most states to the state banks and trust companies. In the absence of any system of centralized banking reserves, such as is provided in most countries by the central bank or by a small group of large and powerful banks with their far-reaching system of branches, each bank was forced to rely chiefly upon its own resources to meet whatever situation might confront it.

The national banking act required each bank to maintain a legal reserve against deposits, but for all banks except those in certain large cities a portion of this reserve might be in the form of deposits with national banks in certain of the larger cities. Quite naturally the banks in the country districts and smaller towns and cities developed the custom of depositing as much of their resources as they felt they could spare in periods of slack business with banks in the large cities of their respective regions. These city banks in their turn made deposits in other larger cities. Thus were set up, as it were,

little rivulets of currency springing from sources all over the country, uniting in the cities to form larger streams, joining again into still larger, till eventually a mighty flow of money poured into the banks of a few financial centres, the greatest of all being New York City. This flow of deposits was encouraged by the willingness of the large city banks to pay interest at a moderate rate upon deposits of other banks. The funds so obtained were loaned by the city banks to their customers, a considerable part being used for stock market speculation. On the other hand, when a succeeding period of business activity opened up larger opportunities for the outlying banks, or when a crisis threatened, they demanded the return of their reserves, and the flow of currency was reversed. This forced the city banks to call in their loans and contract their deposits and caused a state of monetary stringency in these cities. This ebb and flow, besides its minor variations, had a great annual cycle because of the need of currency for harvesting and marketing the crops.

It will be seen that, in a sense, the great city banks, particularly in New York, undertook to perform the services of a central reserve reservoir. But they were fatally handicapped by their great number, their independence of each other, their lack of any branch organization, and their inability to expand and contract their credit in the form of notes. Moreover they were unwilling to accept the responsibility of their position to the extent of keeping adequate reserves. The result was inevitable. Alternating periods of surplus currency and monetary stringency were the normal thing; inability of the outlying banks to obtain the return of their reserves in time to avoid disaster was common; bank failures and business failures were numerous; financial panics and crises were made more frequent and more severe. The situation grew steadily more intolerable with the development of the country's wealth and industry, until at last a remedy was found in the federal reserve system, which became law on December 23, 1913.

The Federal Reserve: Organization: Board of Governors. The federal reserve system does not supplant the national banking system but is superimposed upon it. It was intended to furnish the United States with a centralized banking organization, performing in general the functions of a central bank, and to give to the banking system as a whole a unity and a coherence which it previously lacked. At the head of the Federal Reserve stands the Board of Governors of the Federal Reserve System (known until 1935 as the Federal Reserve Board) with its headquarters in Washington. At present the Board consists of seven members appointed by the President with the advice and consent of the Senate. The President designates one of the members to serve as chairman of the Board of Governors for a term of four years.

Federal reserve districts and banks. The whole country is divided into twelve "federal reserve districts," with a "federal reserve bank" located in a certain city in each district. Each federal reserve bank is permitted to establish branches and agencies located in cities in its district and abroad. At present there are 25 branches. The accompanying map (Figure 23, page 404) shows the boundaries of the districts and the location of the federal reserve banks and their branches.

A board of nine directors is entrusted with the management of each federal reserve bank. Six of the directors are chosen by the member banks, three to represent the interests of the member banks and three the interests of the business and industry of the district, while the remaining three directors are appointed by the Board of Governors. One of the latter is designated "federal reserve agent" by the board and serves also as chairman of the board of directors of the bank.

Member banks. Each federal reserve bank is a corporation, and its stock is all owned by the individual banks within its district which have come into the system. These banks are known as "member banks." Every national bank is required by law to be a member, but membership is optional for the banks and trust companies incorporated under the laws of the

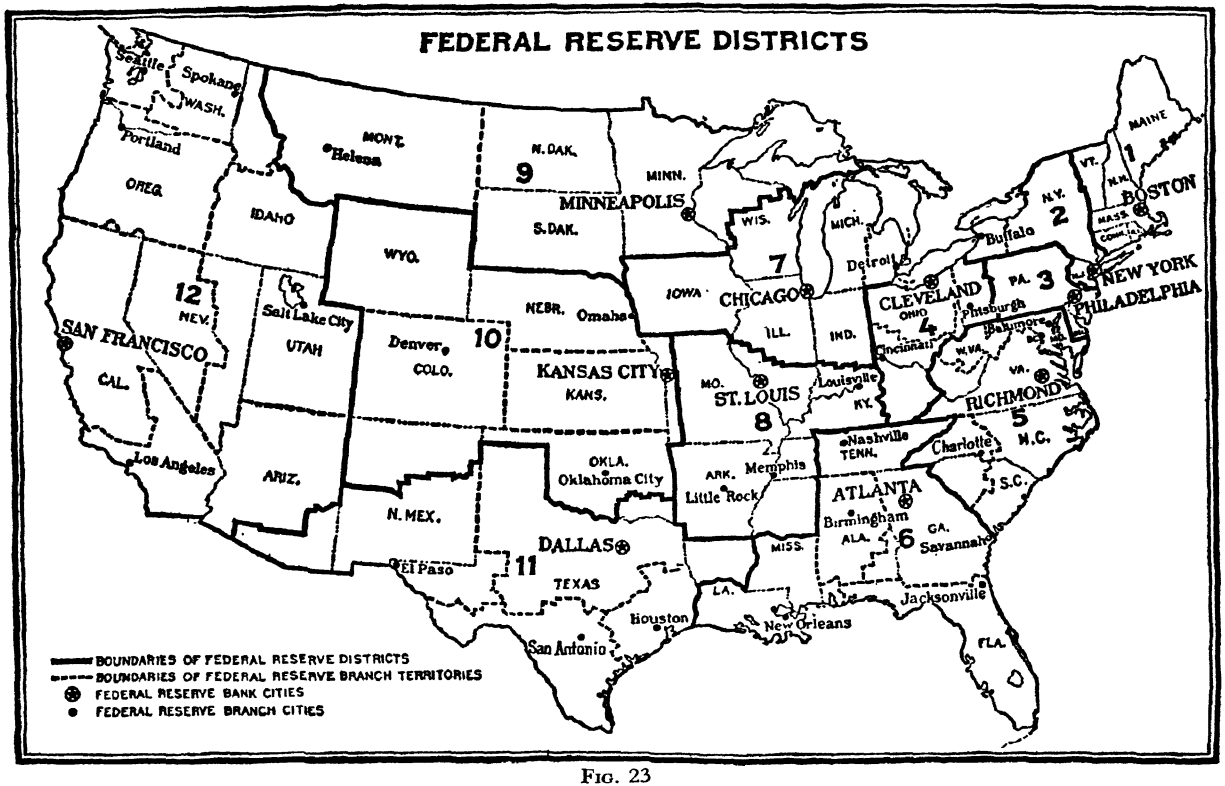

FEDERAL RESERVE DISTRICTS
Seattle
Spokane
WASH.
Portland
OREG.
MONT.
Helena
N. DAK.
9
MINN.
MINNEAPOLIS
S. DAK.
IDAHO
WYO.
WIS.
MICH.
7
Detroit
Buffalo
N.Y.
2
VT.
N.H.
MAINE
1
BOSTON
MASS.
Conn.
R.I.
NEW YORK
PA.
3
PHILADELPHIA
Pittsburgh
Baltimore
W.VA.
VA.
12
NEV.
SAN FRANCISCO
CAL.
Salt Lake City
UTAH
Denver
COLO.
10
NEBR.
Omaha
IOWA
CHICAGO
ILL.
IND.
CLEVELAND
OHIO
4
Cincinnati
KANSAS CITY
KANS.
MO.
ST. LOUIS
8
Louisville
KY.
RICHMOND
5
N.C.
Nashville
TENN.
Charlotte
S.C.
Los Angeles
ARIZ.
N. MEX.
OKLA.
Oklahoma City
ARK.
Memphis
Little Rock
ATLANTA
Birmingham
ALA.
6
GA.
Savannah
DALLAS
11
El Paso
TEXAS
MISS.
LA.
Houston
New Orleans
Jacksonville
San Antonio
FLA.
BOUNDARIES OF FEDERAL RESERVE DISTRICTS
BOUNDARIES OF FEDERAL RESERVE BRANCH TERRITORIES
FEDERAL RESERVE BANK CITIES
FEDERAL RESERVE BRANCH CITIES

FIG. 23

states. On December 31, 1939, there were 15,031 banks in the United States; of these 5,187 national banks and 1,175 state banks, a total of 6,362 banks, were members of the Federal Reserve. Less than twelve per cent of the state banks and trust companies were members, but these represented over sixty per cent of the total loans and investments of all the state banks and trust companies, showing that the largest and strongest institutions have pretty generally come into the system.

Functions of the Federal Reserve: Rediscounting and discounting. The banking business of the federal reserve system is conducted by the twelve federal reserve banks. They rediscount paper for their member banks and for each other, they hold deposits of the member banks, they issue notes, and they act to a certain extent as fiscal agents of the United States government. They are permitted also certain dealings with the general public, but their principal business is with their respective member banks, with each other, and with the government.

Rediscounting by the federal reserve banks is in harmony with the general principles with which the reader is familiar. The law makes certain restrictions as to the kind of paper that is eligible for rediscount, prescribing, up to 1935,[1] that it must be of short duration and based upon commercial, industrial, or agricultural transactions or upon loans secured by United States government obligations as collateral and bearing the endorsement of a member bank. A member bank rediscounting paper with the federal reserve bank of its district may receive the proceeds either in the form of a deposit with the bank or in money — usually federal reserve notes.

A member bank may also procure federal reserve notes or other money from the federal reserve bank or increase its deposit account in the bank by getting so-called "advances," that is, by discounting its own note backed by government securities or eligible paper as collateral.

[1] The important changes made in this feature by the banking act of 1935 are described in the next chapter.

Federal reserve notes. Two kinds of notes may be issued by the federal reserve banks, federal reserve notes and federal reserve bank notes. The federal reserve notes may be issued only on the basis and to the amount of gold certificates, commercial paper eligible for rediscount, or government bonds (up to June 30, 1941) pledged with the federal reserve agent to secure them. Of this security at least 40 per cent must normally consist of gold certificates, though on occasion the gold certificate reserve may fall below this limit, subject to a progressive tax upon the amount of the deficiency. Actually at the present time gold certificates constitute practically the entire collateral held against the outstanding federal reserve notes.

These notes, issued in the process of the bank's rediscounting and open-market operations or in payment of its deposit liabilities, are thus capable of furnishing an elastic currency. They now constitute the largest single item in the monetary circulation.

Federal reserve bank notes. End of the national bank notes. It was evidently the intent of the authors of the federal reserve act that the notes of the federal reserve banks should ultimately displace at least the greater part of the existing national bank notes. The law therefore repealed that part of the national banking act which required each national bank to invest a certain part of its capital in United States bonds. It further provided that, within certain limits, the national banks might sell to the federal reserve banks at par bonds which they had deposited to secure their notes and might retire the corresponding notes. The federal reserve banks, upon purchase of such bonds, were permitted to issue their own notes secured by the bonds. These notes are called *federal reserve bank notes* and are in all essential respects like the national bank notes which they replaced. As an alternative to issuing federal reserve bank notes, the federal reserve banks were permitted to exchange the two per cent bonds, thus bought from the banks, with the United States Treasury for

new three per cent bonds or notes without the circulation privilege. As a matter of fact this program was never carried out; the interference of the World War and the later reluctance of the national banks to surrender their circulation made it abortive. The federal reserve bank notes have never become an important part of the monetary system.

In the year 1935 the Secretary of the Treasury called for redemption all outstanding bonds having the note-issue privilege. For this purpose the Treasury used the greater part of the gold "profit" from devaluing the gold dollar which was not devoted to the "stabilization fund." This action wiped out the entire basis for national bank notes, and now as fast as they come into the Treasury they are redeemed and canceled. Thus closes one chapter in American monetary history. This action has simplified the currency system to some extent and reduced somewhat the advantage of the national bank charter.

Reserves against deposits. The federal reserve banks receive ordinary deposits only from their member banks, foreign banks, and the United States government and certain of its agencies, but deposits for the purpose of exchange and collection are received also from other federal reserve banks and from state banks and trust companies which are not members of the system. Against its deposits each federal reserve bank is required normally to keep a reserve of gold certificates (formerly gold) or lawful money of at least 35 per cent.

Each member bank is required to keep with the reserve bank of its district a deposit not less than a certain percentage of its own deposits; namely, 5 per cent for time deposits and $22\frac{3}{4}$ or $17\frac{1}{2}$ or 12 per cent for demand deposits, depending on whether the member bank is located in a "central reserve city," a "reserve city," or elsewhere. This is called the member bank's *lawful reserve*. The Board of Governors has power to alter these reserve ratios, provided that they may never be set lower than 13, 10, or 7 per cent or higher than double these percentages. All moneys in possession of the member

bank, as well as its demand balances with other banks, serve of course as working reserve, but such funds are not part of its lawful reserve as technically defined.

Open-market operations. The dealings of the federal reserve banks with the public, called *open-market operations*, consist of buying and selling certain specified items — bankers' acceptances, government securities, and gold — in the public markets. Purchases by federal reserve banks in the open market evidently tend to relieve tension in the money market, either by adding to the existing volume of money by payment for their purchases with federal reserve notes, or by increasing member banks' balances with the federal reserve banks, and thus augmenting the power of member banks to make loans and to expand their deposits. Similarly the sale of securities or of commercial paper on the open market will reduce either the amount of money in circulation or decrease member banks' balances with the reserve banks and so decrease the ability of member banks to expand their own deposit currency and perhaps even force them to contract it.

The federal reserve banks have come to rely more heavily upon open-market operations, together with credit rationing and moral suasion, than on changes in the discount rate as methods of controlling credit. This stands in sharp contrast with the practices of the Bank of England, the Bank of France, and the other great central banks, whose influence over the national discount market is exercised largely through manipulation of their own discount rates.

Centralized reserves. The federal reserve system is fairly well adapted to furnish a central reservoir and centralized control for the country's banking reserves. The only reserves against deposits now required by law of the member banks are their balances with the federal reserve banks. The federal reserve banks have, on the other hand, maintained very large reserves, generally far in excess of the legal requirements. Their gold holdings were almost from the start extraordinarily swelled as a result of the suspension of gold payments

throughout most of the world, the enormous demand for American exports, and other events of the World War. In the wake of the War the balance of exports in excess of imports remained large; moreover the political instability of Europe and other parts of the World inclined many people to try to preserve their fortunes by shifting the fluid part to America.

Mobile reserves. Not only do the federal reserve banks have a large reserve, in absolute amount as well as in relation to their liabilities and to the total monetary gold of the world, but this reserve, as well as the reserves of the member banks, is effectively mobilized. The banking reserves of the country are concentrated, available, and controlled in the public interest. Through dealings among the reserve banks themselves, directed and controlled by the Board of Governors, it is possible to direct the flow of reserves to the particular districts where for the time being the need is urgent. Whereas in former days sound liquid banks were sometimes forced to fail because they could not obtain quickly the money due them from other banks, such an event is now virtually impossible. Any soundly managed member bank can convert resources into cash sufficient to meet any run, either by rediscounting with the federal reserve bank certain eligible paper or by borrowing on its notes with United States government securities as collateral. On the other hand, a bank which has no eligible paper or government securities — a condition clearly indicating unsound management — has no means of approach to the federal reserve bank.

A national system of clearing. The federal reserve system has also become a national clearing house for settling the claims of the reserve banks against each other and for collection of checks, drafts, and other claims of individual banks against out of town banks both in its own and in the other federal reserve districts. While the system is far from universal, there were on December 31, 1939, 12,375 banks, 82 per cent of the total, which had agreed to remit at par checks sent by the federal reserve banks for payment.

The government's financial business. The reader will recall that, following the banking collapse in 1837, Congress established the Independent Treasury system, and that later a variable number of national banks were used as depositories of public funds. The federal reserve act partly removed the distinction in favor of national banks by extending the privilege of holding government deposits to state banks that were members of the Federal Reserve. Of greater significance was the provision that the funds of the United States might be deposited with federal reserve banks and that these banks might be required to act as fiscal agents of the Treasury. In 1921 the sub-treasuries were abolished and the Treasury has made progress towards curtailing its reliance upon the local banks and concentrating its fiscal business in the hands of the federal reserve banks.

The American banking system in 1929. The federal reserve banks have proved themselves an important addition to the American banking system. They assisted in the difficult task of financing the country during the World War, and they have given the banking system more semblance of order and centralized control than it ever had before, but there still remain serious defects, both in the structure and in the functioning of the American banking system, which the Federal Reserve has failed to remedy. Even in the years of prosperity prior to the crisis of 1929 the existence of these defects was recognized, but their seriousness was not then generally appreciated. It is important that we examine closely the details of American banking in this period in order to understand the causes of a failure without parallel in the history of banking and to appraise intelligently the reforms which have been made in recent years.

Composition of the banking system. Reference has been made to the fact that the United States has always been a nation of small independent banks. The numbers and resources of the various classes of banks which make up the local banking structure, as of June 30, 1921, and June 30,

1929, are shown in the following table, which contains also statistics of membership in the federal reserve system.[1] Corresponding figures as of October 2, 1939, will be found in the next chapter.

BANKING SYSTEM OF THE UNITED STATES, JUNE 30, 1921 AND 1929

I. All Banks

	Number		*Resources (in billions)*	
	1921	1929	1921	1929
National banks	8,154	7,536	$20.5	$27.4
State banks	18,875	14,437	14.2	16.8
Loan and trust companies	1,474	1,608	8.2	16.1
Stock savings banks	978	747	0.5	1.5
Mutual savings banks	623	611	6.0	10.0
Private banks	708	391	0.2	0.2
Total	30,812	25,330	$49.6	$72.0

II. Federal Reserve Membership

	Number		*Resources (in billions)*	
	1921	1929	1921	1929
National banks	8,154	7,536	$20.5	$27.4
State member banks	1,595	1,171	10.0	18.5
Total member banks	9,749	8,707	$30.5	$45.9
Non-member banks	21,063	16,623	$19.1	$26.1

Omitting private banks and savings banks, the country's banking business was handled in 1929 by something over 23,500 banks. About a third were national banks, having average resources of about 3½ million dollars and owning slightly over 40 per cent of total bank resources, while the remainder were state banks and trust companies with average resources of around 2 million dollars. This is in marked contrast with the situation in most other countries; in Canada, for example, where in 1928 there were 10 large central banks, with average resources of nearly $350 millions.

[1] *Annual Report of the Comptroller of the Currency, 1929.* The figures used in the remainder of this section are also derived from this source.

Even these figures do not show as convincingly as do the figures on bank capitalization the extent to which the small bank prevailed. In 1929 only 28 per cent of the banks of the country had capital of $100,000 or over; of the remaining 72 per cent, 21.9 per cent had less than $25,000, 21.5 per cent had exactly $25,000; with the rest between $25,000 and $100,000. Over 80 per cent of the banks were located in towns of 10,000 or less population, but these banks accounted for only 23.4 per cent of the total capitalization.

Branch banking and chain banking. Without doubt the characteristic decentralization of banking in the United States would not have persisted as it has had branch banking been generally permitted. Only in recent years have there been any important steps away from the prevailing condition of local decentralization. In California and certain other states the law has permitted branch banking, and great systems of branches have been developed, state-wide, county-wide, or city-wide. Competition of these state banking institutions finally forced Congress to grant to the national banks certain branch-banking powers. The McFadden act of 1927 provided that in any state which permitted branch banking both national and state members of the federal reserve system might, with certain limitations as to national banks, establish and maintain branches within the limits of the city where the parent bank was located.

By 1929 there were 818 banks operating 3,440 branches in the United States, thus making the number of banking offices greater by this number than the total number of banks given in the table above. The great majority of the branch offices were situated in the local communities of the respective parent banks. While these banks with branches formed an important part of the banking system, owning about 40 per cent of the loans and investments of all banks, branch banking had not proceeded extensively enough to alter the essential character of the American local banking system.

Another development of considerable importance was the

establishment of elaborate "chains" or "groups" of banks, held together by various devices, the most common being the holding company. This was a device which in fact permitted branch banking, but without the sanction of state or federal law. These bank holding companies were not subject to federal control and supervision, until made so by the banking act of 1933. In June, 1929, there were 275 such chains or groups of banks in the United States, controlling 1,802 unit banks, which in turn owned 40 per cent of all the branches in the country. The loans and investments of all banks controlled by chains or groups amounted to about 15 per cent of the loans and investments of all banks. Nevertheless, important as they were, the chains did not permeate the banking structure sufficiently to be a major factor in the situation.

Weaknesses of decentralized banking. A system of small, independent, local banks contains inherent weaknesses. The multiplication of banks divided minutely the country's banking capital and deposits. Division among so many banks made many of the local banks so small as to defeat effective management and service and cause weaknesses, both of the individual banks and of the system, that were bound in the end to spell disaster. Towns were jealous of one another, and so each had to have its bank. Dissatisfied customers of that bank or ambitious local capitalists or adventurers would start a second or a third bank in a town that could not decently support one. Banking became one of our most overexpanded industries. Competition between these banks, many in a failing condition, was severe, ruthless, and unintelligent. High interest rates were paid on deposits, and voluminous services were performed without charge or at a loss. Interest rates on loans were frequently exorbitant, by sheer necessity, and loans could be made only locally and to necessitous borrowers at such rates; loans and discounts in the better types of paper in the central money markets were denied by similar necessity, and the investment portfolio had to consist largely of nominally high-yield bonds and mortgages regardless of safety or

liquidity. The banks competed in "service to the community," which largely meant buying local securities and making loans on local projects based on sanguine faith rather than intelligent analysis of the borrowers' ability to repay. Thus the bank's portfolio of loans and investments was composed of local items, and the life of the bank was made contingent upon the stability, permanence, and success of the small locality. Bank buildings were commonly disproportionate to the rest of the town in architecture and equipment, and the staff had to be excessive. The salaries, on the other hand, were too small to be attractive to men of adequate training and capacity. Anyone could go into banking with little capital and less of experience and training. This situation was of course not characteristic of all the banks, or even of all those in small places. It was, however, general enough to present a very serious situation.

The small-town state banks were on the whole indifferent, if not hostile, to the federal reserve system. In 1929 only seven per cent of all the state banks belonged to the system, and these were the larger banks. The national banks in small towns or cities had a difficult time competing with the state banks, for in many states the banking laws were more liberal and the supervision more lax. This made the way of the national bank supervisors difficult, since severity was frequently answered by threats to denationalize and take out a state charter. The difficulty was partly adjusted by lowering the national bank standards (by the act of 1927 and other amendments) as well as by less punctilious examinations.

Affiliates. Another defect in the structure of American banking was the close affiliation of commercial banking with other forms of enterprise. Not infrequently close relationships between commercial banks and companies dealing in securities or in real estate were created by the holding company device; in other cases the bank itself owned the controlling stock in an affiliated company. It is the general concensus of opinion that a commercial bank should be divorced from investment business, real estate promotion, and other such interests,

especially if it has a pecuniary interest in the profits of the other company. Indeed it is difficult to see how an officer of a bank can bring an unbiased judgment to bear on the wisdom of making a loan to a customer which will enable him to purchase an issue of stock or a piece of real estate offered for sale by a company affiliated with the bank.

Expansion of credit. Among the financial features of this period which had a bearing upon the banking situation in 1929, we note first a gigantic expansion of credit in the United States, marked by an increase of about fifteen billion dollars in loans and investments of commercial banks, with the total for all member banks of the federal reserve system rising to thirty-six billion dollars. Such an expansion was made possible by the inflow of gold from countries off the gold standard or on a gold-exchange standard and by the "easy money" policies of the federal reserve banks. This federal reserve policy was dictated partly by political and industrial conditions in the United States and partly by a program of coöperation between the Federal Reserve Bank of New York and the Bank of England. On the basis of the excess reserves came a great expansion of bank credit. Most of this increase in credit was neither needed nor used for normal short-term commercial operations but went into speculation in real estate, speculative construction, and the taking up of a tremendous issue of securities, both domestic and foreign.

Lending and investing abroad was a rather novel practice for Americans, but at the end of 1932, after considerable reductions from both default and resale, the private foreign investments of the United States amounted, at conservative estimate, to fifteen billion dollars, of which seven billions were represented by security investments in the ordinary sense and eight billions by direct ownership of capital assets.

The annual lending of immense sums abroad had permitted the merchandise export balance of the United States to expand, in spite of rising trade barriers and against the influence of net payments due the United States because of its creditor

position. This made possible large exports of heavy mechanical equipment and of agricultural products, which served to maintain these types of enterprise in precarious balance, despite the overexpanded condition of their productive capacities in relation to the domestic market and in relation to the reviving and expanding productive capacities of other parts of the world. The new annual loans also served to conceal the inherent difficulties — both as to budget problem and transfer problem — of payments on the intergovernmental debts and reparations in a world of rising trade barriers and badly distributed monetary stocks of gold.

The financial crises of 1929 and 1931. Despite persistent depression in agriculture, industry in the United States had with minor exceptions been fairly prosperous since 1920, but in 1929 the tide turned. Funds in the United States were increasingly in demand for speculation in the domestic security markets. It was becoming evident that the structure of international finance was unsound. With foreign lending reduced, the export industries of the United States were either losing their foreign markets or saving some of them by taking new credit risks and selling on open account; and the unnatural export balance of the United States was declining. By the summer of 1929 most lines of business activity in the United States were showing some recession, and this was fatal to a domestic structure of speculative credit which was based on expectation of everlasting expansion, of accelerated profit taking, and of security values which could move only up. Liquidation made a sudden start on its predestined long journey, foreign lending came to a standstill, and the postponed and now intensified problems of post-war economic readjustments presented themselves for fresh consideration.

In the summer of 1931 a crisis was reached in international financial relations, with successive breakdowns in Austria, Germany, and Great Britain; this was despite considerable international coöperation, in which the United States participated by moratoria on the allied war debts and on com-

mercial debts, as well as by special advances to foreign central banks. Great Britain left the gold standard in September, 1931, and was followed by upwards of thirty countries, leaving a slender "gold bloc" consisting of France, Holland, Belgium, Switzerland, and Poland. There was universal withdrawal of balances of short-term capital from foreign banks, especially from the United States. Among the consequences of this action were renewed credit liquidation, heavy cash withdrawals from the banks, and increased failures in the United States.

Bank failures. From 1921 through 1932 10,816 banks suspended. This is the worst record of bank failures in the world's history, despite the fact that the United States was the wealthiest nation in the world and had gained most and suffered least from the World War. With a sound banking system this crash would have been utterly unnecessary. Neither England nor Canada was completely insulated from the influence of the forces we have been studying, yet not a single bank failed in England because of the depression, nor in Canada.

Analysis of the bank suspensions since 1921 shows that the rate of failure was higher among non-member than member banks, among banks in the smaller ranges of capitalization, among banks in the towns and cities of small population, among banks in agricultural areas as compared with industrial territory, and among state banks as compared with national banks because of their smaller size and looser supervision.

Since the beginning of the depression of 1929–36 bank failures have been more evenly distributed over the country than they were before. In this latter period the industrial areas were as badly affected as the agricultural regions, or even worse, and the investment portfolios of all banks were disastrously impaired by the radical decline in the values of securities and real estate.

CHAPTER XXIII

American Banking since 1929

Banking legislation since 1929. Great importance attaches to the efforts of Congress and the state legislatures since 1929 to mend the weaknesses of the banking system which were demonstrated by the stock market boom of 1926–29, the panics of 1929, 1931, and 1933, and the depression of 1929–36. This legislation, especially that of the federal government, not only amended the structure and the practice of the Federal Reserve and the member banks but also put into effect new theoretical conceptions of the nature and functions of central banks.

The federal legislation of this period had six general objectives: (1) to meet the banking emergency wrought by the panic; (2) to promote recovery from the depression; (3) to strengthen the hand of the Board of Governors and the federal reserve banks in obviating or restraining excessive booms in the securities markets like that which broke in 1929; (4) to centralize authority and responsibility and to increase the capacity of the Board of Governors to control money and credit; (5) to improve the banking structure and the practice of member banks, with a view to greater stability and better service to agriculture, industry, and commerce; and (6) to fill certain gaps in the national financial organization which, in the opinion of the administration, private initiative had left unfilled.

Pre-"New Deal" remedies: National Credit Corporation and Reconstruction Finance Corporation. The increasingly critical condition of American banks, evidenced by the failure of 1,072 banks in the five months beginning with September, 1931, and the intensification of hoarding, led the United States government to take various measures to support the credit situation. The first of these was the organiza-

tion in October of the National Credit Corporation for self-help among the banks. This involved the pooling of resources so that loans might be made on call to distressed members. The National Credit Corporation did not however prove to be of much service, and in February, 1932, it was replaced by the newly created Reconstruction Finance Corporation.

The RFC had an initial capital of 500 million dollars, subscribed by the government, and was authorized to borrow on its own obligations up to three times the amount of its capital. It was empowered to make loans to banks, building and loan associations, insurance companies, mortgage companies, federal agricultural banks, agricultural and livestock credit corporations, and to the railroads and farmers. Later in the year the lending power of the corporation was broadened to permit loans for self-liquidating public works undertaken by states, municipalities, and other public or semi-public agencies. The central purpose of the corporation was to use government credit for thawing frozen private credit.

Glass-Steagall act. Early in 1932, as a result of continuing large withdrawals of foreign balances in gold, with most of the countries of the world already off the gold standard, there was considerable discussion of danger to the gold standard in the United States. The Glass-Steagall act of February, 1932, was proclaimed as a safeguard against this doubtful danger. Enacted as an emergency measure with a time limit of one year (later extended), this act permitted: (a) advances by federal reserve banks to member banks on paper usually ineligible; and (b) the substitution of government obligations for commercial paper or for that amount of gold (above the 40 per cent minimum gold requirement) used as security for federal reserve notes.[1] By this latter provision the act added 750 million dollars to the "free gold" of the federal reserve system. The federal reserve banks were now in a position, without fear of gold withdrawals, to pursue an intensified

[1] The right to use obligations of the United States as collateral for federal reserve notes expires on June 30, 1941.

"easy money" policy, and they did so with a program of heavy purchases of government obligations.

Banking collapse and government action. In the early months of 1933 the American banking system virtually broke down. Among the causal events leading up to this crisis the following are especially deserving of mention: accumulation of huge deficits in government financing, a substantial decline in commodity prices, bold discussion in Congress of various fiat (irredeemable) money proposals, the suspension between 1929 and July, 1932, of approximately a fifth of all the banks operating in the earlier year, further increase of bank failures after November, 1932, loss of gold by export, and an unprecedented popular movement toward the hoarding of gold. In January and February, 1933, there were mass closings of banks by cities and states. Finally on March 5 came the proclamation of President Roosevelt closing all the banks in the country for a period of four days beginning on March 6 and placing an embargo on gold withdrawals and exports. Congress was called in special session on March 9 and on that day passed the emergency bank act, giving the President emergency powers over the banking system.

The bank closing was indefinitely extended in order to give the Secretary of the Treasury time to classify the banks as to their condition for reopening. In the course of a week the condition of the individual banks was checked, and the banks which were declared fully solvent were permitted to resume regular operations, aside from paying out gold. Many other banks were opened immediately, but with some restrictions on their operations, which permitted them to accept and make payment on new deposits, but not to make payment on old deposits until their cash position had improved. Other banks were kept closed until their impaired capital positions could be corrected. So great was the confidence created by this process of examination and controlled reopening that the operating banks found practically normal banking conditions; no serious currency withdrawals were even threatened.

To allay the fears of a money panic caused by the ever increasing hoarding, the federal reserve banks were given power to issue federal reserve bank notes against government obligations (up to the face value) or against notes, drafts, bills of exchange, and bankers' acceptances (up to 90 per cent of the face value) deposited with the federal reserve bank as security. This was distinctly a temporary expedient, and only slight use was made of the privilege, although the effect of the act was to remove certain safeguards which had been established for normal times against undue extension of credit. Under the act the federal reserve banks could take over unlimited issues of government obligations and pay for them by a dollar for dollar issue of new federal reserve bank notes; and they were further empowered to make direct loans to others than member banks on promissory notes of the borrowers if secured by government obligations as collateral.

Under the supervision of the Secretary of the Treasury and the Controller of the Currency, banks in need of new capital might sell cumulative preferred stock, without double liability, or debentures or capital notes, to the RFC. The sale of these securities improved the capital structure of many banks and enabled them to reopen, to qualify for membership in the Federal Deposit Insurance Corporation, and to lend in larger amount and with less hazard.

For banks unable by any of these provisions to get into satisfactory condition for reopening, conservators were to be appointed, with all the powers of a receiver to protect the interests of the depositors and to prepare for reopening or eventual liquidation if necessary.

The emergency bank act of 1933 also gave the President control of all transactions in foreign exchange in any national emergency, gave the Secretary of the Treasury power to call in all gold and to impose penalties for non-compliance, and put the operations of member banks in the federal reserve system under the prescriptions of the President and the Secretary of the Treasury in any emergency.

Provisions for meeting future panics. Following the emergency, two important banking acts were passed in 1933 and 1935, which have a more or less permanent effect upon the banking system. The banking act of 1935 adopted the principle of the Glass-Steagall act of 1932 for alleviating crises and made it a permanent part of the federal reserve system. This act provides that a federal reserve bank, under rules and regulations prescribed by the Board of Governors of the Federal Reserve System, may make advances to a member bank on its time and demand notes which have maturities of not more than four months and which are secured to the satisfaction of the federal bank. Each such note must bear interest at a rate not less than ½ per cent higher than the highest discount rate in effect at the federal reserve banks on the date of the note.

This authority is a continuing privilege to the member bank to use any of its assets as security for a loan from the federal reserve bank at any time, provided that it can satisfy the reserve bank of the desirability of the loan and the soundness of the collateral and is willing to pay the penalty rate on such a loan. Such a privilege is consonant with the practice of the Bank of England, which makes "lombard loans" at rates above its "bank rate" on prime paper.

Structural reforms of 1933 and 1935: Deposit insurance. In addition to emergency features, the banking legislation of 1933 and 1935 contained many important provisions intended to bring about permanent reforms in the banking system. We may examine the more important of these under the two heads of structural reforms and functional remedies.

Prominent in the former group was the establishment of a nationwide system of insuring bank deposits. Though growing directly out of the banking emergency of 1933, this device was heralded as a preventive of similar panics in the future. The system established is not a government guaranty of deposits but the self-insurance of their deposits by the banks themselves. Every national or state member bank is required

to be an insured bank subject to the provisions of the law. The plan is open to non-member banks, but no state bank which during the calendar year 1941 or any succeeding year has average deposits of $1,000,000 or more may be an insured bank or continue to have any of its deposits insured after July 1 of the year following any such calendar year, unless the bank is a member of the Federal Reserve.

The Federal Deposit Insurance Corporation was authorized at its discretion to operate a separate fund for the benefit of the mutual savings banks and their depositors. Conforming to the authority, such a fund is now operated in a wholly separate manner for the mutuals alone, and the capital assets of the corporation, as well as all its expenses, are allocated between the funds on an equitable basis. Such separate treatment of mutual savings banks was warranted by the fact that their deposits are all time deposits and that the mutuals operate, as experience shows, with a very low loss ratio. The scheme has not been popular with the mutuals; the corporation in 1938, for instance, was insuring the deposits of only 48 mutual savings banks out of a total of 554.

The maximum amount insured for any depositor in one bank is $5,000. The assessment rate is one-twelfth of 1 per cent per annum for commercial banks, applied to the average net deposits for the preceding six months of the calendar year. The corporation may, after notice, suspend an insured bank from the insurance protection and publish the fact if it continues unsound or unsafe practices or knowingly or negligently permits violation of the law or regulations. Any insured bank, upon due notice, may voluntarily withdraw from the insurance plan. Whenever a bank ceases to be a member of the federal reserve system, its insurance stops forthwith.

On December 31, 1938, of the 18,220 commercial banks and branches in continental United States 17,073, or 93.7 per cent, were insured. These banks had 97 per cent of all the deposits; but of course not all these deposits were insured, since accounts are insured only up to $5,000.

Judgment of deposit insurance. The deposit insurance plan has met with severe criticism from bankers and economists, and indeed the experience in the United States with various schemes for the guaranty of bank deposits would seem to warrant misgivings as to its right to a permanent place in the banking structure, however essential it may have been considered in 1933 as a means of overcoming the quite natural timidity of depositors and in counteracting the trend towards hoarding. A number of states in the United States have experimented with guaranty of bank deposits, but in every case the standards in banking practice were lowered. In every case the plan quickly broke down and was abandoned. The continued operation of the schemes would have put all solvent banks in bankruptcy.

Whether similar results will follow the adoption of the insurance plan by the federal government is of course impossible to predict. In a national scheme there is a much greater diversification of risk than in a state scheme, where all the banks are frequently dependent on a single type of loan and in danger of all getting into difficulty together if any economic adversity strikes the leading industry of the region. There are those who find encouragement in the thought that the plan will drive state banks into the federal reserve system, believing that depositors will shun those banks not covered by deposit insurance and that thereby the banking system as a whole will be unified and strengthened.

The jurisdiction of the Federal Deposit Insurance Corporation offers of course a new avenue of approach to the exercise of a centralizing influence over most of the 8,722 state banks which are not members of the federal reserve system. So long as they feel the need of deposit insurance, the failure of these banks to join the Federal Reserve need no longer hinder their practical subjection to many of the regulations of that system. The possible centralizing influence is further strengthened by the opening of membership in the federal reserve system to Morris plan banks and others of the same type and to mutual

savings banks that qualify. Again it may be expected that the supervision of the banks under the Federal Deposit Insurance Corporation will be stricter than has been imposed upon the state-chartered banks by some of the states. Rigid standards set up by federal inspectors are expected by some to compensate for the laxness in management which others believe the plan encourages and condones. Nevertheless deposit insurance is at best a poor substitute for a thorough reorganization of the banking system, national and state, such as would make insurance or guaranty plans unnecessary.

Capital requirements. As a step toward the elimination of difficulties arising out of the keen competition in a banking system composed of very many small and frequently undercapitalized independent banks, recent legislation has altered capital requirements. Minimum capital for national banks has been increased in all categories, becoming $50,000 for a bank in any place having a population of less than 6,000. State banks are excluded from membership in the Federal Reserve unless the national bank capital requirements are met. The banking act of 1935 further requires that newly organized national banks must have surpluses equal to 20 per cent of their capital stock and that every national bank must build up its surplus out of earnings to a level of 100 per cent of capital stock instead of the 20 per cent level formerly required. In one respect this accumulation of larger surplus is offset by the loss of the security afforded to depositors by the double liability of shareholders, which has now been abandoned.

Branch banking. Branch banking received modest encouragement in the banking act of 1933 by the grant to national banks of authority to establish and operate branches either in the city or throughout the state in which the bank is located, provided that similar authority to maintain branches is specifically granted to state banks by the laws of the state in question, and subject to severe restrictions, intended to limit the establishment of branches in small places or by small

banks. A number of states have liberalized their banking laws in consequence of this action, and branch banking promises to spread when the depression lifts.

Holding company affiliates. The banking act of 1933 gave some attention to the related problem of chain banking; that is, the problem arising from the control and direction of a group of banks by a holding company. The act requires such a holding company, called a holding company affiliate, to secure a permit from the Board of Governors before it may vote the stock which it owns in a member bank. To obtain the voting permit the holding company must submit to examination like any national bank, must build up assets other than bank stock, and must dissolve any connection with a security selling organization. Member banks are restricted in the making of loans to a holding company affiliate, in the purchase of its obligations, and in the acceptance of such obligations as collateral for loans to others. Of any group of member banks affiliated with a holding company only one may participate in the nomination and election of directors of a federal reserve bank. Some measure of control was thus extended over a form of banking organization which had been open to considerable abuse.

Divorce of commercial and investment banking. The reader was reminded in the last chapter that the keenly competitive search for banking profits had developed unwise policies through the granting of credit by commercial and investment banks. Commercial banks devoted more and more of their resources to strictly speculative loans, to real estate loans, to investment holdings, and even to dealings in securities. The shift of funds away from strictly commercial banking functions was intensified under the conditions of business depression.

The banking act of 1933 sought to correct these defects. Limits were placed on loans by any member bank to its own officers, loans to security affiliates were forbidden, and the amount of the obligations of such affiliates which might be

accepted as collateral on loans was restricted. Another group of provisions was intended to effect a divorce between commercial and investment banking, their marriage having resulted in some undesirable offspring during the speculative boom and the following depression. In this connection, member banks were forbidden to engage in the investment banking business, a limit was placed on the amount of any one issue of securities that a member bank might hold on its own investment account, and it was stipulated that no member bank might be affiliated in any way with an organization engaged principally in the issue or distribution of securities, and that no officer or director of a member bank might be an officer, director, or manager of any organization engaged primarily in the business of buying, selling, or negotiating securities, unless with permission of the Board of Governors. No enterprise engaged in investment banking was permitted to engage in the business of receiving deposits. The banking act of 1935 slightly relaxed some of these prohibitions and gave the banks a limited privilege of underwriting and dealing in securities.

Functional remedies of 1933 and 1935: Interest on deposits. Certain functional remedies are contained in the recent banking legislation. The act of 1935 prohibits member banks from paying interest on demand deposits and directs the Board of Governors of the Federal Reserve System to limit by regulation the rate of interest paid by member banks on time and savings deposits. In addition the Federal Deposit Insurance Corporation must by regulation prohibit the payment of interest on demand deposits in insured non-member banks and similarly must limit from time to time the rates of interest paid on time and savings deposits. The corporation and the Board of Governors have coöperated in fixing the same rates in any territory, and they have also secured the cooperation of the state banking departments in setting rates for non-insured banks.

This change in the operation of the banking business is of

considerable importance. The interest allowed on deposits was frequently a chief competitive weapon among banks in their attempts to secure an adequate volume of business to provide a fair return on capital. There is no justification for interest on the demand deposits of a commercial bank; such payments merely encourage the keeping of unnecessarily large balances in a form which is most dangerous for the bank.

Speculative use of credit. Still another group of provisions in the act of 1933 attempts to set limits to the speculative use of credit. The requirement that federal reserve banks shall discount for members without discrimination is changed to say that they may discount for members; the reserve banks are hereafter to watch the use of credit by members, and the Board of Governors may put an embargo on credit to any member found making an undue use of credit for speculative loans. Limitations on the making of loans on security collateral include a prohibition of the practice of placing such loans with brokers or dealers in the capacity of agent for a non-banking organization. Leading bankers have remained skeptical of the effectiveness of these provisions to accomplish adequately their intended purpose. The stock market boom of 1936–37, with the precipitate crash of 1937–38, seems to support this skepticism.

Centralized control of the banking system. The Board of Governors had not, up to 1933, attained to that dominance in the banking system which was regarded by many, including the sponsors of the "New Deal" banking legislation, as desirable. A number of changes were accordingly made, designed to bring about greater integration within the Federal Reserve System itself through a strengthening of the control of the Board of Governors. This control was fortified by the power granted the board to alter reserve requirements, to regulate the payment of interest on time and savings deposits, to exercise supervision over credit for speculative purposes, to establish regulations governing the standards for rediscounts and loans by federal reserve banks to members, to set the redis-

count rates of the banks, and to control through the Open-Market Committee all open-market operations and impose its decisions upon the federal reserve banks. The control was further strengthened by the power granted in 1933 to order the removal of any director or officer of a member bank if after warning he continued to violate any banking law or continued "unsafe or unsound practices."

Control of credit: Concept of credit control. The founders of the Federal Reserve sought, by strict definition of paper eligible for rediscount, a *qualitative* control of credit. In contrast the greater emphasis in recent years has been laid on *quantitative* control. The "New Deal" has seemed less anxious about the direction which credit is to take and more zealous to increase the quantity outstanding, to flush the bank reserves and ease the money market, to enlarge purchasing power and thereby to promote recovery. The philosophy of the Federal Reserve has thus been shifted, in emphasis at least, to quantitative control, exercised chiefly through open-market operations and reserve manipulation.

Open-market operations. Experience has taught that the most effective instrumentality of control, permitting positive action by the authorities rather than passive waiting for requests from member banks for advances or rediscounts, is the open-market purchase and sale of government securities and acceptances. The new importance attached to quantitative control required that the open-market machinery be brought more fully and directly under the centralized control of the Board of Governors. This was accomplished by the banking acts of 1933 and 1935. A Federal Open-Market Committee has been set up, consisting of the members of the Board of Governors of the Federal Reserve System and five representatives of the federal reserve banks. The committee adopts and transmits to the federal reserve banks regulations relating to the open-market transactions of such banks. The open-market operations of every federal reserve bank must be in accordance with the direction and regulations of this com-

mittee. The time, character, and volume of all transactions in paper eligible for open-market operations must be governed with a view to accommodating commerce and business and with regard to their bearing upon the general credit situation of the country. The securities eligible for open-market operations are extended to include all those fully guaranteed by the United States government.

Regulation of reserve ratios of member banks. Second in importance among the instrumentalities of quantitative control of credit is regulation of the member banks' reserve ratios; *i.e.*, the percentages of deposits which they must hold as reserve with the federal reserve banks. Under the banking act of 1935 the Board of Governors of the Federal Reserve System may, "in order to prevent injurious credit expansion or contraction," change by regulation the percentage requirements, provided that it cannot set them below 13 per cent, 10 per cent, or 7 per cent of demand deposits, depending upon the location of the bank, and 3 per cent of time deposits and that they cannot be increased to more than twice those amounts. If a bank required to carry a 10 per cent reserve were directed to jump its reserve to 20 per cent, its loaning capacity would be reduced, and a lowering of the reserve percentage later would enhance this capacity. Of course, instead of curtailing loans when the reserve requirement is increased, the bank may sell securities from its portfolio and transfer the proceeds to its reserve account.

This device is even more direct and powerful than open-market operations in creating or absorbing excess reserves; moreover by reducing reserves and forcing members to become indebted to the federal reserve banks, the board can make changes in the rediscount rate more effective in credit regulation. Such control is a new device in central bank technique; how useful it may prove to be is problematical.

Loans on real estate. Loss of interest in qualitative control is further indicated by a provision of the banking act of 1935 which greatly extends the authority of national banks to loan

on real estate security. This represents a departure from the notion of keeping these banks strictly commercial. A national bank is now authorized to make loans secured by first lien upon improved real estate in amount not exceeding 50 per cent of the value of the real estate; the maximum period is five years, except for amortized loans, which may run for ten years if the installments are such as will amortize at least 40 per cent of the principal of the loan within the ten-year period. These restrictions do not apply to insured real estate loans made under the provisions of Title II of the national housing act. The aggregate amount of real estate loans may not exceed the capital and surplus of the bank or 60 per cent of the amount of its time and savings deposits, whichever is the greater.

The relaxation of restrictions on the making of real estate loans was an act of doubtful wisdom. Such loans are lacking in liquidity and are in consequence a type of investment which the authorities have generally considered the commercial banks should avoid. The wholesale bank suspensions and other banking difficulties in the United States in recent years were aggravated by the unwarranted extent to which many banks had diverted funds to loans on real estate. Between the middle of 1918 and the middle of 1929 member banks increased their real estate loans by nearly 600 per cent, while their total loans and investments were increasing by a little less than 100 per cent. At the latter date real estate loans constituted 9 per cent of the total loans and investments of these banks. To encourage still further the making of such loans is definitely to weaken the commercial banking structure.

The present banking situation. The foregoing survey of recent banking legislation shows that some of the weaknesses of the American banking system described in the immediately preceding chapter have been remedied but that at the same time certain new elements of weakness have been introduced. The table on the following page shows the composition of the banking system as a whole as of October 2, 1939.

BANKING SYSTEM OF THE UNITED STATES, OCTOBER 2, 1939

	Number	*Deposits*	*Loans and Investments* (In Millions of Dollars)
Member banks			
National banks	5,196	$25,248	$21,289
State banks	1,143	14,039	11,786
Total member banks	6,339	39,287	33,075
Non-member banks			
Mutual savings banks	552	$10,520	$10,333
Other non-member banks	8,170	7,630	6,546
Total non-member banks	8,722	18,150	16,879
Total, all banks	15,061	57,437	49,954

While the total number of banks has been reduced from 25,110 in June, 1929, to 15,061 in October, 1939, there still remains that condition of many independent local banks of small average resources which has always represented a fundamental weakness in the American banking system. There remains the imperative need of reduction in the number and increase in the average size of the commercial banks. These results could be brought about either through consolidation of competing banks or through a thoroughgoing extension of branch banking. Some form of branch banking on a wide regional or national basis would appear to be both salutary and inevitable, and certain progress in this direction has already been noted. In some manner the competitive chartering of banks by both federal and state governments should give place to a uniform national system. At the head of such a system should stand, as now, the Federal Reserve; possibly it could be more efficient if organized as one central bank with branches. Certainly the Federal Reserve should be removed so far as practicable from political influence to the end that it may serve the commercial banks and the national treasury, in a manner illustrated by the Bank of England.

Other forms of government banking: Agricultural credit agencies. We may conclude this survey by examining certain collateral banking functions which have been undertaken by

the federal government. In the special banking field of agricultural credit the federal government has also directly assumed new or amplified banking functions. The various agricultural agencies were consolidated in a new Farm Credit Administration under a single Governor, for the purpose of improving the efficiency of the credit service. The Administration, in addition to the Governor, consists of the Land Bank Commissioner and the Intermediate Credit Commissioner, each of whom heads a centralized system of credit institutions. The institutions thus consolidated consist of (1) the twelve federal land banks — long-term mortgage banks, owned by some 3,000 coöperative national farm loan associations scattered over the country; (2) the joint stock land banks — long-term mortgage banks, similar to the federal land banks, but privately owned and operated and to be liquidated, no new loans being permitted; (3) the twelve federal intermediate credit banks — operating in the same districts, housed in the same buildings, and governed by the same directorate as the federal land banks, but lending for periods from one to three years on agricultural products and livestock; (4) twelve production credit corporations, supplied with government funds and authorized to invest in the stock of, make loans to, and in large measure control the activities of local production credit associations, which lend to farmers for general agricultural purposes; (5) the Central Bank for Coöperatives and twelve regional banks for cooperatives, with funds largely supplied by the government, authorized to make loans to coöperative associations which promote the orderly marketing of farm crops by holding, storing, processing, and marketing farm products and by making loans to their members on the security of farm products deposited with them for eventual marketing.

The federal government is thus prepared to perform functions which in times past were considered to be exclusively within the province of the banks, but which were not satisfactorily performed by them during the long period of bank

closings and liquidations in agricultural areas, if ever. It seems that much of the financing of agricultural operations — of long, short, and intermediate maturities — is permanently taken from the state and national banks and the former loan institutions.

The Congress went still further and created the farm mortgage corporations, to issue and use two billions of dollars of refinancing bonds guaranteed by the government as to both principal and interest. The funds of the corporations are used (1) to refund bonds which had already been issued by the federal land banks for the purpose of refinancing farm mortgages, thus exchanging new bonds fully guaranteed for the former bonds, which were guaranteed only as to interest, and (2) to refinance mortgages of additional mortgagors who might be induced to exchange mortgages for the newer type bonds.

Loans to home owners. The home loan mortgage business had never been adequately provided with funds or effective machinery for granting, servicing, or rediscounting loans. One significant effect of the failures in 1929 and after of the state and national commercial banks, as well as of the mortgage bond and mortgage guaranty companies, following as they did the excessive real estate developments between 1922 and 1929, was the complete prostration of the construction industry and the collapse of real estate values. It was generally felt that recovery from the depression was defeated, if not made impossible, unless this situation were relieved. Into this vent stepped the government, with a series of credit institutions:

(1) The twelve federal home loan banks, established in 1932, as reserve or discount institutions. These are owned by mortgage companies, the federal savings and loan associations, the state building and loan associations and other lending institutions, and the United States government. The basic function of these banks is to help the members finance their lending operations, especially in emergencies.

(2) The Home Owners' Loan Corporation. In order to save the owners of small homes from losing their equities in their homes, where that danger might exist, and in part to liquidate mortgages tied up in closed banks, a program of refinancing with government funds, similar to the program for farm mortgages, was applied to the field of urban home mortgages by the home owners' loan act of June, 1933. This was to be accomplished through the Home Owners' Loan Corporation, owned by the government and authorized to issue a total of $4\frac{3}{4}$ billion dollars of bonds with interest and principal guaranteed by the government and, during the emergency, to exchange them for distress mortgages, as well as to make supplemental cash loans when needed for such items as accrued taxes and repairs. The operations of the corporation were viewed as strictly emergency financing and the corporation discontinued the taking of new loans in 1936.

(3) The federal savings and loan associations — organized under and holding membership in the federal home loan bank system, and doing business similar to the building and loan associations. The associations are under federal charter and have capital subscribed in part by the government, but they are under local management. They are intended to fill a permanent place in areas not already adequately served by private mortgage companies.

(4) The Savings and Loan Insurance Corporation — created to insure the shares in the federal savings and loan associations and other members of the federal home loan banks.

(5) The Mutual Mortgage Insurance Fund, which under the Federal Housing Administration establishes a system of mutual mortgage insurance whereby certain mortgage investments, in the hands of approved mortgagees, may be granted a substantial insurance against loss through default of payments of interest or principal. These are long-term loans, commonly called "Title II loans." The national housing act of June 27, 1934, also provided a scheme of insurance

of credit employed in the renovation and modernization of homes. These loans are spoken of as "Title I loans"; they are of short term and designed to stimulate building and help end the depression. The legal provision for the insurance of such loans expired in 1937 for a time but was revived in January, 1938, because of the return of depression.

(6) The national housing act provided for the establishment of national mortgage associations, under federal charter, with power to buy and sell first mortgages and to borrow money for such purposes by the issuance of securities. They were designed to absorb the insured Title II mortgages in case the public should not readily buy them. One such institution, the Federal National Mortgage Association of Washington, D. C., was created by the RFC in 1938, and others may follow.

Criticism of government banking. To a certain extent specialized banking activities of the government may be justified as necessary salvage operations in an emergency characterized by general timidity and frequently paralyzing losses among private credit agencies. A great deal of hardship for the families of farmers and of urban home owners has been averted or ameliorated. At the same time it must be recognized that putting government capital into competition with private capital on an extended scale does not tend to restore the active confidence of the latter in the prospects for profitable operations. Officials of the "New Deal" have frequently said that the government wishes to get out of banking activities as soon as may be justified by indications that bankers are prepared to "resume their responsibility." But the history of the assumption of functions by government suggests the possibility that much of the operation of specialized banking services by the government of the United States may be destined to continue indefinitely, a prospect which is scarcely calculated to inspire confidence on the part of private lending agencies.

CHAPTER XXIV

Money and Prices

The value of money. As we have defined and used the terms "value" and "price," the value of anything is the quantity of some other thing that would be given in exchange for it, and the price of anything is the quantity of money that would be given in exchange for one unit of it. In practice it is customary to express values as well as prices in terms of money. But what then is the value of money itself?

The value of any commodity may be expressed in terms of any other commodity, but this is clumsy and confusing, and by common consent we have come to express all values in terms of money. In the same way the value of money might be expressed in terms of any other commodity. This would likewise be clumsy and unprofitable, but there is not here the same escape. We cannot obviously express the value of money in terms of itself. Moreover the concept of the value of money, to be useful, must not be its relation to some one commodity. What we are interested in is the relation of money to all other commodities and services, to economic goods in general. When we say money has depreciated in value, we mean that a unit of it will not buy so much; and we do not have in mind any particular commodity; we think of it as less valuable with respect to anything and everything that could be bought. So when money is of high value we mean that it will buy much, or conversely that much of other things must be offered to obtain any given quantity of money. This is evidently nothing else than what people mean by the "purchasing power" of money. We thus arrive at a satisfactory working definition of the value of money and one that agrees closely with the popular concept. *The value of money is the quantity of*

other things in general that will be exchanged for one unit of money; i.e., its purchasing power.[1]

Value of money not constant. The universal custom of measuring all values in money, of noting how the prices of things rise and fall expressed in money, is responsible for the common and persistent notion that the value of the monetary unit is a fixed thing and that all price changes are due to changes in the values of the priced commodities. A moment's consideration of the definition of money should convince anyone of the error of this notion. Value is always a relationship between two things. If one bushel of wheat exchanged for two bushels of corn yesterday and for three bushels of corn today, we may say that the value of a bushel of wheat has *risen*, from two bushels of corn to three bushels, assuming thus that the value of corn is a constant. But we could just as correctly assume the value of wheat constant and say that the value of a bushel of corn has *fallen*, from one-half bushel of wheat to one-third bushel. Without knowledge of the facts of the particular case, the change in relation may have been due to a change in either corn or wheat alone or to changes in both of them. Under a money economy value and price are relations between money and other things, and it is always possible that a change in the value of anything may be due to a change in the value of money.

Measuring the value of money. But how is the value of money, or its purchasing power, to be measured? The answer must be found in that which expresses at once the relation of money to all other commodities and services. This is price, not the price of any particular thing, but a composite price of things in general, or what we have called the "general

[1] It will be noted by the careful reader that, as we have defined the term, value may relate to any quantity of the good in question (not merely to a unit). When we speak of the value of anything, as of land, we must specify the quantity, as a ten-acre farm, or a city lot. When referring specifically to the value of a unit of anything, we generally use the term price. In referring to the value of a unit of money we cannot say price, since we have chosen to express prices in terms of money. Hence, unless otherwise stated, we shall use the phrase "value of money" to mean the value of one unit of money.

price level." It is change in the general price level which indicates change in the value of money. A change in a particular price, while other prices remain substantially the same, is presumably due to changes affecting that particular commodity. But when the general price level rises or falls, we may safely assume that the cause is a change in the value of money. High prices and cheap money mean the same thing. Low prices and dear money are synonymous. A change in the general price level means an opposite change in the value of money.

It follows that to measure and record the value of money we must have some way of expressing the general price level. For this purpose there have been developed *index numbers*, which are capable of reflecting with a high degree of accuracy changes in the general price level or the purchasing power of money. Before proceeding further in our study of the value of money and its relation to prices, we must acquire a working knowledge of the technique of index numbers.

Index numbers: Price relatives. We know that all prices do not move in unison. Some rise, while others fall, and still others remain stationary. A complete picture of all price changes in any given period would show a kaleidoscopic confusion of thousands of separate and apparently unrelated prices, rising or falling or standing still. Yet there is such a thing as a general price movement, just as there is a movement of a whole swarm of bees while the individual bees are dashing about in all directions in the utmost apparent confusion and without seeming relation to each other.

For a simple example, let us consider the average wholesale prices of three commodities, bituminous coal, wheat, and copper, in the three years 1913, 1914, and 1915. Here are the actual facts:

PRICES

	1913	*1914*	*1915*
Coal	$1.27 per ton	$1.17 per ton	$1.04 per ton
Wheat	.91 per bu.	1.04 per bu.	1.34 per bu.
Copper	.15 per lb.	.13 per lb.	.17 per lb.

Evidently the price of coal declined during this period, the price of wheat rose, and the price of copper fell from 1913 to 1914 and rose in the next year. To measure these separate price movements we may call one year, say 1913, the base year and express each price of each commodity as a percentage of the price of that commodity in the base year. For example the prices of coal were: in 1913, $1.27; in 1914, $1.17: in 1915, $1.04. Calculating the percentages we have; $\frac{1.27}{1.27} = 1$, or 100%; $\frac{1.17}{1.27} = .92$, or 92%; $\frac{1.04}{1.27} = .82$, or 82%. If then we let 100 stand for the price of coal in 1913, the price in 1914 relative to it will be 92, and the 1915 price, 82. Such numbers are called *price relatives*. By similar calculations we find the price relatives for wheat and copper with respect to their prices in 1913, and the whole result is shown in the following table:

PRICE RELATIVES

	1913	*1914*	*1915*
Coal	100	92	82
Wheat	100	114	147
Copper	100	87	113

Such a table might be continued to include similar price relatives for all other important commodities, giving thus a record of the price changes of each separate commodity. But what we now desire is to know how prices in general have changed; we are seeking a record of changes in the general price level. For this there is required a number which shall express for each year the general or average price relative, evidently the average of the separate price relatives for each year.

The average, simple and weighted. Now there are many kinds of averages, and the particular average chosen will determine the character of the index number. For example we might employ the simple arithmetic mean of the separate price relatives of each year. In our example the average for

1913 is of course 100. For 1914 the simple arithmetic average is $\frac{92 + 114 + 87}{3}$, or 98. For 1915, it is 114. Index numbers have been constructed on this principle, but they have a serious defect. The several commodities which enter into trade are not of equal importance, and an index number which takes no account of such differences is not an accurate record of the changes in the general price level or the purchasing power of money. The value of money is a good deal more affected, for instance, by a change in the price of wheat than by an equal change in the price of clover seed.

To avoid this defect of the simple arithmetic mean it is necessary to use a weighted average which takes proper account of the relative importance of the several commodities. For example suppose that in a given year the price of corn increased 4 per cent, while the price of butter increased 10 per cent, and that corn was twice as important as butter. The simple arithmetic average would show a combined increase of 7 per cent; $\frac{4 + 10}{2} = 7$. But this is not a true picture, since no account is taken of the greater importance of corn. The average is corrected by putting the figure for corn in twice; thus $\frac{4 + 4 + 10}{3} = 6$. This is a weighted average of the percentage increases. Weighting may be accomplished in various ways, the most satisfactory method being to weight the price movement of each commodity according to the total value of that commodity exchanged in a given year. Thus if twice as much money is expended for commodity A as for commodity B in the selected year, this is taken to indicate that A is twice as important as B and should be given double the weight of B in the index number. The simplest way to accomplish this is to calculate first the total value of all the goods exchanged in a given year and then to compare with

this the value of exactly the same quantities of these goods reckoned at the prices of each of the other years.

An index number of the aggregative type. We are now ready to apply this averaging technique to the prices in our previous example. Here are the facts for the first year:

YEAR 1913

	Prices in 1913	*Quantities exchanged in 1913*	*Values at 1913 prices*
Coal	$1.27	447 million tons	568 million dollars
Wheat	.91	555 million bu.	505 million dollars
Copper	.15	812 million lbs.	122 million dollars
Total value			1,195 million dollars

It cost $1,195,000,000 to buy these quantities of these commodities in 1913. If the same quantities of these commodities cost more or less in 1914, it must have been because of changes in their prices, and if we could determine what the cost was in 1914 we should have a very good measure of the average movement of these prices. This may be readily calculated, as follows:

YEAR 1914

	Prices in 1914	*Quantities exchanged in 1913*	*Values at 1914 prices*
Coal	$1.17	447 million tons	523 million dollars
Wheat	1.04	555 million bu.	577 million dollars
Copper	.13	812 million lbs.	106 million dollars
Total value			1,206 million dollars

Evidently commodities which cost $1,195,000,000 in 1913 were worth $1,206,000,000 in 1914. There must have been a small rise in the average price level of these three commodities.

For the year 1915 similar results are obtained, as follows:

YEAR 1915

	Prices in 1915	*Quantities exchanged in 1913*	*Values at 1915 prices*
Coal	$1.04	447 million tons	465 million dollars
Wheat	1.34	555 million bu.	744 million dollars
Copper	.17	812 million lbs.	138 million dollars
Total value			1,347 million dollars

Again there has evidently been a rise in the average price level.

The final step in deriving a real index number of the prices of these three commodities is to express the total value of each year as a percentage of the total value in 1913. Thus: $\frac{1,195}{1,195} = 1.00$, or 100%; $\frac{1,206}{1,195} = 1.01$, or 101%; $\frac{1,347}{1,195} = 1.13$, or 113%. These percentages are the index numbers, and the series may be thus expressed:

Year	*Index number of prices*
1913	100
1914	101
1915	113

The formal definition of an index number of prices is as follows: *An index number of prices is a number which expresses the general price level for any given year or other period relative to the general price level of some particular period which is taken as the base period and whose general price level is commonly represented by 100.*[1]

The particular type of index number which we have here developed represents what is known as "an index number of the *aggregative* type." The year 1913 is the base period, its index number being 100. The weighting is according to the values exchanged in 1913. Though the same in our example, the year chosen for the weighting need not necessarily be the base year.

Practical index numbers. The example used in the foregoing explanation was purposely made simple by employing only three commodities. A practical index number must take account of a large number of articles and may be calculated for any number of years, but no new principle is

[1] The term "general" in this definition does not imply that the prices of all commodities are included in any given index number. The meaning is that the index number represents the result of averaging price relatives for all commodities which have been considered. It is only in this sense that any index number shows a "general price level." Actually it is possible to construct index numbers which are capable of serving almost as well as though they were in fact based upon all commodities.

thereby involved. Also it should be noticed that other averages than the weighted arithmetic mean may be employed. Of the many possible forms of index numbers, the aggregative type both illustrates sufficiently the theory of index numbers and is probably the most useful of the types in actual practical use.

Two of the most useful series of index numbers and the ones most generally used in the United States are those published by the United States Bureau of Labor Statistics. The first of these was originally published in 1902, since

WHOLESALE PRICES AND VALUE OF MONEY IN THE UNITED STATES (1890–1939)

Year	*1913 base*	*1926 base*	
	Index Number	Index Number	Purchasing Power of the Dollar
1910	101	70.4	142.0
1911	93	64.9	154.1
1912	99	69.1	144.7
1913	100	69.8	143.3
1914	98	68.1	146.8
1915	101	69.5	143.9
1916	127	85.5	117.0
1917	177	117.5	85.1
1918	194	131.3	76.2
1919	206	138.6	72.2
1920	226	154.4	64.8
1921	147	97.6	102.5
1922	149	96.7	103.4
1923	154	100.6	99.4
1924	150	98.1	101.9
1925	159	103.5	96.6
1926	151	100.0	100.0
1927		95.4	104.8
1928		97.7	102.4
1929		95.3	104.9
1930		86.4	115.7
1931		73.0	137.0
1932		64.8	154.3
1933		65.9	151.7
1934		74.9	133.5
1935		80.0	123.5
1936		80.8	124.2
1937		86.3	115.8
1938		78.6	127.2
1939		77.1	129.7

which time it has undergone various changes from an unweighted mean of price relatives to its present form, which is an example of the aggregative type described on page 442. It shows the index number of wholesale prices for each year from 1890 to 1926 and for each month from January, 1913, through 1926. The number of commodities used has varied from time to time. For the first year, 1890, there were 192 commodities or separate series of quotations; in 1914 the number had increased to 297; in 1926 it was 404. The base year is 1913, and the weighting, utilizing the latest

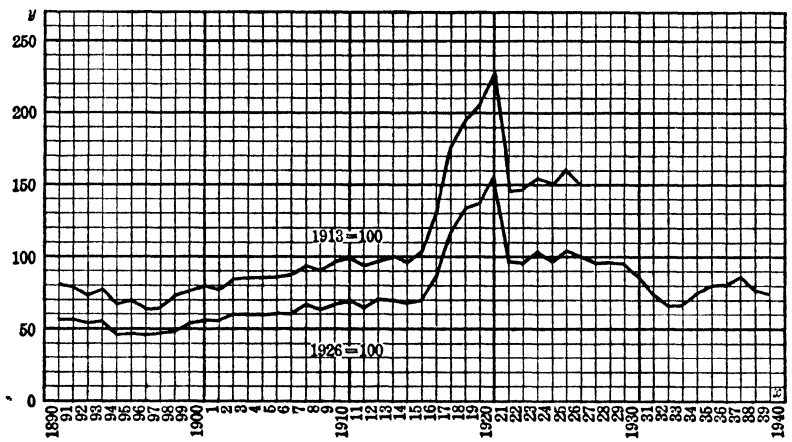


FIG. 24

Census reports of quantities exchanged, is made in accordance with the values exchanged in the year 1919. There is now also a new series, which uses 1926 as the base year and is weighted in accordance with the values exchanged in the three years, 1923, 1924, and 1925. This series employs 550 commodities or price series. Annual and monthly numbers are given. The table on page 444 gives these two annual series from 1910. The figures in the second column under the 1926 base are the reciprocals of the corresponding index numbers, thus showing the purchasing power or value of the dollar in each year relative to its value in 1926 (which is called 100).

For many purposes the most useful way to present an index number is by means of a graph or curve. The Bureau of Labor Statistics' annual index numbers of wholesale prices are thus shown in Figure 24 on page 445, the point for 1939 being based upon the average of the monthly index numbers of January through May.

The problem of the price level. We are now ready to inquire into the forces that determine the value of money and the general price level. These forces operate through three factors; namely, the volume of trade, the quantity of currency, and the velocity of circulation of currency. These factors are all subject to more or less continuous and independent changes.

Volume of trade. The chief service performed by money is the making of exchanges. From the analogy of demand and supply and the value of commodities in general, it should be evident that the value of money is dependent in part upon the volume of exchanges, since it is this that measures the need of money. The more money work there is to be done, as measured by the volume of exchanges, the greater is the need of money, and the greater will be the value of a unit of it; and *vice versa.* As our first step we conclude that the value of money tends to vary directly with the amount of trade.

Quantity of currency. It is by now a commonplace that the value of a unit of anything (*i.e.*, its price) tends generally to vary inversely with the quantity available for use. In a season when there is an unusually large wheat crop we expect, other things being equal, to see the price of wheat comparatively low. A crop failure generally leads to a high price of wheat. So it is with money; the more dollars there are to do a given amount of money work, the less is the value of one dollar; and *vice versa.* In other words, the value of a unit of money tends to vary inversely with its quantity.

All of the money work however is not performed by money, and in determining the value of money account must be taken of the other forms of currency. The only important one is

bank deposits subject to check. The quantity of bank deposits has the same relation to the value of money as the quantity of money itself. The conclusion of the previous paragraph should therefore be modified to read: the value of a unit of money tends to vary inversely with the quantity of currency.

However the quantity of money and the quantity of bank deposits are not wholly independent magnitudes. As we have learned, the amount of deposits which the banks have outstanding is governed in part by the amount of their reserves. When the ratio of reserves to deposits is high, the banks are likely to encourage loans and so increase their deposits. When the reserve ratio is low, loans are curtailed and deposits checked. There tends thus to be at least a rough relation between reserves and deposits. And since the division of the total stock of money between bank reserves and other uses depends upon habits of the people which have a certain degree of persistence, we can discern a relation between the total quantity of money and the deposits of the banks. Moreover causes which would lead to increases or decreases in bank deposits, such as business activity or depression, tend also to have the same effect upon the volume of money, principally through changes in the amount of bank notes, and this contributes also to maintaining the relation between money and deposits.

While not neglecting this relation, we must however avoid the opposite error of overlooking the possibility of considerable independent variations in the respective quantities of money and bank deposits. And we must therefore include both of these magnitudes in our analysis of the quantity of currency as a factor in determining the value of money.

Obviously the value of money can be affected only by such currency as is actually performing the monetary function of a medium of exchange. When, as sometimes occurs, the people are hoarding unusual quantities of money, such money is for the time being withdrawn from use and ceases to be a part of the quantity of currency that affects the value of money.

Of course there is always a considerable amount of money in pocketbooks, in the homes of the people, and elsewhere held for later use. The difference between quick spending and hoarding is after all one of degree. So long as the money so held is the normal amount as dictated by custom, such money should be considered as actually employed as part of the medium of exchange. Only an abnormal amount withdrawn for hoarding may properly be regarded as outside the sphere of influence upon the value of money. The variable influence of money normally withdrawn, in varying degree, from active circulation is properly taken into account in connection with the velocity of circulation of money, to which we shall give attention in the next section.

In like manner, the value of money is affected by bank deposits only so far as they are available for making current purchases. This is why savings bank deposits and time deposits generally are not included.

Velocity of circulation. Unlike most other commodities, money is not consumed when used. A coin or note, after facilitating one exchange, is ready with unimpaired efficiency to perform another exchange.[1] The efficiency of a given quantity of money to make exchanges is thus materially affected by its rapidity of turnover. One dollar used or "turned over" twelve times in a given period comes to the same thing as twelve dollars used once apiece. The number of times on the average that a unit of money changes hands during a certain period of time, say a year, is called the "velocity of circulation" of money. The amount of money work that can be done in a given time is the product of the quantity of money multiplied by its velocity of circulation, and changes in velocity have the same effect on the value of money as proportional changes in its quantity. This principle is of course equally true of other forms of currency, in particular of bank deposits. It has been estimated that money changes hands on the aver-

[1] The fact that paper money wears out is not of significance, since the worn-out notes are always automatically replaced by new ones.

age in the United States about 25 times a year. The velocity of circulation of bank deposits is estimated to be somewhere between 50 and 60 times a year.

The quantity theory of money: Algebraic relations. Summarizing these separate conclusions, we find that the value of money tends to vary inversely with the quantity of currency and its velocity of circulation and directly with the volume of trade. Since it is the general price level which measures the value of money, varying inversely with it, our conclusion may equally well be stated thus: *the general level of prices tends to vary directly with the quantity of currency and its velocity of circulation and inversely with the volume of trade.* This principle has come to be known as the "quantity theory of money."

The reader may possibly be helped to a clearer understanding of these relations by putting the several quantities into an algebraic equation. Let M stand for the average quantity of money in circulation during the year, and let M' denote the average amount of bank deposits subject to check and all other forms of currency. (The other forms are relatively insignificant and may be disregarded.) Let V and V' denote the velocity of circulation per year of money and of other forms of currency respectively. Let T stand for the volume of trade; that is, the total number of units of goods and services exchanged for currency during the year. Let P represent the general price level; that is, the average price paid for all the units of goods and services so exchanged. We then have the following equation:

$$MV + M'V' = PT$$

The quantity of money multiplied by its velocity of circulation is evidently the total amount of money paid in making exchanges during the year. Likewise the quantity of deposits multiplied by their velocity of circulation gives the total amounts of payments by transfer of deposits.[1] The sum of

[1] Strictly speaking, $M'V'$ includes also any other forms of currency that may have been used and which, for simplicity's sake and because of their practical insignificance in the problem, we shall hereafter ignore.

these two products is the total amount of payments, in all kinds of currency. The volume of trade multiplied by the average price is obviously the total value of all things exchanged for currency. This magnitude and the total amount of currency payments are necessarily the same, and the "equation of exchange," as it is called, is the algebraic expression of this relation.

It is P which measures the general level of prices and (inversely) the value of money, and the effect upon P of any of the other magnitudes is clearly shown by the equation. Changes in M or M', or V or V' tend to cause similar changes in P; changes in T tend to cause opposite changes in P.

Significance of the quantity theory. Few principles of economic science have been responsible for so much learned dispute and so much popular misunderstanding as the theory of just what causes determine the general level of prices and, more specifically, what the relation is between the quantity of money and the price level. The battle has raged most fiercely about the so-called quantity theory of money and the equation of exchange. While anything like exhaustive study of this problem lies necessarily outside the limits of this elementary book, we can with profit undertake some further analysis of the quantity theory and the equation of exchange, with the purpose of seeing clearly just what the principle means and some at least of the things that it does not mean.

The equation of exchange sets forth the necessary balance, over a year or some other limited period of time, between six specified factors. The quantity theory of money expresses the influence upon the price level of each of the other factors as a "tendency"; *i.e.*, as the change in P which would follow a given change in any one of the other factors, "other things being equal." But other things may not — very generally will not — be equal. For example, a decrease in the quantity of money (M) would decrease the magnitude on the left side of the equation; this would require an equal decrease on the right side, necessitating a decrease in the price level (P) — always

assuming no change in any other factor. But a decrease in M might be counteracted by an increase in M', necessitating no change in P, or even leading to an increase in P. Changes in M might likewise be counteracted by opposite changes in V or V'. Again a decrease in the quantity of money might fail to cause the price level to fall because of a simultaneous decline in the volume of trade (T), which would make the necessary adjustment on the other side of the equation. Many other combinations of changes are possible. Failure to understand this mutual balance of the factors involved leads to many erroneous assumptions, the most common and pernicious being the naïve notion that any change in the quantity of money necessarily causes a similar or even proportional change in the price level. While this is in truth the tendency, history records many instances where an increase in the quantity of money has not increased the price level or has not increased it proportionally, and *vice versa*.

Again for the sake of clear thinking, it must be noted that the several factors whose influence upon the price level concerns us are not entirely independent of each other. Changes in these factors may, in part, work themselves out on each other rather than on P. We have already noted the relation which may exist between money and bank deposits (M and M'); so far as this relation is concerned the consequence is to strengthen the influence of changes in M or M' on P. An opposite consequence follows from the fact that a sudden great increase in the quantity of money might have an effect on velocity of circulation, money passing on the average more slowly because of the greater quantity in people's possession. A sudden great decrease might have the opposite effect, causing the limited quantity of money to be turned over more rapidly. As regards the relation between M and M' and T, it is true that a sudden great increase in M or M' would very probably increase temporarily the purchases of the people and so increase T. A corresponding decrease in M or M' might decrease T. These relationships, tending to weaken the

effect of changes in M and M' upon P, would generally be temporary in their effects, ultimately yielding to the habits of the people and other more fundamental influences.

Even more important is the relation between the volume of trade and the quantity of currency. There can be no question that an increase in the quantity of goods exchanged is very likely to have a material effect in increasing the amount of money and still more the amount of deposits, while a decrease in trade has the reverse effects. As we have seen, an elastic currency responds to business needs, especially through changes in the amount of bank notes (which are money) and bank deposits. Changes in T thus may cause similar changes in M and M', and such changes will offset the tendency of the changes in T to cause opposite changes in P.

Finally we must note — although limitations of space prevent its demonstration — the fact that the equation of exchange is chiefly useful in portraying certain long-run relationships. It is less useful in explaining short-run changes. This is in part due to the fact that the equation combines all transactions in one total (T) and averages all prices so as to give one price level (P), whereas it is the various relationships between separate prices, rather than their average as shown by the general price level, which are of chief significance as causes of the short-run disturbances which are constantly appearing in the economic system.

As thus elaborated and qualified, the quantity theory of money is generally accepted as a true statement of the long-run relationships between the quantity of currency and the general price level.

Practical limitations of the quantity theory. The practical utility of the quantity theory of money, either as a means of predicting price changes or as a guide to official control of the price level, is by no means so great as is sometimes assumed. This is due both to the interdependence of the several factors in the equation of exchange, to which our attention has just been directed, and also to the practical difficulty of

discovering the exact magnitudes of the various factors. With the exception of the quantity of money and the general price level, these magnitudes must generally be supplied by means of estimates and calculations which are subject to a wide margin of error.

Far-reaching effects of monetary fluctuations. Fluctuations in the value of money are fraught with the most serious consequences to all the people, inflicting fortuitous loss upon some and bestowing undeserved gain upon others. Rights to future payments are normally expressed in terms of the country's monetary unit. When the time for payment comes the debtor will pay and the creditor will receive money whose value may be quite different from that which existed when the agreement or contract was made. A farmer borrowed $10,000 at a time when wheat was worth $1.00 a bushel. If before the time for payment arrives there has been such a fall in the value of money that prices, including the price of wheat, have generally doubled, the farmer will be able to pay his debt in dollars worth only half as much as those he borrowed. When he made the loan, it required the sale of 10,000 bushels of wheat to produce $10,000. Now he need sell only 5,000 bushels to obtain the $10,000 needed to pay his debt.

All investments which consist of rights to receive stated sums of money, such as bonds, mortgages, annuities, insurance, etc., are subject to the same principle. The investor is compelled in spite of himself to gamble on the value of the dollar. If it goes up, the dollars he receives will buy more; if it falls, his dollars will buy less. Such investors were fortunate gainers from this cause during the long period of falling prices that ended in 1897. After that the most conservative investors saw their resources dwindle from year to year as the index number of prices rose. On the other hand, a period of drastic price decline may injure investors through the failures and bankruptcies of many business concerns whose securities they hold.

Those whose incomes are in the form of salaries and wages

are affected by fluctuations in the value of money the same as are creditors. They have agreed to render services in return for a stated number of dollars, which as they are received may be worth more or less than when the agreements were made. Only as agreements expire and new ones are made can the parties undertake to make new terms conforming to the new monetary value, and these new agreements are equally subject to the vicissitudes of future changes.

It is thus that adjustment lags behind the fluctuations of the monetary standard on all agreements except those which are for immediate payment or for relatively short terms. Here there is the current adjustment, through raising and lowering prices, which produces those fluctuations in the general price level by which the community adjusts itself as best it may to the changing monetary standard.

Monetary reform. It will generally be taken for granted that stability of value is a desirable feature in any monetary system. Stability of value of money means a general price level correspondingly stable. Some would go so far as to hold up as the ideal a monetary unit whose purchasing power would be always the same. In the past those who have understood the economic principles of money and have sought the maintenance of the most stable standard possible have generally had their hands fully occupied in repelling the attacks of those who sought to displace the gold standard by something even less stable. More recently the widespread abandonment of the gold standard, and the present uncertain state of the monetary systems of the world have directed much thought to the fundamentals of monetary theory and the problems of monetary reform. Is a stable monetary standard possible? How may such a standard be obtained? The limits of this elementary book will not permit us to explore this problem much beyond the mere outlining of some of the more significant current suggestions for a stable monetary system.

There are first those plans which would adhere to the

gold standard and look to such regulation, official or otherwise, of the amount of credit money — generally bank notes — and the amount of bank deposits subject to check as would regulate the value of currency through control of its quantity. Such plans have undoubted theoretical merits, while at the same time involving many practical difficulties and facing serious political hazards.

Some years ago Irving Fisher proposed his now well-known "commodity dollar" or "stabilized dollar." Adhering to the gold standard, this plan would make the monetary unit — the dollar — not a certain weight of gold as now, but a variable weight of gold, so adjusted from time to time as to keep its purchasing power virtually constant. Gold coin would not be used in actual circulation. The government would be as now the custodian of the country's monetary gold. There would be in circulation gold certificates and other forms of money reading in terms of dollars (not necessarily much different from the present forms), all of which would be redeemed in gold by the government on demand. The amount of gold to be given for a dollar would be officially determined from time to time on the basis of the index number of prices, so that a gold dollar would always purchase about the same quantity of goods in general. Variations of individual prices with respect to each other, like the movements of the separate bees in a swarm, would in no wise be affected. But the movements of all prices in general would be stopped; the swarm as a whole would become stationary. This plan, especially if simultaneous adoption by most of the leading nations be assumed, might conceivably operate successfully under normal circumstances. Whether political obstacles would prove unsurmountable is a serious question.

There is reason to doubt whether either of the foregoing plans would be strong enough to withstand the forces tending to a drastic decline of prices in times of major crises. On the other hand, the operation of such a plan as the "stabilized dollar," though not perhaps able to cope with drastic short-

run price changes, might have considerable influence in smoothing the long-run disturbances which are known as "long waves." Elimination or moderation of these long-range fluctuations might have a salutary effect on the shorter fluctuations of the business cycle. In that case, the strain of the crisis itself would be easier to withstand.

Managed money. During recent years there has arisen a school of thought which holds that the ideal monetary system should not rest upon standard money at all, but should be a regulated irredeemable paper money system. The value of the monetary unit would be controlled by official action, for which the two most effective instruments are (1) control of the quantity of money and of bank deposits so as to maintain a uniform or otherwise controlled level of domestic prices, and (2) dealings in foreign exchange in order to maintain the value of the money in relation to the monetary units of other countries.

There has been of late some experience of this sort of monetary control, especially with reference to foreign exchange rates. Since their recent departure from the gold standard the leading nations have instituted measures to stabilize the values of their paper currencies in the international exchange market. For example the British government undertakes to keep the value of the pound in terms of dollars, francs, etc., fairly stable. This is accomplished by purchases and sales of gold in the open market. These operations require a fund set aside for this particular purpose. Great Britain in 1932 established the Exchange Equalisation Account, as its stabilization fund is technically known, for the purpose of stabilizing exchange rates and "managing the currency." The fund is controlled by a committee of the Treasury and the Bank of England. It is an immense fund, approximating 575 million pounds and consisting of gold, foreign exchange, and national currency, and is secretly managed. To take a concrete case, let us suppose that in England the value of the pound begins to decline in terms of the dollar. The British committee may

halt the decline if it is able to buy pounds and sell dollars; in other words, if it offers to redeem pounds with dollars at some price. But if it is to redeem pounds in terms of dollars, the equalization account must either have accounts in the United States with private banks or with the federal reserve banks from which it can draw dollars, or else it must possess sufficient gold so that it may secure dollars from the United States Treasury by selling gold to the Treasury.

Similar machinery and stabilization funds have been set up in the United States, France, Belgium, Holland, and Switzerland. The United States, as the reader will recall, devalued the dollar at the time of departure from the gold standard, thereby acquiring a "profit" of nearly 3 billion dollars, from which her stabilization fund of 2 billions was taken. Great Britain, on the other hand, raised her 575 million pounds by an issue of securities (*i.e.*, by a loan). These stabilization funds have become a fundamental feature of the currency systems of these countries, and certain authorities are disposed to favor such control as a permanent part of the financial machinery.

In recognition of such control as this, it has become the habit of late to apply the term "managed money" to the recently adopted irredeemable paper money systems. To avoid misleading inferences it should be remembered that not all irredeemable paper money systems have been characterized by any important degree of "management," while, on the other hand, there may be a good deal of "management" in connection with a gold standard system, through control of discount rates, "open-market operations," etc. There was for example far less management of the Continental bills of credit or the greenbacks than there was of the United States gold standard system in the twenty years before 1933. It is a fact however that the leading nations which went off the gold standard in connection with the crisis of 1929 and the following depression have exercised a considerable control over the international values of their paper money currencies.

In support of "managed money" there has been marshalled a considerable weight of theoretical argument based upon the quantity theory of money. There is theoretical argument also against the plan. In particular, the hazards of a monetary standard controlled by government and so always a potential football for political contest are not to be overlooked. The temptation to pay the costs of government by paper money issues and bank credit, in lieu of the less popular taxation, and the popular sentiment in favor of cheap money which is so prevalent in many countries furnish two powerful forces which as a matter of fact have eventually directed most irredeemable money systems down the path of uncontrolled inflation. Historically the world's experience with irredeemable money has been, almost without exception, disastrous. On the other hand, Great Britain, the United States, and a large number of other nations have now maintained irredeemable paper money systems, involving various degrees of "management," for about a decade at least without bringing disaster.

Before leaving this topic it should be noted that monetary theory is not wholly clear even as to what should be the aim of a managed money system. The idea of a constant price level is not so simple as may at first appear. There is reason to believe that in our existing economic order a stable price level might conceal the presence of factors making for eventual disequilibrium. Neither can it be said that the value of the monetary unit ought to be maintained at any specified level in terms of gold. Since Great Britain went off the gold standard in 1931, the gold value of the pound has in fact fluctuated. It is proposed by some that the managed money system be employed, not to preserve a stable price level, but to regulate the price level so as to bring about desired ends embraced in the government's program. This takes the problem into a realm in which economic science does not yet speak with sureness. For the further study of these important problems the reader is recommended to more advanced and special trea-

tises. Our investigation thus far would seem to justify the conclusion that, in spite of the recognized defects of the gold standard and the attractive features of some of the current monetary proposals, as a matter of practical policy the presumption is still strongly in favor of the traditional gold standard.

EXERCISES

1. Compute an index number of prices for each of the three years, on the 1926 base, using the simple average of price relatives:

	1926	*1932*	*1938*
Wheat (bu.)	1.20	.40	.60
Cotton (lb.)	.15	.06	.09
Coal (ton)	5.00	4.00	7.00
Iron (ton)	12.00	9.00	15.00
Petroleum (bbl.)	1.20	.60	1.00

2. What is the change in the value of money if prices rise 50 per cent? If prices rise 25 per cent? If prices fall 20 per cent? If prices fall 50 per cent?

3. In terms of the "equation of exchange" explain what effect each of the following would tend to have upon the general level of prices in the United States:

(*a*) The issue of a large amount of paper currency by the federal government.

(*b*) Increased productive efficiency of American industry.

(*c*) Open-market purchases of government securities by the federal reserve banks, increasing the size of commercial bank reserves.

(*d*) A decrease in borrowing by business men.

(*e*) Improved facilities for transfer and collection of checks.

(*f*) Expectation of rising prices.

(*g*) A law increasing the size of the reserve ratios which banks must maintain.

(*h*) Floods, droughts, strikes.

(*i*) Increased volume of loans by banks to business men.

(*j*) A sudden wave of thrift, or an increased desire to save.

4. (*a*) In the last stages of inflation in Germany after the World War, prices rose more rapidly than the increase in the quantity of money. Account for this situation.

(*b*) In the three year period, 1930–32, the quantity of hand-to-hand currency in circulation increased by more than 800 million dollars, yet prices declined severely. How, if at all, can this result be reconciled with the quantity theory of money?

CHAPTER XXV

Principles of Government Revenue

Government and the economic organization. No account of the economic organization would be complete without attention to the part played by government [1] in the production, distribution, and consumption of wealth. The existence of government is familiar to everyone, and virtually everyone recognizes its necessity. In fact we incline so much to take government for granted that it may not be amiss to lay the foundation for the study of this part of our subject by noting very briefly some of the more important economic aspects of government.

In modern times the greatest consumer of goods and services is the government. In 1938 the economic activity of the people of the United States brought forth a total income which is estimated as worth 62¼ billion dollars. The government (including all grades, federal, state, and local) collected and expended about 17 billions, being about 27 per cent of the total national income. In taxes alone there were collected about 14 billions. In the prosperous year 1929, when the national income was estimated at 79½ billion dollars, government collected and expended 11⅔ billions.[2]

On the side of production, the government is ever present. For example the United States Post Office furnished services in the fiscal year ending June 30, 1938, which cost over 776

[1] The term "government" is used in this book to include all grades of government of any particular community, for example in the United States of America the federal government, the state governments, and the governments of counties, towns or townships, cities, school districts, and all other political jurisdictions. It is rather common in popular speaking and writing to refer to the United States federal government as "the government" in distinction from the states and the local jurisdictions. The term is not so used in this book.

[2] United States Census, *Financial Statistics of State and Local Governments, 1932*, 1935; National Industrial Conference Board, *Economic Record*, August 31, 1939.

million dollars. In this same year the Tennessee Valley Authority cost 42 millions, and the Panama Canal, 11 millions; on public works the federal government spent the huge sum of 1,849 millions. But the chief contribution of government to production is indirect, consisting in such services as protection from outside foes, enforcement of law and order among the people, promotion of the public health, public education, etc.

In the field of distribution, the government is no less omnipresent and important. We need only to recall that in the year 1936 over one-fourth of the total income of the American people was collected and redistributed by government. In this year all the various governments in the United States (federal, state, and local) paid nearly 6 billion dollars in compensation to 3,625,000 employees, not including work relief employees.[1] These employees comprised almost 9 per cent of all persons employed in the country (not including unpaid family labor).

The functions of government. Lest they be overlooked in the present-day discussion of the more controversial governmental activities, a brief summary of the ordinary functions of government may not be out of order. The primary functions of government include (1) those of a defensive nature: defense against outside foes, requiring army, navy, etc.; justice and security, afforded by police service, courts, and penal institutions; regulation of private industry, in particular monopolistic industries like the railroads and those which are so complicated as to be only imperfectly understood by the public, such as the banks and insurance companies; protection against disease and accident; maintenance of moral standards; and protection against fire, storms at sea, and certain other forces of nature; and (2) developmental functions: education; religion; facilities for private industry,

[1] United States Bureau of Foreign and Domestic Commerce, *National Income, 1929–1936*, 1937. For further light on this topic, cf. William E. Mosher and Sophie Polah, "Public Employment in the United States," *National Municipal Review*, January, 1932, pp. 51–75.

especially highways and bridges, harbors, the monetary system, information services, etc.; and development and conservation of the nation's natural resources. In addition to the foregoing, which are peculiarly governmental functions in modern times, government may enter the industrial field, where private enterprise generally holds sway, and we have a considerable variety of government industries.

Government income. To meet the heavy expenditures necessitated by the performance of its functions, the modern government must have an equally large income. In moderr times there are just four important sources from which government income is to be obtained and which now demand our study. These are (1) industrial earnings (prices), (2) taxes, (3) fees and assessments, and (4) loans.[1]

Industrial earnings. When a government engages in a business enterprise, such as the United States Post Office or the water supply system of a city, it sells its product (goods or service) at a definite price to the consuming public much as any private business concern does. The citizen may take the offered service or not as he wishes, and if he does not take it he makes no payment. Payment is voluntary and in proportion to the amount of the service received. For example the government offers, through the post office, to transmit letters and parcels. If the citizen chooses to take advantage of this service he pays the stated price; otherwise he pays nothing to the post office. In these respects government industrv is similar to private business.

Financing the primary functions of government. If now we transfer our attention from its industrial undertakings to the primary functions of government, we shall find certain fundamental differences in their financing. The primary functions, such as defense, justice and security, protection against disease, education, etc., are undertaken by the government in order to benefit the people as a whole, without the purpose

[1] Only the first three represent revenue in the strict sense. Government loans are discussed in Chapter XXVII.

of giving particular benefit to particular individuals. For example the state and local governments of the United States provide public schools and universities, at great expense, not primarily in order to benefit those persons who want an education for themselves or their children, but because it is believed that a prosperous and enduring democracy is not possible unless there is widespread education among the people. Education is provided therefore to benefit the whole people, not for the sake of those particular individuals who attend the public schools.

In all these cases therefore no attempt is made to pay the cost of the service by charging a price and so collecting from each citizen in proportion to the benefit he has received. The government does not sell education as it sells postal service; education is offered freely to all. The burden of furnishing the money necessary to pay the cost of public education, as of the other primary functions of government, is distributed among the people according to certain rules, which we shall presently study, but there is no necessary relation between the amount of any person's contribution and the service received by him, nor may any person avoid contributing by declining to avail himself of the government's service. Contribution toward the cost of the primary functions of government is compulsory.

It follows that no attempt is made in the case of the primary functions of government to obtain in connection with each function a revenue earmarked to meet the expenditures of that particular function. As a general rule, little or no revenue comes in to the government in connection with the performance of any particular function, and the bulk of the revenue is received without any reference to the particular functions of government for which it is to provide.

Taxes. The foregoing analysis of the purpose of the government's primary functions and the financial contrast between such functions and the government's industrial undertakings has introduced the principal characteristics of taxa-

tion; namely, (1) compulsion, (2) devotion to the general welfare, and (3) apportionment without reference to individual benefit. It will be found helpful to have a precise definition of a tax, for which the following will satisfactorily serve our purpose: *A tax is a compulsory payment by a person to the government, destined to defray the cost of government services performed for the common benefit, and paid without reference to individual benefit.* It is upon taxes that modern governments principally rely for obtaining the money necessary to defray the cost of performing their primary functions, and taxes make up the bulk of the income of most modern governments.

Fees. The primary functions of government are performed, as has been shown, for the common benefit. Nevertheless it often happens that, in performing its primary functions, the government incidentally and unavoidably gives special benefits to particular individuals. Consider for example the granting of patents for inventions. The purpose of this service of government is not to give a special benefit to the inventor; if such were its purpose it would be a case of favoritism, repugnant to the spirit at least of democratic government. The real purpose of the patent is to promote the public welfare by encouraging the discovery and disclosure of useful new devices and by offering inducement to capitalists and entrepreneurs to develop and market them. It is for this reason that the inventor is given for a time a privileged position of monopoly with respect to his discovery. Nevertheless this is a material benefit conferred upon the inventor, and for that reason the government imposes upon him a special charge, justified by and — somewhat roughly it must be confessed — related to the particular benefit which he is presumed to have received. Such a payment is called a fee.

Fees arise also from somewhat different circumstances. There are certain occupations which cannot be left to the free and unrestricted operation of all persons without danger to the public interests. Such are the businesses of banking, insurance, taxicab operation, peddling, etc. Now the per-

formance of this regulatory function necessitates government expenditure, and the person who chooses to enter such an occupation is thereby causing expense to the government. The government therefore charges such person a fee, covering part or all of the cost imposed upon it by his engaging in the regulated occupation. Fees are thus imposed upon those who engage in any activity over which some government regulation has been found necessary. One cannot drive an automobile, or own a dog, or enter the marriage state, without a government license, and in each case he is charged a fee to cover the cost which his act has imposed upon the government.

We now have in mind the nature of the fee, which may be defined in precise terms as follows: *A fee is a payment by a person to the government on account either of a special benefit received from the government or a special cost imposed upon the government in connection with a government service performed for the common benefit.*

Special assessments. Suppose that the city authorities decide to pave a certain street. There will generally result a material advantage to the owners of land fronting that street on account of the rise in the value of such land. The same rise in land values follows the introduction of a sewer system, street curbs, sidewalks, and other improvements. The city government undertakes a new street pavement, not from any desire to benefit particularly the owners of the land, but because the interests of all the people of the city will be promoted by a better surface on that street. Here then is a perfect occasion for a fee and, in America at least, a fee is usually charged, which is none the less a fee in spite of its common designation as a *special assessment.* This is the definition: *A special assessment is a particular kind of fee, imposed when the special benefit consists in an enhancement of the value of land.* The amount of the special assessment may be anything up to the total addition to the value of the land caused by the government improvement. The special assessment furnishes a most interesting example of the fee, since there is here usually an attempt to relate the amount of the payment to the value of the special

benefit conferred much more closely than in the case of fees in general.

Names may be misleading. In distinguishing prices, fees, and taxes, we must have regard to their real nature and not be misled by mere names. For example a fee is never greater than the value of the special benefit received or the special government cost occasioned by the fee payer. If this limit is exceeded the payment is, at least to the extent of the excess, a tax. This is usually the case with the automobile operator's "license fee" collected by the states of the United States. On the other hand, consider the tax upon gasoline, which has come into recent prominence in the United States. This tax is imposed by the state at so many cents per gallon, and the proceeds are devoted chiefly to the maintenance and repair of the highways. When the facts are thus, this so-called "tax" is really of the nature of a fee. It is a payment which may be justified either on account of the benefit received from the public highways or the cost imposed by wear and tear upon the highways, and it apportions with reasonable accuracy the size of the payment to the amount of such benefit or cost.

Distribution of the tax burden: Equal distribution. Almost the entire revenue required to meet the cost of government's primary functions comes in the form of taxes, which are exacted from the people with little or no regard to particular benefits received. Great importance therefore attaches to the distribution of the tax burden; that is, the determination of how much each person must pay for support of the general services of government.

It may perhaps serve to clarify the first steps of the investigation of this problem if we stop for a moment to consider the possibility of equal distribution. This will be recognized as the rule which usually determines the contributions of the members for support of the general services of an ordinary club. Let us see how it would work in the United States. The total tax bill of the people of the United States in the

year 1936 was about 10½ billion dollars. With a population of about 128 millions, the share of each person — man, woman, and child — would be about $82. At this rate a family of five (father, mother, and three children) would have an annual tax bill of $410. If the man's wages or salary and other income were $1,500, his tax bill would be more than a quarter of his income. Another family of five with an income of $10,000, in paying the same amount, would pay only about one twenty-fifth of their income, while the very wealthy would find their tax burdens quite insignificant. To the very poor the burden would be absolutely crushing, if indeed they were able to pay at all.

Now quite apart from any question of justice, it is evident that this rule would be unworkable, since there are large numbers of the people who could not possibly pay at the rate of $80 a year for each taxpayer and all those dependent upon him. And whatever one's standard of justice may be, it is safe to say that virtually everyone would condemn this plan as flagrantly unjust. The important conclusion is that public opinion tolerates — even demands — a plan of tax distribution which makes some persons pay more than others. Having cut loose from the base of equal distribution, what shall be our rule of apportionment?

Benefit or cost. There have been those who believed that the costs of government should and could be distributed on the basis of the benefits received by the individual members of the group, or on the somewhat similar basis of the cost to the government of the service rendered to each person. These are in general the rules that determine prices charged in business dealings, and we have seen that they apply also to industrial enterprises conducted by government. But when we come to the general services of government, performed for the common good rather than for the particular benefit of individuals, there is generally no way of measuring either the benefit or the cost to each person. And even if some measurable relationships were ascertainable, the government could

collect only from those who were able to pay, though its general services must be rendered for all without discrimination.

It should be clear that neither benefit nor cost can serve as a general rule for the apportionment of taxes. At the same time these principles need not be discarded entirely. Just as they serve well as the bases for fees and assessments, so they have a certain bearing at least upon the obligation to pay taxes, if not upon the amount of the individual contribution. This will come up again in our study of particular taxes.

Ability to pay: Meaning. The almost instinctive rejection of the plan of equal distribution of the tax burden is based upon the common feeling that those who have much should pay more than those who have little, that there should be at least some approach to equality of sacrifice, that people should pay taxes in proportion to their ability to pay. There is no doubt a lack of precision in these notions, but they are widely held and with sufficient clearness to indicate another rule for tax apportionment; *i.e.*, the principle of ability to pay, which may be paraphrased accurately enough as the principle of equal sacrifice.

It is not our purpose to undertake a philosophical analysis of the principle of payment according to ability; its broad meaning will be clear enough for our present purpose. As so understood, the ability principle has on the whole commended itself to modern students of taxation and to modern public opinion as the basic principle of tax apportionment, to be supplemented by the principles of benefit and cost so far as they may be applicable and to be modified on occasion by other principles of a non-fiscal character. It must never be forgotten however that the complex practical problems of taxation are not to be solved by the mechanical application of any set of general principles. Modifications of general rules and resort to other principles will be found necessary to meet the exigencies of particular kinds of taxes.

Measurement of ability to pay. The collection of taxes being a practical financial operation, there is necessity of a

financial yardstick with which to measure taxpaying ability. For all practical purposes, just three measures are available: (1) the possession of wealth (or property), (2) the receipt of income, (3) consumption, or expenditure. We may fairly assume that to the student who has reached this point in the study of economics the appropriateness of these measures will be evident.

In the application of the property and income yardsticks to the measurement of taxable ability, it is generally the net amount, rather than the gross, that more accurately measures taxpaying ability. A and B may each own wealth of a value of $50,000. If A has debts of $30,000, representing property rights of others against him and his wealth, while B has no such claims against him, it is evident that the taxpaying abilities of the two, so far as indicated by wealth or property, are not the same, and that A's ability is to be measured by his net possessions of $20,000. Speaking strictly and as we have previously defined the terms, it is not wealth, but net property, that is the real measure. Net property is the difference between all the taxpayer's possessions (including property rights against the wealth and persons of others) and the property rights of others against him and his wealth. Likewise in using the income measure, it is net income; *i.e.*, the difference between gross income and the expenses necessary to secure such income, that is the real measure of taxable ability.

Proportional taxation. There arises at once the question whether the relation between wealth, income, or consumption and taxable ability is one of simple proportion or is something different. Suppose that A has a net income valued at $2,000, while B's is worth $4,000, and C's, $40,000. Does this indicate that B should pay twice as much as A, and that C should pay ten times as much as B and twenty times as much as A? If so we have proportional taxation, to be accomplished by a uniform rate. Suppose the rate is 4 per cent. Then we have this result, in which the tax payments are in direct proportion to the respective net incomes:

Person	*Net income*	*Tax rate*	*Amount of tax*
A	$ 2,000	4%	$ 80
B	4,000	4%	160
C	40,000	4%	1,600

Progressive taxation. If instead of making the tax payments directly proportional to the base, they are made to increase at a rate faster than the increase in the base, we have progressive taxation. For example suppose a scheme of income tax rates such as this:

On the first $1,000 of net income		1%
On the part of the net income:		
over $ 1,000, up to and including	$ 2,000	2%
over 2,000, up to and including	3,000	3%
over 3,000, up to and including	5,000	4%
over 5,000, up to and including	10,000	5%
over 10,000, up to and including	20,000	6%
over 20,000, up to and including	40,000	7%
over 40,000, up to and including	60,000	8%
over 60,000, up to and including	80,000	9%
over 80,000, up to and including	100,000	10%
over 100,000		11%

Now A, with a net income of $2,000, will pay a tax of 1 per cent of his first $1,000 plus 2 per cent of his second $1,000; *i.e.*, $10 + $20 = $30. Similarly B's tax would be calculated thus:

1% of the first	$1,000,	$ 10
2% of the second	1,000,	20
3% of the third	1,000,	30
4% of the fourth	1,000,	40
Total tax on	$4,000,	$100

C's tax would amount to $2,390. These results may be tabulated thus:

Person	*Net income*	*Tax rate*	*Amount of tax*	*Ratio of tax to net income*
A	$ 2,000	1%–2%	$ 30	1.5%
B	4,000	1%–4%	100	2.5%
C	40,000	1%–7%	2,390	5.975%

Under such a set of rates, B's tax would be more than three times that of A, though his income is only double A's. C's income is twenty times A's, but his tax is practically eighty

times as great. This is the result of progressive taxation and is accomplished, be it noted, by giving up the uniform rate in favor of a rate which increases with increases in the base.

It is probably not possible to prove scientifically that progressive taxation is required by the principle of ability. For that we should need a calculus for the measurement of sacrifice which thus far does not seem to have been discovered. However there is little doubt that the consensus of modern opinion, expert and general, inclines to the support of progression.

If there is no scientific proof of the correctness of progressive taxation, still less can there be any scientific determination of the exact scale of progressive rates required to apply the ability principle. In actual legislation the rates of progressive taxes are determined principally by rule of thumb, guided by various principles and interests, harmonious or conflicting.

Regressive taxation. It is possible to apply a principle opposite to progression, by fixing rates which decrease as the tax base increases. This is called regressive taxation. There are few if any who defend this rule as an accurate application of the ability principle, and it never appears consciously embodied in modern tax legislation.

The problem of taxing according to ability. There is no one perfect yardstick for the measurement of taxpaying ability. A general property tax which reached the entire net property rights of every natural person would be an accurate measure so far as it went. It would be defective in that it would fail to take account of the ability obviously conferred by the receipt of income from other sources than property, such as salaries, professional fees, and wages.

On the other hand, the commonly held notion that income is the one perfect measure of taxable ability can easily be shown to be erroneous. Suppose A and B each to be possessed of $1,000,000; A invests in corporate securities yielding an annual income of $40,000; B invests three-fourths of his fortune in idle real estate and derives only an income of $10,000

from the remaining $250,000 invested in corporate securities. The income tax alone will obviously fail to reach the taxpaying ability of B. And it is no sufficient answer to point out that if B's real estate venture proves successful, he will eventually have an increased income, upon which he will be taxed. The venture may prove unsuccessful. And in any case B is not today contributing to the present cost of government according to his ability. Each of these men was free to invest his fortune as he chose. Assuming neither to have any other source of income, their taxpaying abilities before they made their respective arrangements were equal. In this case the property tax would come closer to apportioning taxation according to ability to pay.

Consumption taxes likewise furnish only an imperfect measure of taxpaying ability. A general consumption or sales tax obviously takes more from the rich than from the poor and so recognizes the principle of ability. But such a tax alone is a very imperfect device for reaching taxpaying ability. It fails to reach the ability represented by property that is not yielding a present income (as in the case of B above), and even when there is income, it fails to take account of that part of the income that is saved. Suppose C and D are neighbors with about the same needs and the same scale of expenditures, each spending around $3,000 a year, although C's income is $6,000 while D's is $3,000. C's taxable ability is double that of D — if not more — and is clearly not reached to its full extent by a tax on consumption or spending, which takes only the same amount from him as from D.[1] The relation between tax payment and ability is still more remote when the consumption tax is, as is usual, levied upon a selected list of commodities or services instead of upon all expenditures.

[1] It may be noted that, if income be defined strictly as the actual benefits or services rendered by wealth or free persons (cf. Chapter I), savings are not a part of net income, and the income tax also would fail to reach the full taxpaying ability of C.

The three measures of taxpaying ability differ as to their adaptability to progression. Consumption taxes are seldom imposed directly upon the consumer; they are usually levied against manufacturers and dealers, through whose hands the goods pass. It is impossible in such case to use progressive rates in relation to the total consumption of the individual consumer. Even where the tax is imposed directly upon the purchaser, as in some of the recent state sales taxes, it would be impracticable to use progressive rates. The property tax might conceivably be progressive. But there are serious difficulties, both theoretical and practical, and the property tax is virtually always a proportional tax. It is only the income tax that lends itself readily to progression.

Another concept of progression. In discussions of the problem of tax apportionment, use is often made of somewhat different concepts of progressive and regressive taxation. According to such concepts, taxation is considered progressive when the amount paid by the individual increases at a rate faster than the increase in some assumed general or composite measure of economic power; roughly "the tax bears more heavily upon the rich than upon the poor." In this sense the progressive income tax (even the proportional income tax with its ordinary exemptions, credits, etc.) is progressive. So to a degree is the property tax. In the first instance this tax generally falls lightly or not at all on the poor. However as will appear more clearly after our study of shifting and incidence in the next chapter, a part of the ultimate burden of the property tax is passed on to many who do not pay the tax directly. Consumption taxes, on the other hand, are considered regressive (even though they are imposed at uniform rates), because, in the first place, they are usually levied upon articles of wide general consumption which call for a greater share of the poor man's expenditure than of the rich man's, and secondly they impose no burden on that part of income which is saved, a feature of which the poor can take only slight advantage. This concept of progression, though lacking

in precision, is useful when passing judgment upon a tax system.

Benefit, cost, and ability. General and special business taxes, measured more or less accurately by the amount of business done or the amount of business profits received, are quite common. Such taxes are well known in Europe. In America they are found in the business license taxes of many of the states, particularly in the South, in the liquor licenses commonly imposed by towns and cities, and in most of the special taxes on corporations. When imposed in addition to, not in lieu of, the regular taxes upon either the property or the income of the business enterprise or upon the owner on account of the property employed in the business or the income derived from it, business taxes can with difficulty be related to the principle of ability, but must generally be frankly acknowledged as applications of the benefit or the cost principle. The assumption is that government gives special protection to property employed in business and special services to those persons, natural or corporate, so engaged, or that the cost of government is especially increased on account of business activities of the people.

The poll tax, being usually a uniform contribution from all individuals, has no relation whatever to ability, but is related to the benefit principle, on the somewhat dubious theory that protection of life is granted to all and is of equal value to all.

The fact that business taxes, poll taxes, and possibly certain other forms of taxation are not in full harmony with the principle of ability is not to be taken as an *a priori* condemnation of such taxes, but is rather an indication of the truth of our earlier statement that no one principle can be relied upon to shape all the taxes which any government finds it necessary to impose. It must further be remembered that we cannot pick out a single tax from a tax system and criticize it as though it stood alone. Final judgment must be upon the tax system as a whole. Thus it is possible to combine certain taxes measured according to benefit or cost with other taxes

measured according to ability, as well as to combine proportional consumption taxes with progressive income taxes, inheritance taxes, etc., in such manner that the whole tax system may have reasonable conformity to the principle of ability.

Escape from taxation: Lawful adjustments. No one likes to pay taxes. It is only human nature that every taxpayer should use all lawful means to make his taxes as light as possible and that some taxpayers should resort to measures that are not lawful. Certain modern taxes, such as the property tax, the income tax, and the inheritance tax, are so involved with complexities that it is often difficult for the taxpayer to know just what his legal obligation is. For example there can be no doubt that, under the United States income tax, many persons are paying more than their lawful taxes, through misunderstanding of the law or mere carelessness. Under such circumstances the use by some taxpayers at least of every lawful means, including the employment of tax experts and lawyers, in order to avoid paying more than is legally due, is inevitable. Again a person may escape a heavy tax upon his property by selling his real estate and moving to another town where the tax rate is lower, or a man may avoid an inheritance tax by giving his property to his intended heirs during his lifetime. Finally it is often possible to shift the burden of a tax to other persons through increasing the price charged for some commodity or service. Put in general terms, it may be possible to escape a tax burden that might otherwise have to be borne (1) by exercising care through employment of lawyers, accountants, etc., not to pay more than is legally due, (2) by putting one's self in such position that the tax law is not applicable, or (3) by using the power which a tax may give to raise the price of something sold.

Now so far as all of these expedients are legal, it can raise no ethical controversy to assert that they are entirely justifiable. For the first the legislature itself is principally responsible, through enacting tax laws too complicated or obscure for the understanding of the ordinary individual. In

the second case, if it appears that taxpayers are finding it easy so to adjust their residences or businesses or possessions or other circumstances as to put themselves outside the application of a tax, thus defeating the purpose of the law, that again must be due to defects in the tax itself, for which the law-making body alone is responsible. The case of shifting the burden to others may indicate either failure of the law to place the final burden upon those whom the legislature intended to reach or simply the result foreseen and intended by the legislature. If the first case, the miscarriage is not to be ascribed to any dereliction, legal or moral, on the part of the taxpayer, but simply to lack of understanding of economic principles on the part of the lawmakers. As to the second, we have the simple and common case of indirect taxation, which will be examined presently, but which requires no further comment in this connection.

Tax evasion. In popular discussion there often appears a certain failure to distinguish between such legal and proper escape from the burdens of taxation and the evasion of taxes legally due. However much this popular confusion may be excused by the obscurity of tax laws and the complexities of the whole modern tax situation, the student must always keep the distinction in mind. Escape from taxation may be lawfully accomplished in a variety of ways; the term "evasion" had better be reserved for violations of the tax laws.

When tax evasion occurs in sporadic instances, the situation is similar to any other case of violation of law and is of no special concern to the economist or the tax student. When, as is sometimes the case, there is wholesale evasion, the responsibility must be either upon the legislature for enacting a law incapable of enforcement or upon the administrative personnel. To ascribe it to a low state of public morality does not face the problem. The legislature has to deal with human nature as it is and is responsible for enacting laws which can be enforced and for providing the necessary machinery for efficient administration.

The shifting of taxes: Meaning. Taxes are *shifted* when the person upon whom the tax has been imposed finds that the presence of the tax has given him the power, which he did not before possess, to raise the price of some commodity or service which he has for sale, thereby reimbursing himself in whole or in part for his tax payment and to that extent passing the burden on to someone else. An example familiar to everyone is the United States internal revenue tax upon the manufacturers of tobacco products. A comparatively heavy tax is thus imposed, but the manufacturers, who are the immediate taxpayers, regularly pass the burden of the tax on to the consumers of tobacco through an increase of the price. The degree of shifting which is possible varies, according to the nature of the tax, all the way from those whose burden is entirely or almost entirely shifted, such as the tobacco tax just cited, to those whose shifting is impossible, such as a tax upon monopoly profits or upon economic rent. Shifting and *incidence* (meaning the final resting place of the tax burden) must never be overlooked when considering the problem of the distribution of the tax burden or when criticizing any particular tax. Obviously the person who finally bears the burden is more to be considered than he who pays only to shift the burden to others.

Direct and indirect taxes. Taxes are usually put into two classes according as the burden is or is not normally shifted by the one from whom the government collects the tax. Direct taxes are those whose burden remains on the whole upon those from whom the tax is originally collected. Indirect taxes are those whose burden is normally shifted. Since the possibility of shifting may be present only partially and in various degrees, it is not always possible to draw a sharp exclusive line between the two classes of taxes. However the characteristics of the more important kinds of taxes are so clearly marked that there is usually little difficulty in placing any given tax in its proper class. This distinction, be it noted, is an economic one. It does not agree precisely

with the legal definition of a direct tax, as laid down by the Supreme Court of the United States. There is no occasion here for dispute. We have simply to deal with two different concepts, the economic and the legal, and must be on our guard only against the danger of confusion from the use of the same technical terms in two different senses.

The purpose of taxation. Of course the primary purpose of taxation is the raising of money whereby the government may pay the cost of performing its primary functions. While the discussion thus far has ignored any other purpose, it is essential to recognize that taxation may be imposed for other purposes and to examine the economic bearing of the more important of these purposes. The taxing power is among the most powerful and far-reaching of the attributes of sovereignty. Even when applied only for the purpose of securing government income, its indirect economic effects may be — indeed certainly will be — very great. When consciously used for the accomplishment of other ends its power can scarcely be exaggerated. Taxation may be used to impoverish one individual and enrich another. It may foster or kill whole industries and destroy the prosperity of entire regions. It may divert the course of industry and trade, affect the movement and location of population, interfere with the distribution and consumption of wealth. With such a power in its hands, it is not surprising to find the government occasionally using it for the accomplishment of ulterior ends, nor is it strange that numerous programs of social reform through taxation are constantly being urged.

The most familiar illustration is the protective tariff. The tariff, or customs, is a tax, imposed originally and primarily for the sake of revenue. At various times and places it has been used for the purpose of checking the import of certain foreign products in order to relieve home producers of foreign competition and so to build up the home industries. This purpose has been so very prominent in the United States during most of its history that there may be a question whether

protection is not really the primary purpose of the American tariff. Another example is the ten per cent tax imposed by the United States upon the notes of state banks. This tax was introduced with the sole purpose of legislating the state bank circulation out of existence. Revenue was not expected and has not appeared, since the tax has been successful in accomplishing its real purpose. Heavy license taxes imposed upon dealers selling alcoholic liquors are partly for the sake of revenue but also for the purpose of restricting and regulating the retail liquor establishments. Among suggestions for the use of taxation for "reform" are the proposal to limit or abolish roadside billboards by heavy taxes, the demand that inheritance taxes be used to produce a more equal distribution of wealth, and the time-honored "single tax" of Henry George.

In general it may be said that the use of a tax for another purpose tends to impair its effectiveness as a revenue machine. This is obviously true to the limit of such a tax as the prohibitory levy upon state bank notes, which produces no revenue at all, and it is especially applicable to the protective tariff. This subject will be pursued further in connection with our study of particular taxes, especially the tariff.

CHAPTER XXVI

Taxation

Among the various taxes employed by modern governments the following are the principal forms: property tax, income tax, poll tax, excise or sales tax, customs tariff, death tax. It is the purpose of the present chapter to sketch, necessarily in brief and elementary terms, the essential features of these several types of taxation.

The property tax: Evolution. The property tax is among the oldest of taxes; history shows us numerous examples in ancient and medieval times. At a time when there was not the bewildering variety of wealth and property that exists today, the early property taxes were based chiefly upon land and buildings. Later attempts to include various other forms of wealth and property rights as they gradually evolved did not prove successful, and the property tax in most European countries long ago returned essentially to its original base — real estate.

In America, on the contrary, where the property tax has since colonial days been the chief mainstay of state and local finance, the attempt is still generally made to include in its base most or at least very many of the complex modern forms of wealth and property rights. When thus inclusive, the tax is known as the *general property tax.*

The American property tax: General description. The property tax is prescribed by state law in each of the forty-eight states and the District of Columbia. Although no two states have exactly the same system, there is still enough of similarity to permit of a composite picture of the typical structure and operation of the property tax.

The first step is the *assessment; i.e.*, the discovery and valuation of all the taxable property in the jurisdiction of the

taxing government as of a certain date each year. This function is performed generally by officers of the counties or towns, acting under the state law. Assessment includes opportunity for appeal and revision in case of taxpayers dissatisfied with the original assessment. Thereafter there is a process of *equalization*, performed by a county board where assessment is by the towns and in any case by a state board, the purpose being to insure that the assessment in the several towns and counties has been uniformly done; *i.e.*, at full value or at the same percentage of full value. The assessment, as thus equalized, furnishes the official tax list or "grand list."

In most states revenue is obtained from the property tax by the state, the counties, the towns or townships, and the school districts, and often by various other local taxing districts. Each such political body determines the sum which it will require from the property tax. By the process of dividing this sum by the total value of all taxable property within its own jurisdiction, it determines its "tax rate." This process is called *apportionment* of the tax. The tax due from any particular taxpayer is determined by multiplying the assessed value of his property by a tax rate which is the sum of the tax rates of all the taxing jurisdictions to which his property is subject.

Position of the property tax. The property tax is the main source of revenue of the states and local bodies of the United

Jurisdiction	*Total revenue receipts*	*Property tax receipts*	
		Amount	Per Cent of Total
State governments	$2,208 mil.	$ 323 mil.	15
Counties	1,314 mil.	877 mil.	67
Cities, towns, villages, and boroughs	3,374 mil.	2,057 mil.	61
School districts	1,515 mil.	1,119 mil.	74
Townships	248 mil.	195 mil.	79
Other civil divisions	194 mil.	112 mil.	58
Total	$8,852 mil.	$4,685 mil.	53

States, as is clearly shown in the table on page 481, relating to the year 1932.

The year 1932 is the latest for which detailed figures such as the above are available. According to estimates of the United States Treasury Department, issued in August, 1939, the property tax in 1938 produced 4,745 million dollars for all the state and local governments. This was equal to four-fifths of the total tax collections of the federal government and accounted for one-third of the entire national tax bill (for all grades of government).

Assessment: The problem. The principal problems of the property tax centre about the assessment. The assessor is required (1) to discover all taxable property and (2) to determine its value. The first process, easy for real estate (land and buildings), heavy machinery, furniture, stock in trade, and similar forms of wealth, becomes difficult in case of such tangible personal property as watches and jewelry. Even such wealth as household furniture, books, pictures, clothing, etc., can be discovered only by a personal search of the taxpayer's home, a duty to which the ordinary assessor is loath to resort in the face of the natural repugnance of the taxpayer. When we come to intangible property — stocks, bonds, bank deposits, etc. — discovery becomes a practical impossibility.

Valuation presents even more difficulty. The typical local assessor, generally without expert knowledge of values, is required by law to appraise, within a period of a few weeks, all of the wealth in his town or county, including farms, factories of various kinds, complicated machinery, raw materials, "goods in process," and finished products of manufacture, stock in trade of the stores of utmost variety, and household effects of all sorts. In this task he will obviously not receive much aid from the taxpayers themselves, and he will be under constant pressure to make his valuations low.

Expert assessment. This situation is met, more or less adequately, in many of the larger cities and in some small cities. A staff of trained assessors, with the necessary clerical and

other assistance, is built up. The work is so arranged as to continue through the year, permitting of long-term or permanent employment, and giving to the assessor a steady "career" job, with adequate salary. Accurate maps of the entire region are made and skillfully used. Complete descriptive data regarding land and buildings are prepared and recorded. Evidence regarding all sales is gathered. Scientific rules for determining values with respect to location and various other standards are used. Even the airplane is pressed into service for the quick mapping of extended areas in the suburbs or country districts. In many an American city, where such conditions prevail, a high degree of efficiency, in the assessment of land and buildings at least, has been attained.

Some of the states have obtained a certain degree of centralization of the work of assessment in the office of a state tax commissioner or board, which has enabled them to secure full-time services of engineers and other experts to aid in the work of local assessment.

The ordinary local assessment. But in spite of assistance now and then from the state tax commission, assessment is still essentially a local function, and the ordinary town or county or small city will seldom have anything like an adequate assessment organization or procedure. There is not enough work to warrant a full-time job with attractive salary. Expert qualifications cannot be and are not required of the candidate for the assessor's job. There is lacking both the equipment and the personnel requisite for expert assessment.

Under such circumstances as these, we need not be greatly surprised at the result. The assessor generally does about the best he can. In the first place he has last year's lists, of which he takes full advantage; copying last year's figures will suffice for the majority of cases where he has reason to believe there has been no material change. He will travel over his district, making brief inspections of the outside of new buildings, casting cursory glances at the merchants' shelves, asking a few questions, picking up what gossip he can as to the worldly

goods of his fellow townsmen. Using such information as this, reënforced by a good deal of often shrewd guesswork, influenced almost always by the popular satisfaction which will follow when everybody feels that he has been assessed considerably under true value, influenced now and then no doubt by the desire to be kind to a friend and to stand well with those of political and social influence, seldom downright corrupt, the assessor manages somehow to check up on the lists handed in by taxpayers, to make out the lists of those who fail to file them, and finally to turn in a town tax list which will not cause too many appeals to the board of relief, nor be too much censured by the county or state board of equalization, nor cause so much local dissatisfaction as to endanger his reëlection.

Practical results. Under these circumstances it is not surprising that the accomplishment of the property tax falls far short of the ideal of the law. United States Census statistics show that the total assessed value of real estate is not more than half the true value legally taxable and that less than one-fifth of the true value of taxable personal property is assessed. These totals moreover give no picture of the great inequality among individual properties, which are assessed all the way from zero (where the property is not discovered) to more than 100 per cent of the value. As to intangible property the breakdown in most parts of the country is nearly complete.

As certain properties escape, through failure to be listed or through undervaluation, the tax rates must be proportionally higher, and the burden is increased on other properties. On properties whose obvious face values do not permit of undervaluation, existing high tax rates (30 mills, 40 mills, or even as high as 80 mills) may be quite confiscatory and virtually force the taxpayer to concealment. Prevailing undervaluation moreover results in essential denial of the legal remedy for unjust assessment through appeal to the courts, since the taxpayer who is assessed at a higher ratio than his

neighbors may still be assessed at less than true value and so less than the law requires. The attempt to correct wholesale undervaluation by the process of equalization generally fails; the county and state boards of equalization are rarely equipped to make an accurate check of the original assessments in the towns or counties.

Confusing wealth and property. When a property right is expressed by a document, such as a stock certificate, a bond, or a mortgage, the property tax commonly includes the document as taxable property. This is double taxation. The present attitude of American state tax laws on this matter may be roughly summarized as follows: (1) shares of stock of domestic corporations are taxable to resident owners in somewhat less than half the states; shares of foreign corporations are similarly taxable in somewhat more than half the states; (2) bonds of corporations, whether domestic or foreign, are taxable to resident owners in about three-fourths of the states; (3) mortgages are generally exempt; when secured by property within the state, nearly all the states grant exemption, and a large majority exempt even mortgages secured by outside property.[1]

Recognition of the peculiar position of intangible property, which has been pointed out, has led a number of states to provide for the taxation of such property at rates much lower than those applied to tangible wealth. It has been the hope that the taxpayers, when granted a reasonable rate, would voluntarily list their intangibles, and the experience of several states has shown some increase in the amount of such property assessed following a reduction in the rate. There is no evidence however to show that any state has by this expedient succeeded in assessing the whole or even the greater part of the intangible property subject to its tax laws.

Taxation of corporations. The taxation of corporations presents several special problems. In the first place, there is

[1] Cf. National Industrial Conference Board, *State and Local Taxation of Property*, 1930.

the complex problem of double taxation resulting from taxing both the corporations on their property and the owners on their stocks and bonds, to which reference has been made. Secondly it is the corporations which are the principal owners of those complicated categories of property whose valuation proves too much for the ordinary town or county assessors. In the third place, the wealth of many great corporations is located in more than one taxing district, being spread over several towns or counties or even states. The property cannot be assessed piecemeal, and no local assessor has authority to assess the whole. This situation is peculiarly evident in the case of the railroads, telephone and telegraph companies, and other public service companies. These problems have been met in many states by taking the assessment from the local bodies and entrusting it to a state commission. In addition special methods are frequently applied, very generally to the public utility corporations and the banks and less commonly to the general business corporations.

Reform of the property tax. The first step toward a sound property tax is the abandonment of all taxation of intangible personal property. This is indicated both by the theoretical nature of intangible property rights and by the complete and inevitable breakdown in administration. This step has been recommended by numerous tax authorities, including a committee of the National Tax Association, and it has been already accomplished by some ten states.

Next in order is the abandonment of the attempt to tax all those classes of tangible personalty where experience has shown the practical impossibility of a real assessment. This would include household furniture, private libraries, musical instruments, clothing and jewelry, and similar objects of essentially personal or domestic character. We may as well admit that these classes of property never have been and never will be properly assessed and give up attempting the impossible.

A special problem is presented by machinery and goods

and materials of manufacturers and merchants. We have the choice of either giving up the taxation of such goods, as has been done by New York, or consigning their assessment to the state tax commission, with authority to use when advantageous specified rules of assessment which do not necessarily arrive at market value. This second remedy has recently been recommended by a special tax commission in Connecticut.[1]

These changes as to personal property would leave for assessment by the local assessors only real estate and a few forms of bulky, easily discovered, and easily valued tangible personal property. Since the classes of property which it is proposed to exempt are the ones which now generally escape, the loss of revenue from this reform would not be great.

There are, as we have seen, certain cities whose assessment of lands and buildings has attained a high degree of efficiency. The organization requisite to produce this result is possible in any town or county where the finances are on a large enough scale to warrant the expense, and for the smaller and less wealthy towns and counties a consolidation of the work of assessment in the hands of larger units or groups of units is possible. There is no reason why any state should not provide assessment machinery accurate enough to accomplish its purpose; *i.e.*, the equitable distribution of the cost of government.

The property tax, with all its defects, is still the backbone of the American tax system. During the years of depression since 1929, it proved the anchor to windward that saved American public finance from a vastly greater shipwreck than it actually suffered. Its relative contribution, as compared with other taxes, increased materially during the depression. Fortunately the defects of the property tax are not incapable of remedy, along lines which have been pointed out.

The income tax: Historical background. In contrast

[1] *Report of the Connecticut Temporary Commission to Study the Tax Laws of the State and to Make Recommendations concerning Their Revision*, 1934, Chapter X.

with the property tax, the income tax is a modern tax. The first real income tax was introduced in England in the year 1798. It was intended as only a temporary measure, though it actually lasted till 1816. Restored in 1842, the income tax has since then been a permanent part of the British revenue system. Italy, Germany, and many other nations introduced the income tax at various dates during the latter half of the nineteenth century.

In America the states made no use of this tax before 1840, and such state income taxes as were introduced thereafter in the nineteenth century were abortive and of no permanent significance. The federal government first made use of the income tax at the time of the Civil War. This was a war measure; its yield declined rapidly after the war, and it was repealed in 1872. In the year 1894 the national government made its second excursion into the field of income taxation; this act was declared unconstitutional before it went into effect. It was not until 1913, following an amendment of the federal Constitution, that the income tax became a permanent part of the tax system of the United States government. About this time state income taxation also (beginning with the Wisconsin law of 1911) acquired a fresh lease on life. Today the income tax is the principal tax of the federal government and a source of revenue, more or less important, in more than half of the states.

The United States income tax: General description. In the present federal income tax (revenue act of 1939) the law distinguishes between individuals, corporations, and partnerships. As to the former, all residents of the United States are in general subject to the tax upon all their taxable incomes. Non-residents are generally taxable upon income derived from sources in the United States.

The individual income tax is upon the *net income*, obtained by making certain *deductions* from the *gross income*. A personal *exemption* of $1,000 for a single person and $2,500 for a married person or head of a family is allowed. There is also a

credit of $400 for each child under eighteen years of age and each person incapable of self-support who is supported by the taxpayer. An *earned income credit* is granted with a view to taxing earned incomes somewhat more lightly than incomes from investments. In the present law (revenue act of 1939) there is a normal tax of 4 per cent and a surtax from which there is an additional exemption of $4,000; the surtax rates progress from 4 per cent to 75 per cent, the latter rate applying to the excess of net income above $5,000,000.

Corporations are separately taxed. Under the revenue act of 1939, corporations with taxable income not exceeding $25,000 pay at rates graduated from 12½ per cent to 16 per cent. For corporations with taxable net incomes in excess of $25,000 the rate is 18 per cent. The law contains many complex and frequently arbitrary provisions affecting the definition of taxable income.

Partnerships as such are not taxable on their income. Their net income, whether actually distributed or not, must be allocated to the several partners and included by each in his individual income.

Individual income tax base. The result of the personal exemptions and credits for dependents permitted in the United States is to give the individual income tax a very narrow base. On the incomes received in the year 1936, returns were filed (to August 31, 1937) by 5,334,513 individuals, of whom only 2,816,657 had taxable incomes. With a population of 128½ millions in 1936, returns were filed by slightly over 4 per cent of the population and income tax paid by less than 2¼ per cent. This does not mean of course that the income tax reached less than 2¼ per cent of the people in 1936. But making allowance for married couples and minors and other dependents, it is clear that the federal income tax reaches only a small minority — probably not over 10 per cent — of all the people.

In most other countries the personal exemptions and credits are much less generous than in America. Yet in any case, so

long as any substantial minimum is exempted, the individual income tax is bound to rest upon a narrow base.

Corporation dividends. The taxation of the entire net income of corporations may be regarded as double taxation to the extent of the part that is distributed in dividends to stockholders, since the latter are then taxable on their dividends as individual income. Until 1936 this was recognized, and a partial correction was made by requiring individuals to pay only the surtax (not the normal tax) on their dividend income. The effect of the present law is to place a heavier tax on income from investment in the stock of corporations.

Undistributed income tax. The United States revenue act of 1936 introduced a novel feature in a special heavy tax, at rates from 7 to 27 per cent, on the undistributed earnings of corporations, in addition to the normal tax on all their taxable net income. The chief purpose was to force a greater distribution of earnings, retarding the accumulation of corporation surpluses, and bringing more income under the individual income tax. From the taxation standpoint, there is no justification for this device. As regards the corporations, the effect is to introduce an influence upon the disposition of corporate earnings which may be contrary to the judgment of the responsible directors and managers. Special injury is likely to be done to corporations which by law or contract are restricted in their payment of dividends, to those that are seeking to retire debts or to pay off past losses out of current profits, and to the young, small, and growing businesses generally. The undistributed earnings tax was in large measure removed by Congress in the revenue act of 1938 and completely eliminated in the act of 1939.

Progressive taxation of corporations. The United States revenue acts of 1935 and 1936 introduced another novelty in the form of progressive tax rates upon the incomes of corporations. Progression is generally applied in the taxation of individual incomes, upon the principle of distributing the burden in accordance with ability to pay. But this principle

appears to have no valid application to the taxation of corporations. The equity of taxation imposed upon corporations can be judged only in relation to its effects upon natural persons. The burden of a corporation income tax rests chiefly upon the stockholders. If it were true that the large corporations were owned by wealthy stockholders and that the stockholders of the small corporations were poor or moderately well-to-do, there might be a rough justification for a progressive corporation income tax. But this premise does not agree with the facts. The shares of some of our largest corporations, such as the American Telephone and Telegraph Company, the United States Steel Corporation, and many others, are very widely distributed among small stockholders. On the other hand, many wealthy men are sole or chief owners of incorporated businesses which, while representing substantial individual fortunes, are comparatively small as corporations go. Under such circumstances the graduated tax operates to defeat the ends of justice. Recent amendments have left only a vestige of the system of graduated rates, which are now limited, as the reader has observed, to the small corporations only (those with net income not exceeding $25,000).

The graduated income tax doubtless receives some support from those who wish to discourage the large corporations. The rather questionable merits of this method of regulation would scarcely seem to justify the resulting injustice to the smallest investors.

Income tax rates. The World War brought a revolutionary increase in income tax rates. In Great Britain, for example, during the fifteen years before the war, the basic income tax rates varied between 3½ and 6¼ per cent, with a super tax of 2½ per cent after 1910. The present United States income tax was inaugurated in 1913 with progressive rates from 1 to 7 per cent. Today income tax rates have everywhere soared to hitherto unimagined heights. In the United States the present maximum rate (normal tax and surtax) of 79 per cent is the highest tax rate known to modern history, and to this must

be added additional income taxes in about half of the states.

Since the World War the practicability of such rates has come under serious question, especially in the United States. It is possible for income tax rates to be so high as to discourage investment in industry. Investors and entrepreneurs may conclude that they are not warranted in taking all the normal risks of industrial enterprise and in addition facing the prospect that in case of successful ventures the government will take the major part of the profits, while the burden of losses will fall entirely upon them. During the past few years this has been the attitude of many American investors and business men, and it is believed by many that this has been one of the major causes of the continuation of industrial depression. In consequence the tax revenue is less than it probably would be with lower rates; in other words, the rates have been pushed beyond "what the traffic will bear."

There is another important aspect of excessive income tax rates. Even with the best possible administration, which the United States confessedly does not have, the success of an income tax must depend materially upon a prevailing spirit of fairness and coöperation on the part of the taxpayers. So long as this spirit is general, administrators of the law and the courts can cope with the actual criminal evaders. When this spirit disappears, when there is a general feeling that only the quixotic few pay their full taxes, when the majority become lawbreakers, then no administrative machinery that human intelligence can devise will be able to master the problem. This condition is threatened if rates are made too high.

Elasticity of the income tax. In the early days of the income tax it was rightly regarded as a peculiarly elastic source of revenue, adaptable to normal variations in fiscal needs, and a tower of strength in time of war or other national emergency. The recent trend toward very high rates has seriously impaired the elasticity of the income tax. Moreover this tax is peculiarly sensitive to general economic conditions. This was rudely disclosed with the coming of the depression after 1929.

The financial exigencies of government led to increased demands upon the income tax, to which the tax, with its existing high rates, was not able to respond. In the United States, for example, even with increased rates, the yield of the income tax fell off disastrously. In 1930 the income tax produced 2,410 million dollars, contributing 58 per cent of the total revenues of the federal government. In 1933 its yield was 746 millions, being only 36 per cent of a total revenue which had itself dropped to less than half the level of 1930.

Shifting and incidence of taxes on property and income: Taxes on economic rent or land value. We have already learned in a general way that shifting of the burden of a tax by the immediate payer to someone else is conditioned upon a power, given by the tax itself, to raise the price of some commodity or service which the immediate payer has to sell. The simplest theoretical case is that of the tax upon the landowner's income from economic rent. Having in mind the nature of economic rent (which was developed in Chapter XV), we readily reach the conclusion that a tax upon economic rent gives the landowner no power to increase the rent he may be charging. The tenant (assuming he is already paying the full economic rent) can pay no more rent without being forced to operate at a loss. Rather than do this he will leave this land and either go out of business or betake himself to the marginal land, where there is no rent and no tax on rent. On the other hand, neither the tenant nor the landowner who is using his own land is given any power to charge a higher price for his crop by a tax on economic rent. The price is determined by demand and the cost of production on the no-rent land, neither of which is affected by the tax in question. We conclude that a tax on the economic rent of land is not shifted but is borne by the immediate payer, the landowner.

One necessarily reaches the same conclusion as to a property tax upon land alone, since the value of the land is the capitalized value of its economic rent. A property tax on

land can certainly give the owner no power to sell the land at a higher price; nor does it give him power either to exact more rent from a tenant or charge higher prices for agricultural products, since the marginal land, having no economic rent and so no value, is no more affected by a property tax than by a tax on rent.

Capitalization of taxes. The tax on economic rent or on land value furnishes the simplest illustration of the principle of "tax capitalization." If an acre of land is yielding an economic rent tax-free of $10 a year and the rate of interest at which such incomes are discounted is 5 per cent, the land will be worth $200. If now, assuming no other material change in the tax system, an annual tax is imposed on economic rent at the rate of 10 per cent, the landowner will have left an income of only $9 a year, which, capitalized at 5 per cent, makes his acre of land worth $180. Taking 10 per cent of the annual income is equivalent to taking in one lump sum 10 per cent of the value of the land. No one will buy this land for more than $180, and the purchaser, receiving $10 in rent and paying $1 in tax, obtains 5 per cent on the investment, the same as he would have received if he had paid $200 for an acre not subject to the tax on its rent. Of course precisely the same result follows the imposition of an annual property tax on the value of the land. Taxes on economic rent or land value thus tend to rest as a permanent burden upon those who owned the land at the time of the imposition of the taxes and not to be passed on to later purchasers. The principle of capitalization applies not only to land but to any other capital instrument where it is not possible to shift the tax — for example a tax on bonds.

Taxes on dwelling houses or their rent. The rent paid for a dwelling house is distinguished from the rent paid for the location by the obvious fact that buildings are constantly wearing out and being replaced with new ones, whereas land is permanent and practically fixed as to quantity. Before the imposition of a tax on the rent of dwelling houses, there was

a certain normal return obtainable from renting dwelling houses; now that return has been reduced by the amount of the tax. Investors will tend to stop building new houses, preferring to put their capital into other ventures where the normal return is obtainable. The diminished supply of houses for rent, as old houses wear out and are not replaced by new ones, will cause an increase in the rents which house owners can exact and tenants will pay, until the rents are high enough to encourage new building. Our general conclusion then is that the final incidence of a tax either on dwelling houses or on their rent tends to be upon the occupiers, whether they be owners or tenants, the tax being shifted when imposed upon the owners of rented houses.

Taxes on business property or income. The property tax on wealth used in business, such as factory and store buildings, warehouses, railway equipment, machinery, goods and materials, etc., generally means an addition to the cost of the products of business enterprise. Under competitive conditions anything that increases cost of production will tend to raise prices. Even under monopoly conditions, there will generally be some tendency to raise prices. It is thus that a property tax upon wealth, other than land, used in production is generally shifted through higher prices of products and so finally rests as a burden upon the consumers of the products. In the case of the tax on competitive business profits there may be a certain degree of shifting through the charging of higher prices. But since the tax encroaches only upon net income or profit, and takes only a fractional part of that, an income tax at moderate rates will not generally drive business men to sacrifice such profits as remain to them. However very heavy rates, such as have recently been imposed in the United States, especially when accompanied by legal definitions which make the tax base different from true net income, undoubtedly do deter investors and entrepreneurs from risking their capital in business and so tend to raise the prices of the products of industry.

In the case of a monopoly an income tax upon profits cannot possibly be shifted. The monopoly, having already set its price so as to reap the maximum profit, has nothing to gain by a change in price, even though it face a tax up to 100 per cent of its profits. It is better to retain any given fraction of a larger net income than the same part of a smaller income which would result from a change in price.

The poll tax. This is a tax, generally at a uniform rate of one to three dollars, upon every adult person, or every voter, or some other broadly inclusive class of the people. It is expensive to collect, except from those who are already subject to the property or income tax, and while it is (on paper at least) a recognized part of the American local tax system, it is generally honored more in the breach than in the observance, and the revenue derived from it is usually small.

Of late the poll tax has come into somewhat more favorable recognition from two viewpoints. Appreciation of the great importance of having everyone feel some financial stake in his government and some direct interest in its finances has led to the proposal of a universal income tax, which would allow no personal exemption and would start with a minimum tax of from one to five dollars. Such an income tax would obviously absorb the existing poll tax. This plan has much to commend it. Thus far it has produced only a few tentative steps in actual legislation.

On the practical side, the poll tax has recently proved itself useful in bringing home to the people financial responsibility for certain government expenditures conferring special benefits upon specified classes of the people. Thus the state of Massachusetts, after the World War, granted a bonus to war veterans and raised the necessary funds by means of a state poll tax. Connecticut in 1935 inaugurated a system of old age pensions, with provision for raising the money by means of a town administered poll tax.

Consumption taxes. Next in importance to the property tax and the income tax comes a group of taxes whose central

characteristic is the exaction of a tax contribution from the consumers of certain goods in proportion to their consumption. They appear in a great variety of forms, sales taxes, excise taxes, licenses, customs tariffs. For the purpose of our study we may divide consumption taxes into two main groups: (1) excise taxes (including sales taxes and business license taxes), and (2) customs tariffs.

Excise taxes: General account. An excise is in general a tax levied upon the manufacture or sale or consumption of commodities or the exercise of an occupation or enjoyment of a privilege within the limits of the jurisdiction imposing the tax. What will chiefly interest us are the taxes levied upon the domestic manufacture and sale of commodities, such as are commonly known in the United States as "internal revenue" taxes when imposed by the federal government and "sales taxes" when imposed by the states.

Excise or sales taxes played only an insignificant rôle in the United States until the time of the Civil War. Starting in 1860, Congress had before the end of the war built up an elaborate system of sales taxes, covering nearly every commodity of wide consumption, levied now upon manufacturers, now upon dealers, often involving double or multiple taxation of the final product, and carrying a wide variety of rates.

After the war this elaborate system was generally repealed, leaving as a permanent part of the federal revenue system only the excises on liquors and tobacco and a few minor items, such as playing cards, oleomargarine, etc., whose retention was largely for sumptuary purposes or to favor the producers of a competing commodity. The liquor and tobacco taxes are levied upon the manufacturers. At the time of the Spanish War in 1898–99 and again during the World War, the United States government obtained large additional revenue from excise taxes upon other selected commodities and services. On each occasion these excise taxes were generally repealed soon after the war. The depression following 1930 similarly occasioned some fresh imposition of excise taxes.

Excise taxes are generally used by national rather than by local governments. This is principally because of the tendency of such taxes to drive business outside the borders of the taxing jurisdiction. In the United States the several states have generally been deterred from extensive use of excise taxes by these considerations, coupled with the constitutional prohibition of customs tariffs and excises burdening interstate commerce. The states are not prevented however from levying moderate excise taxes upon industries catering to a strictly local market or upon local or retail sales, though caution has always to be exercised lest the people be driven to make their purchases outside the state.

One of the most striking features of post-war finance all over the world is the rapid and widespread development of sales taxation, forced by the fiscal needs of governments harassed by increased demands for expenditure and dwindling sources of revenue. Among the American states there has been a really extraordinary development of this form of taxation. The excise tax on gasoline is now universal. Excise taxes or license taxes are imposed upon sales of liquor by all states which do not prohibit its sale. As to consumption generally, by the close of 1937 more or less general sales taxes were in force in some twenty-five states.

Shifting and incidence of excise taxes. The excises illustrate the operation of shifting in its simplest form. Consider for example the case of the United States internal revenue tax upon cigars. The tax is imposed upon all manufacturers of cigars at a certain rate per hundred and is thus an addition to the normal unit cost of production. Manufacturers are therefore able to increase the selling price by practically the amount of the tax and so to pass the burden on to the wholesale dealers to whom they sell their product. These in turn pass the tax on to the retailers, who pass it on in the prices charged the consumers. Sometimes an excise tax may be imposed upon the dealers, wholesale or retail, instead of upon the manufacturers, or upon the consumers themselves at the

time of purchase of the taxed articles, the law requiring the dealer to collect the tax and pass it back to the government. In all such cases the price paid by the consumer is of course increased by the amount of the tax. When the excise is based upon an industrial occupation, instead of upon a commodity, the same result follows; the tax becomes an added cost of conducting the particular occupation and is passed on to the consumers by means of a higher price for goods or services. The retail liquor seller's license is a familiar example.

It is not to be inferred that an excise tax is wholly without burden upon the manufacturers and dealers; indeed the re-

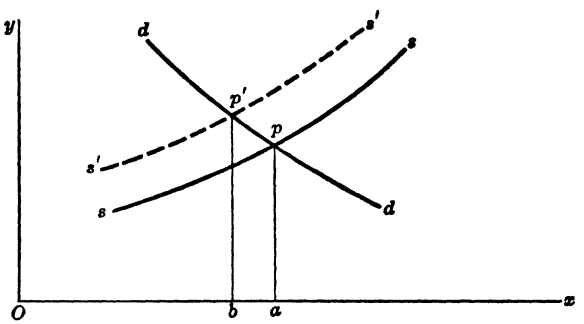


FIG. 25.

verse is generally true, due to the normal shape of the demand curve. Let *dd* and *ss* (Figure 25) represent respectively the demand and supply curves of a certain commodity, determining the selling price *pa*. Now suppose an excise tax is imposed upon the manufacturers or dealers. The unit cost being thereby increased, there will be a new supply curve *s's'*, higher than the former curve. The intersection of this curve with the demand curve, of course not changed by the tax, will determine a new price *p'b*. This price is higher than the old price, but except in the case of constant cost industries it is generally not increased by the full amount of the tax. What is of more importance, the quantity sold has fallen off (from *Oa* to *Ob*). The manufacturers and dealers will, after the process of shifting, be able to charge enough to give them the

ordinary rate of return, but only on a smaller volume of business.

The customs tariff. Customs or tariff duties are taxes imposed upon those who bring goods across the frontier, the base being the goods so moved. The Constitution of the United States prohibits all duties on exports and also practically forbids the levy of any customs duties by the states. From the beginning of the American nation in 1789 to the Civil War, the tariff was almost the only tax imposed by the national government, and from then to 1913 it divided honors about equally with the excise or internal revenue taxes. Today the tariff yields about one-tenth of the federal revenue.

Economic effects of the tariff. It may help the reader to grasp the salient characteristics of the tariff, together with their immediate consequences, if we set up a simple numerical example. Let us suppose that a certain kind of cloth is being imported, free of duty, and sold in the home market at $1.00 a yard and that 1,000,000 yards are sold in a given year. We assume that the cost of making this cloth at home is $1.50 a yard. Obviously the reason for the import is that this is the cheapest way to obtain the cloth, and of course there will be no home manufacture. The simple situation resulting is shown in column 1 of the following table:

	1	*2*	*3*	*4*
	No Tariff	10% Tariff	50% Tariff	75% Tariff
Number of yards imported	1,000,000	900,000	300,000	———
Number of yards domestic	———	———	300,000	600,000
Total number of yards consumed	1,000,000	900,000	600,000	600,000
Price per yard to consumers	$1.00	$1.10	$1.50	$1.50
Cost to consumers	$1,000,000	$990,000	$900,000	$900,000
Additional cost because of tariff	———	90,000	300,000	300,000
Revenue to government	———	90,000	150,000	———
Additional cost in excess of revenue	———	———	150,000	300,000

Now suppose that next year a tariff is in effect at the rate of 10 per cent, which will make the cloth cost $1.10 in the home market. This price being still less than the home cost,

there will be no home manufacture, and all the cloth consumed will still be imported. In accordance with the normal demand curve, and assuming other circumstances unchanged, we assume that the amount consumed will fall off somewhat with the increase of price and that 900,000 yards will be taken in this year. The result is shown in column 2 of the table above. Consumers pay $990,000 for cloth which, were it not for the tariff, they could have obtained for $900,000, and this additional cost goes to the government in revenue. Here is obviously the case of a simple tax on imports.

Next year a 50 per cent tariff is in effect, and the imported cloth will be sold, if at all, at a price of $1.50. This price is just equal to the cost of making the cloth at home, and the foreign and domestic products are in competition on an equal basis. Presumably they will divide the market between them in some ratio, which for the sake of simplicity we may assume will be an equal division. The result is shown in column 3 of the table, assuming a still further drop in the amount taken because of the rise in price to $1.50. Consumers now take 600,000 yards, paying for them $300,000 more than they would have paid for this quantity had there been no tariff. This additional cost however does not now all go to the government; the government collects 50 per cent upon the 300,000 yards imported, giving it a revenue of $150,000; but it gets no revenue from the 300,000 yards of domestic manufacture. We meet here for the first time the phenomenon of *protection*, by which we mean *encouragement given to a home industry by means of a tariff*.

Now let us proceed to the last case, that of the *prohibitive* tariff. A duty of 75 per cent being in effect, the foreign cloth can be sold only at a price of $1.75. Since the home industry can furnish the cloth (at least 600,000 yards according to our assumption) for $1.50, there will be no possibility of selling the imported article. This situation is shown in column 4 of the table. So far as the consumers are concerned, the situation is the same as in the previous case, except that all their

purchases are of the domestic cloth. The government however now gets no share at all of the $300,000 additional cost imposed by the tariff upon the consumers. This is the extreme of protection.

To the reader who has fully grasped the meaning of the foregoing example, the most important principles of the customs tariff will now appear almost as a matter of course. They may be summarized in the following language (the reader will find it helpful to refer again to the numerical example above):

(1) A tariff will cause the price of the dutiable article to rise in the home market by about the full amount of the duty, up to a duty which is equal to the difference between the home and the foreign unit cost of production.[1] Above that the duty is prohibitive, and the further result depends on whether there is competition or monopoly in the home industry. If competition prevails, further increments of duty have no effect on the price; under monopoly conditions the price may rise by more than the difference in unit production cost up to the full amount of the duty.

(2) A tariff imposed at a rate less than the difference between the home and the foreign unit cost of production will have no immediate effect on home industry and will produce a revenue equal (disregarding costs of administration, etc.) to the added burden placed upon consumers.

(3) A tariff equal to or greater than the difference in unit production costs gives protection, and the additional burden placed upon consumers will be in excess of the revenue obtained by the government. If the tariff is in excess of the difference in unit production costs it is prohibitive, in which case the additional burden on consumers is accompanied by no revenue whatever to the government. Stated in slightly different words, any tariff which gives protection to domestic in-

[1] The word "about" is introduced in this sentence in order to allow for the possibility of a change in the foreign cost of production resulting from the decreased purchase of the imported goods. The practical importance of this would generally be slight.

dustry necessarily causes a loss to consumers greater than the gain obtained by the government.

(4) The additional burden placed upon consumers by a tariff which gives protection will be equal to the difference between the home and the foreign cost of producing all the goods (domestic and foreign) consumed, except when a prohibitive tariff accompanies a monopolistic condition in the home industry. In that case, the consumers may be burdened by an amount in excess of the difference in the entire cost of production.

The revenue tariff. The traditional country of "free trade" was Great Britain. From the reform of her tariff system in the middle of the nineteenth century until after the World War, Great Britain followed a policy of tariff for revenue with virtually no protection. In order to make her tariff truly a revenue producer without disturbance from protective influence, Great Britain adopted certain devices which are universally effective for this purpose. In the first place, her duties were placed upon a comparatively small number of articles, these being generally commodities which in the nature of the case could not profitably be produced in Great Britain even in the face of high tariff duties. Such are tea, coffee, sugar, cocoa, spices, various fruits, and other tropical products. The second device for safeguarding the revenue tariff from defeat through the checking of imports and the development of home industries is to impose an internal excise duty on each home-made good to match the tariff duty on the imported article. Finally there is the extreme safeguard of prohibiting outright the home industry, which is employed by Great Britain in the case of tobacco growing. By such means as these it is possible to build up a tariff system of extreme simplicity, easy to administer, having no disturbing influence upon the country's trade and industry, taking nothing from the people beyond what goes into the public treasury, and withal yielding a handsome revenue.

Fiscal weakness of the protective tariff. Few countries however outside of Great Britain have chosen this kind of tariff. Virtually everywhere protection is consciously sought to a greater or less degree, and the employment of devices to neutralize the protective tendency of duties imposed for the sake of revenue is rare. At the extreme protective pole stands the United States, where, in spite of the great importance of the tariff as a source of revenue, the protective purpose has always been strong or even predominant.

Throughout the history of the United States, the tariff revenue has been notoriously irregular, out of relation to governmental needs, and not amenable to legislative control. In 1807 the tariff revenue was double the year's expenditures; in 1837 the tariff, still the sole tax revenue of the federal government, produced less than one-third of the amount necessary to cover the year's expenditures. The tariff failed dismally to meet the war emergencies of 1812 and 1861. Since the Civil War the steadying influence of the excise taxes has somewhat mitigated the gyrations of the tariff revenue, but throughout the entire national history up to the World War dependence upon the tariff has been, more than anything else, the cause of that succession of uncomfortable deficits and embarrassing surpluses which testify to an imperfect financial system.

The explanation is simple. The tariff being looked upon as an instrument for giving protection to certain industries, its rates cannot be raised or lowered to meet fiscal conditions without opening up the whole political conflict over protection. Furthermore the relation between tariff rates and tariff revenue is extremely complex, and it is impossible to predict accurately the effect of a change. By checking the flow of imports an increase of rates may diminish revenue instead of increasing it; a reduction of rates may increase revenue.

Death taxes: General description. The death of the owner gives to government a convenient opportunity to impose taxes from time to time upon the transfer of property of every de-

scription, an opportunity which is availed of by sovereign governments quite generally the world over.

The typical modern *inheritance tax* contains the following technical features. At the outset property left to certain religious, educational, or charitable institutions is generally entirely exempt from the tax. Of the remaining taxable estate, a certain amount is usually exempt, this amount varying according to the degree of relationship between the decedent and the beneficiaries. Above the exempt minimum, the tax is calculated at rates which are usually doubly progressive: (1) the rates start low for bequests to the closest relations and increase step by step as the degree of relationship becomes more remote, and (2) within each class of beneficiaries, the rate is progressive according to the amount transferred. Among the states of the United States the top rate on direct heirs varies all the way from 3 per cent to 16 per cent; on the most remote relations and strangers the maximum rates run as high as 40 per cent.

A tax levied upon the entire estate without relation to the amounts of the distributive shares or the relationship of the beneficiaries is called an *estate tax.* As an example we may take the United States estate tax, as amended by the revenue act of 1935. The tax, as regards resident decedents, is imposed only upon estates whose amount exceeds $40,000 and is based upon the "net amount" of the entire estate, calculated by subtracting a lump sum of $40,000 besides other deductions. With respect to the net estate so defined the rate begins at 2 per cent upon the first $10,000 and increases progressively until the maximum rate of 70 per cent is reached, applying to that part of the net estate in excess of $50,000,000.

Some fundamental questions. A bequest, except where received by a dependent of the testator, comes as a more or less extraordinary windfall, different from one's ordinary regularly recurring income. This is usually taken to justify quite heavy taxes upon estates and inheritances as compared with the other kinds of taxes on property, income, etc. Similar con-

clusion is drawn from the prevailing concept of property rights, which holds them to be distinctly less sacred when it comes to disposition of one's possessions after death, especially that part which is not passed on to those immediately dependent upon the testator. There are those who would go still further and use the inheritance tax as an instrument to check the rise or the perpetuation of large family fortunes, and there can be no doubt that the very high rates in the upper brackets of some death taxes — witness the United States estate tax, and especially its amendment in the revenue act of 1935 — are partly in response to a demand for equalizing wealth by heavy levies on the large fortunes.

Death tax rates, at least as regards bequests to indirect heirs and strangers, might be pushed very high before reaching any limit set by the principle of ability to pay. But before this limit is reached there is almost certain to be a limit set either by administrative considerations or by broad social policy. There are many kinds of business arrangements and legal devices whereby death taxes may be escaped while still keeping within the letter of the law. There is always the possibility of escape through gifts before death, a door which no development of gift taxes is likely completely to close. The higher the rates the more will such avenues of escape be resorted to. The second limit is reached when the tax becomes so high as to discourage the accumulation and husbanding of capital or to interfere with its most productive use. While it is probable that this effect of the inheritance tax has been frequently exaggerated, there would eventually come a time when the public loss from decline of capital and disturbance to business and investment would exceed the possible benefits from the revenue thus contributed.

CHAPTER XXVII

Public Finance

Cost of government and national income. In the year 1938 there were expended by government in the United States about 17 billions of dollars. The federal government's share was about 7⅞ billions; the 48 states spent about 4⅛ billions; and the various local governments in all the states accounted for about 5½ billions. These are huge sums, beyond the power of the imagination to visualize. Somewhat easier to grasp is the fact that government expenditure in that year took more than a quarter of the total income of the people of the United States, estimated at 62¼ billion dollars. If perchance the situation in 1938 may be considered abnormal, on account of depression and "New Deal" expenditures, we may go back to the year 1929, when the total expenditures of government were about 12 billions and represented about 15 per cent of a national income of 79½ billions. It will be noted that between these dates there had been both a drastic decline in the people's income and a heavy increase in the cost of government.

Growth of public expenditure: General summary. The constant growth of public expenditures is an outstanding phenomenon of modern history. Starting as far back as reliable records will take us, this universal tendency appears. It applies to all types and grades of government: monarchies and republics, peaceful nations like Switzerland as well as militaristic Germany, imperial governments, national governments, states, and municipalities. The increase in government expenditures before the nineteenth century is on the whole more apparent than real, the great advances in the absolute figures being substantially offset by such factors as diminishing reliance on goods and services directly rendered, growth

of population, increase in territory, increased wealth of the people, and depreciation of the monetary standard. But since the beginning of the nineteenth century there has undoubtedly been real growth.

United States federal government. The following table will give an idea of the growth of expenditures of the federal government during 149 years of national existence.

ORDINARY EXPENDITURES OF THE UNITED STATES GOVERNMENT[1]

Year	*Total expenditures* (In Millions)	*Per capita expenditures*
1792	$ 5	$ 1.28
1800	11	2.03
1810	8	1.13
1820	18	1.89
1830	15	1.18
1840	24	1.42
1850	40	1.71
1860	63	2.01
1870	310	8.03
1880	268	5.34
1890	318	5.05
1900	521	6.85
1910	694	7.54
1915	761	7.66
1919	18,515	176.33
1920	6,403	60.57
1925	3,063	26.60
1930	3,440	27.93
1935	7,010	51.42
1939	9,210	70.31
1940	9,637	73.01

For the first seventy years of the national existence of the United States, the expenditures of its government were on a modest scale, the increase from 11 millions in 1800 to 63 millions in 1860 being no more than sufficient to keep pace with the growth of the country's population, as is shown by

[1] The figures for total expenditures through 1939 are from *Annual Report of the Secretary of the Treasury on the State of the Finances, 1939*, pp. 362–65. The 1940 figures are as estimated in the *Budget of the United States Government for the Fiscal Year 1941*. Postal expenditures are included only to the extent that they exceed the postal revenues.

the figures for per capita expenditure. The figures for 1870 show the effect of the Civil War cost, and the later figures illustrate the universal truth that war not only adds a war-time burden of extraordinary costs but sets a new scale of peace-time expenditure from which no people has ever been able to return to the pre-war status. Other influences contributed further to the growth of public expenditures, so that from the Civil War to the World War the cost of the federal government averaged more than four times as much as before the Civil War, after making full allowance for the growth of population.

The gigantic expenditures of the World War reached their maximum in 1919. Expenditures then dropped, but they were never again as low as three billions. The new level reached in 1925 was four times as high as the expenditures of 1915, or three and a half times as high after allowing for the increase in population. In 1932 begins another decided upward swing, the result chiefly of depression and "New Deal" expenditures, which will be analyzed at a later point in this chapter.

All grades of government in the United States. It is only since the latter part of the nineteenth century that we have data enabling us to bring state and local expenditures into the picture of the growth of all government costs in the United States. The table on page 510 is the result of researches of the National Industrial Conference Board.[1]

For the year 1938 the National Industrial Conference Board offers the following preliminary estimate: federal, 7.5 billions; state and local, 9.3 billions; total, 16.8 billions.

It will be noted that division of the total cost of government between the three grades of government had not greatly changed in the generation before the depression. The federal

[1] *Cost of Government in the United States, 1926–1927*, 1928, pp. 2, 63; *Cost of Government in the United States, 1933–1935*, 1936, pp. 4, 30; *Cost of Government, in the United States, 1935–1937*, 1938, p. 4; and *National Income in the United States, 1799–1938*, 1939, pp. 6–7.

TOTAL GOVERNMENTAL EXPENDITURES, UNITED STATES

	1890	*1913*	*1929*	*1936*
	Total expenditures (millions of dollars)			
Federal	291	692	3,046	8,547
State	77	383	1,943	2,433
Local	487	1,844	6,720	6,038
Total	855	2,919	11,709	17,018
	Per capita expenditures (dollars)			
Federal	4.61	7.17	25.06	66.55
State	1.22	3.97	15.99	18.94
Local	7.72	19.10	55.30	47.01
Total	13.56	30.24	96.35	132.51
	National income (millions of dollars)			
Total	10,701*	31,450	79,498	65,246
	Ratio of expenditures to national income			
Total	8.0%*	9.3%	14.7%	26.1%

* National income as of 1889.

government accounted for about one-third of the total, the states for ten to fifteen per cent, and the local governments for something more than half. This is significant as showing that, in spite of the powerful influence of war in causing expansion of federal expenses, the costs of state and local government had not fallen behind. The climax of total government cost was in 1919. After that the federal government reduced its expenses. But in the meantime the expenditures of the state and local governments continued to rise, and after 1925 those of the federal government were also growing, so that the total cost of government was again on the increase. The effect of the depression was to increase substantially the relative share of the federal government in the total and to reduce correspondingly the share of the local governments. In 1936 the distribution was: federal government, 50.3 per cent; states, 14.3 per cent; local governments, 35.4 per cent. According to preliminary estimates of the National Industrial Conference

Board, the distribution in 1938 was: federal government, 45 per cent; state and local governments together, 55 per cent.

In 1890 a sum equal to 8 per cent of the income of the American people was devoted to defraying the cost of government. In 1929 the percentage was about 15. In 1936 the effects of depression were apparent, drastically reducing the incomes of the people and increasing rather than decreasing the cost of government. The ratio had risen to about 26 per cent. The profound significance of this development should be obvious. As time goes on a continually greater part of the income of the people is being taken and employed by government. In return government renders valuable services. Whether this distribution of the national income is giving the maximum of utility to the people and, if so, how far the portion devoted to government may continue to increase without a decrease in utility are questions to which the student may well give serious thought. They raise problems in economics, in political science, and in social philosophy which, though most inviting, must be regarded as outside the field of this elementary treatise.

Functional analysis of government costs in the United States: Federal government. In discussing the cost of government we shall have to direct our attention chiefly to the United States. As is true of all national governments, the largest item of expenditure is for defense — war. Besides war the leading expenditures of the United States government are rivers and harbors and other public works, marine and land transportation, various facilities for national industry, maintenance of justice and security, regulation of private business, sanitary regulations, immigration and naturalization service, and the maintenance of the legislative, judicial, executive, and administrative branches of the government. The table on page 512 is a summary functional analysis of the expenditures of the federal government in the fiscal year 1937.[1]

[1] National Industrial Conference Board, *Cost of Government in the United States, 1935–1937*, 1938, pp. 17, 19.

This table, while representing fairly the character of the federal expenditures of recent years, is not typical of conditions before 1932, owing to the fact that since that year the expenditures of the United States government have included very large sums devoted to efforts to combat economic depression and upon the various projects of the "New Deal." These expenditures will be further considered later in this chapter.

EXPENDITURES OF UNITED STATES GOVERNMENT, 1937

Function	*Millions of dollars*	*Per cent of total*
Regular operating expenditures		
Legislative, judicial, and civil establishments	814.7	10.2
National defense	888.6	11.1
Veterans' pensions and benefits	1,128.2	14.1
Interest on the public debt	866.4	10.8
Other	48.1	0.6
Total regular operating expenditures	3,746.0	46.8
Public Works		
Public highways	350.6	4.4
Public buildings	76.3	1.0
Rivers and harbors, improvement	142.4	1.8
Grants to public bodies, including administration	272.9	3.4
Other	237.2	3.0
Total public works	1,079.4	13.5
Relief		
Direct relief	184.3	2.3
Work relief (WPA and CWA)	1,896.7	23.7
Civilian Conservation Corps	385.8	4.8
Total relief	2,466.8	30.8
Miscellaneous		
Loans (net excess of credits)	−307.1	−3.8
Subscriptions to stock and surplus	47.1	0.6
Agricultural adjustment program	515.8	6.4
Social security	447.7	5.6
Railroad retirement	5.5	0.1
Total miscellaneous	709.0	8.9
Total expenditures	8,001.2	100.0

State and local governments. Among the expenditures of American state and local governments education stands

first. Next come highways, care of dependents and defectives, maintenance of justice and security, protection against disease, and interest on debt. The following statement is a functional classification of expenditures of the states, the counties, and the other local governments.

GOVERNMENT COST PAYMENTS, STATE AND LOCAL GOVERNMENTS, 1932 [1]

(In thousands of dollars)

Function	*States*	*Counties*	*Other local divisions*	*Total*
General departments:				
General government	123,333	251,150		
Protection to person and property	86,875	44,231		
Health and sanitation	37,292	32,778		
Highways	235,690	236,350		
Charities, hospitals, and corrections	274,920	182,120		
Schools	602,065	178,406		
Libraries	2,405	3,772		
Recreation	8,054	7,618		
Development and conservation of natural resources	72,965	.....		
Miscellaneous	57,794	44,767		
Total general departments	1,501,393	981,192	3,566,523	6,049,108
Public service enterprises	6,330	489	301,125	307,944
Interest	112,258	118,875	612,735	843,868
Outlays	885,854	311,270	1,164,571	2,361,695
Total	2,505,835	1,411,826	5,644,954	9,562,615

The American revenue system: Federal government. Light will be thrown upon this subject by a brief historical sketch of the revenue system of the United States federal government. The American colonists were always sensitive to taxation. Taxation was a sore point between them and Great Britain, and when the Revolution released them from British taxation they were loath to give to their new national government adequate taxing powers. The Articles of Confederation broke down largely because they gave the government no

[1] United States Bureau of the Census, *Financial Statistics of State and Local Governments: 1932*, 1935, p. 28.

power to tax the people, and the Constitution of the United States contained important restrictions upon the taxing power of the national government. Chief among them was the provision that direct taxation may be employed only if apportioned among the states according to their population. This rule made the practical task of levying a direct tax almost prohibitively difficult. Some timid experiments were made soon after the adoption of the Constitution and again at the time of the Civil War. But these direct taxes were ineffective and produced little revenue, and the attempt has never been repeated.

From the start the national government relied upon indirect taxation. And of the two principal types of indirect taxes, the customs tariff was at the first chosen rather than the more obvious internal excise taxes, again on account of popular antagonism to taxation. While there was some half-hearted use of excise taxes up to the close of the War of 1812, the tariff was practically the only tax resource of the national government until the time of the Civil War. It proved, as we have learned, an erratic source of revenue at best and failed completely to meet the emergency needs of war.

The Civil War was of such duration as to give the people time to overcome their traditional repugnance to internal taxation. Starting in 1862, excise taxes were enacted, which finally laid burdens upon almost every article of industry and commerce, and in the financial reorganization after that war the excise system, though greatly reduced by freeing the majority of taxed commodities, was not entirely abandoned as it had been after the War of 1812. A group of excises, principally upon liquors and tobacco, remained and provided not only an important source of peace-time revenue but also the skeleton upon which it was possible to build quickly an effective system of war taxes in time of future need, of which good use was made on the occasions of the Spanish War and the World War.

For the half century following the close of the Civil War, the tariff and the excises shared about equally (tariff revenue

being generally a little greater than excise revenue) the task of producing the revenue needed by the federal government, no other form of taxation, and indeed no other source of revenue of any sort, being of comparable significance. For example during the generation ending in 1913, the combined receipts from customs and excise provided more than 90 per cent of the net ordinary revenue (*i.e.*, excluding loans) of the national government in every year except one.[1]

The federal tax system was revolutionized by the permanent addition of the income tax — on corporations in 1909 and on individuals in 1913. Shortly thereafter the estate tax, regarded heretofore as a temporary war measure, became a fixed part of the system when Congress failed to repeal the World War estate tax. The ordinary receipts of the federal government for two recent fiscal years (the twelve months ending June 30) are shown in the following table: [2]

REVENUE RECEIPTS OF THE FEDERAL GOVERNMENT

	1929		*1938*	
	Millions of dollars	*Per cent of total*	*Millions of dollars*	*Per cent of total*
Customs	603	14.9	359	5.8
Internal revenue (miscl.)	547	13.6	2,658	42.6
Income and profits tax	2,331	57.8	2,635	42.2
Estate tax	62	1.5	382	6.1
Other ordinary receipts	493	12.2	208	3.3
Total ordinary receipts	4,036	100.0	6,242	100.0

Nothing could show more clearly the revolution in the revenue system of the federal government. Up to 1913, as we have seen, the national government had relied for its revenues almost exclusively upon indirect taxation, leaving the whole field of direct taxes to the states and their local subdivisions. Thereafter the federal government not only invaded the

[1] In 1898 this percentage was reduced on account of a large fortuitous income from sale of public property in the Kansas Pacific Railroad and the Union Pacific Railroad.

[2] *Annual Report of the Secretary of the Treasury on the State of the Finances, 1929*, pp. 375 and 420; *1938*, pp. 2–3.

direct tax field; it made direct taxation its chief reliance. In 1929 more than half of the total ordinary receipts came from the income tax. Barely over one-fourth came from the two great indirect taxes, customs and excise, which for a generation before 1913 produced more than 90 per cent of the ordinary receipts.

The depression following 1929 caused a drastic decline in income and excess profits tax receipts, which in 1933 produced only 746 millions, less than one-third of their yield in 1929. This started the pendulum swinging back again toward indirect taxation, though not toward the customs tariff, which continued to decline both absolutely and relatively. It was the excises which came to the Treasury's aid, so much so that in 1938 they were producing nearly five times as much as in 1929 and were even approaching the relative position they held in the days before the income tax appeared.

Revenue system of the American states. It is out of the question to attempt, in this general study, any detailed account of the tax systems of the forty-eight American states. The following table shows the principal forms of revenue used and the total sums obtained by all the states from each source in the fiscal year 1932.

REVENUE RECEIPTS OF THE STATES, 1932[1]
(In thousands of dollars)

Property tax	323,477
Inheritance tax	143,959
Income tax	47,852
Poll tax	5,321
Licenses and permits	979,385
Other taxes	141,856
Special assessments	26,060
Fines, forfeits, and escheats	5,833
Subventions and grants	249,282
Donations, gifts, and pension assessments	32,034
Highway privileges, rents, and interest	87,996
Earnings of general departments	154,701
Earnings of public service enterprises	10,179
Total	2,207,935

[1] United States Bureau of the Census, *Financial Statistics of States and Local Governments: 1932*, 1935, p. 7.

Of the combined state revenue receipts, amounting to 2,208 million dollars, 1,642 millions, or 75 per cent, came from taxation. The heading "licenses and permits" in the table includes the gasoline taxes, motor vehicle licenses, taxes on business, etc. The combined revenue from this group of taxes and licenses furnished 60 per cent of the total tax receipts. The property tax came next, contributing about 20 per cent. The inheritance tax accounted for about 9 per cent, and the income tax, 3 per cent. Of the non-tax revenues the principal item was subventions and grants from other governments; *i.e.*, from the federal government above and from the minor subdivisions of the respective states. The reader will note the small importance of earnings from public service enterprises.

Revenues of American local governments. The following table shows the principal sources of revenue employed by the local subdivisions of the American states and the amount contributed by each source to the several divisions of local

REVENUE RECEIPTS OF LOCAL GOVERNMENTS, 1932 [1]
(In thousands of dollars)

Revenue receipts	*Divisions of government*					
	Counties	Cities, Towns, Villages, and Boroughs	School Districts	Town-ships	Other Civil Divi-sions	Total
Property tax	877,142	2,057,495	1,118,744	195,431	112,495	4,361,307
Licenses and permits	132,050	116,635	184	5,652	2,458	256,979
Other taxes	11,643	77,159	2,108	6,575	126	97,611
Special assessments	32,146	226,430	.....	5,389	31,140	295,105
Subventions and grants	139,077	137,784	317,641	18,830	5,255	618,587
Earnings of general departments	88,956	87,415	63,517	8,535	8,817	257,240
Earnings of public service enterprises	821	487,001	28	2,143	29,826	519,819
Other	31,997	183,675	12,428	5,094	4,140	237,334
Total	1,313,832	3,373,594	1,514,650	247,649	194,257	6,643,982

[1] United States Bureau of the Census, *Financial Statistics of State and Local Governments: 1932*, 1935, p. 7.

government throughout the country. These figures are for the year 1932, the latest year for which such data, furnished by the United States Census, are available.

As we have already learned, the property tax here stands pre-eminent; it furnishes two-thirds of the total revenue receipts of all the local governments and 87 per cent of their tax revenues.

Public borrowing: Purposes. A government borrows in order to counteract a deficit, realized or anticipated. The circumstances which cause the deficit and so compel the government to borrow are various.

In the absence of extraordinary events which may arise to affect either the expenditure or the income of a government, there is normally an approach toward agreement between the totals on the two sides of the account. Even in the best regulated governments however an exact balance is seldom to be looked for. Moderate deficits and surpluses occur from time to time, and while in the long run they tend to offset each other, the periods of deficit require borrowing to make both ends meet. Such borrowing in order to meet a deficit in the ordinary running expenses is common to all governments. Especially in the United States, where budgetary procedure and financial administration are far from perfect, federal, state, and local governments are frequent borrowers for this purpose. The loans contracted by the federal government to meet the huge deficits of 1931–40 are discussed in later sections of this chapter.

No nation can meet the extraordinary expenditures of a great war out of its ordinary sources of revenue, however heroically it may expand its tax system and other revenues. Moreover the magnitude of war costs is seldom foreseen by the people and their leaders at the war's beginning, realization of the extraordinary contributions which will be required comes slowly, and the war is generally well along, sometimes approaching its close, before the maximum financial effort of which the people are capable is exerted. Often a weak or un-

intelligent leadership purposely avoids calling upon the people for heavy tax contributions. For these reasons war generally causes large government loans. The American Revolution was almost entirely financed by loans. The cost of the Civil War was about three billion dollars, of which four-fifths was borrowed, the balance being met mainly by taxation. The most gigantic loans in the history of public finance were occasioned by the World War. The United States borrowed about 24 billion dollars, or three-fourths of the cost imposed by the war. Great Britain expended 42 billion dollars on account of the war, obtaining 36 billions by means of borrowing. Germany, under the guidance of a weak and mistaken leadership, borrowed almost the entire amount of her war expenditure.

Local governments and members of federal governments, like the states of the United States, do not have to finance wars, and since they are seldom confronted with other serious crises, emergency borrowing is comparatively rare with them.

When a government engages in an industrial enterprise of any magnitude it usually finds it necessary to borrow the capital required. When the Prussian government decided to take over the railway system, it borrowed the money to purchase the existing railroads from their private owners. The United States government borrowed in order to construct the Panama Canal. The state of New York borrowed large sums to enlarge and improve the Erie Canal. American cities occasionally borrow in order to establish water works systems and other municipal industries. Governments also borrow, not infrequently, the money necessary to pay the cost of buildings and other public works needed for the performance of their primary functions. States and towns frequently borrow to build hospitals, asylums, and bridges; towns and cities, to build school houses and other public buildings.

The security back of public debts. Government debts differ from ordinary private obligations in that there is usually no pledge of security to guarantee payment. The exception

to this statement relates to weak governments with doubtful credit which, when borrowing from foreign capitalists, are sometimes required to pledge certain public property or certain sources of revenue. Such pledges are however the exception, not the rule, and in the case of a strong nation no pledge whatever is ordinarily given beyond the word of the government. In case of default of the ordinary public debt, the creditor has no recourse except either to bring suit in the debtor government's own courts — which is not of much help if the government has deliberately chosen not to pay — or, if a foreigner, to seek to persuade his own government to take diplomatic or, in the last resort, military action against the defaulting government.

High standing of government obligations. It may appear strange, in the face of the foregoing, that the unsecured debts of sovereign powers generally enjoy such high standing. As a matter of fact, government obligations, especially those of strong national governments, which are almost invariably unsecured, are ordinarily held in high favor by the investing public. Under normal conditions governments can generally borrow at lower rates than are paid by other borrowers, and government obligations are generally quoted on the markets at higher prices in proportion to their yield than any other investments. Bonds of respectable governments stand in the very front rank of "gilt-edged" securities.

The high standing of public obligations rests on certain characteristics of government. In the first place, the life of a government is not limited as is the life of an individual, and it is considered more enduring even than the lives of most corporations. Of chief significance however is the fact that the government's ability to meet its obligations, both interest and principal, is not dependent upon the success of any particular business venture. Most borrowing by individuals and corporations is for investment in business. If the business fails the borrowers may default on the interest or on part or the whole of the principal. The government's ability to meet

its obligations rests upon the taxing power, which may always be exerted to obtain the necessary money provided the debt is not too large. The absence of security is generally counteracted by a most scrupulous regard for its financial obligations, by which most governments seek to protect their credit.

It must be confessed that the history of the past generation has tended somewhat to impair the high standing of government obligations. The World War placed upon certain nations a financial strain which was more than they could bear. As a result of budget deficits and reckless issue of paper money, Russia, Germany, and certain other great nations virtually repudiated their earlier public debts. Other nations saw their obligations sink far below par and suffered serious loss of credit; among these were France and Italy. Great Britain and the United States are the two conspicuous examples of great nations which came through that struggle without serious loss of national credit. Since then monetary depreciation has lowered the value of public debts in many nations. The action of the United States in 1933 in simultaneously repudiating the gold contract in its national bonds and depreciating the value of the paper money in which alone the bonds may now be redeemed was not favorable to the credit of the United States government. The very size of the debts of some of the leading nations — including the United States — is an element of weakness. Government obligations, though still standing high, do not generally command that confidence which they held a generation ago.

The national debt of the United States. The War of the Revolution was financed principally by loans, and when the new national government of the United States started operations in 1789 it had a heritage of 52 millions of public debt, a sum which appears trivial enough to us of the present generation, accustomed to speak casually in terms of billions, but which the people of that day regarded as a heavy burden upon the shoulders of the new government.

Nevertheless one of the first acts of the government, under

the far-sighted leadership of Alexander Hamilton, was the assumption, as an addition to the national debt, of the debts of the several states incurred in the prosecution of the Revolutionary War. This brought the debt up to 76 millions in 1794. American public opinion, from the very beginning, has looked with disfavor upon a large national debt and has always favored rapid payment. By 1812 the debt of the United States had been reduced to 45 millions. The War of 1812, likewise financed principally by loans, brought the national debt up to 128 millions. Then commenced a period of rapid debt payment, culminating in 1836 in the complete extinguishing of the national debt except for a few thousand dollars, which were unpaid because the holders had not presented the bonds, and for whose payment the Treasury had the cash in hand. This represents a unique performance in the records of public finance — a national debt completely paid off.

The country was not long out of debt. Deficiencies in current revenues and the Mexican War restored the debt, which just before the Civil War in 1860 amounted to 65 million dollars. The borrowing of this war brought the national debt to 2,845 million dollars on August 31, 1865. Once more the nation set about the task of debt payment and, by a performance only less remarkable than its previous record, had brought the debt in 1893 down to 839 millions, a reduction of more than two-thirds. Since then there have been numerous debt operations, increases as well as reductions, the principal causes of new borrowing being the resumption of specie payment of the greenbacks in 1879, the currency disorders of 1893–96, the financing of the Spanish War, and the digging of the Panama Canal. In 1916, the year before the United States entered the World War, the national debt amounted to about a billion and a quarter. The World War was financed to the extent of three-fourths of its cost to the United States by borrowing, principally by the issue of the Liberty loans, Treasury notes, certificates of indebtedness, and war savings certificates. The total borrowing on account of this war

makes everything preceding pale into relative insignificance; on August 31, 1919, the national debt amounted to $26,596,701,648. Once again the traditional American policy of debt reduction was vigorously pursued. Large appropriations for payment of the debt were included in the regular annual budgets; in addition substantial surplus revenues were regularly available for debt retirement, and the debt by December 31, 1930, had been reduced to 16,026 millions, a reduction of two-fifths of the debt in eleven years. The subsequent growth of the national debt will be noted in connection with our examination of "New Deal" finance later in this chapter.

The budget. Under sovereign constitutional governments all taxes and other revenues and all expenditures must be authorized by law. This requires that at certain stated intervals the law-making body — be it Parliament, Congress, state legislature, or other law-making authority — must take account of the state of the public finances, determine what services shall be performed by the government, estimate the probable cost of such services, and provide taxes and other income to meet the anticipated expenditures. In determining its fiscal policy, the legislature will find it helpful to take account of the income and expenditure of the current and immediately preceding years. Such account, together with the estimates and plans for the coming year, is called the *budget.*

The subordinate grades of government — counties, cities, towns, school districts, etc. — also have use for the budget. Although they have no sovereign power, they have been delegated a considerable authority over their own finances, under which the local legislative body — board of aldermen, city council, town selectmen, or whatever it may be — determines the expenditures and at least the rates of taxation for the district.

Budget procedure has been brought to a high degree of perfection by certain governments, the most notable being the British Parliament. In the United States the budget has

been generally more honored in the breach than in the observance by all grades of government. Its efficient development has been hindered in the federal government by the constitutional separation between the executive and legislative arms of government and the bicameral Congress, features which have been generally carried down into the political organization of the states; and there have been other contributing causes of the poor record in America. Only since the World War has the federal government developed any semblance of a budgetary procedure, and, barring a few notable exceptions, the states and municipalities have been no less backward. Even now the United States budget falls far short of the standards set either by theory or by the practice of such nations as Great Britain.

American financial policy during the past decade (1929–1939): General character. The crisis of 1929 and the succeeding period of depression witnessed a governmental program in the United States which embraces many features of interest to the student of economics. Certain of these projects are examined at various places in this book in accordance with their subject matter. The present appears to be the proper place to investigate the important financial aspects of the "New Deal," as this program came to be called.

In its attitude toward finance, the philosophy of the "New Deal" presented two or three faces, distinct and not free from contradiction. First and most obvious, the operation of the several parts of the program occasioned the expenditure of enormous sums of money. A reckoning of this account appears later in this chapter. On the other hand, there was apparent an inclination to regard spending as something more than defraying the cost of desirable public functions. Government spending was considered desirable in itself, as a means of "putting money in circulation," stimulating purchasing power, "priming the pump," and restoring normal industrial activity. From this viewpoint, all government spending carried a favorable presumption, tending to brush

aside traditional notions of economy and to relegate to the background consideration of the balanced budget.

The government was, at the start at least, not wholly unmindful of economy. One of the first acts of the Congress that assembled in special session in March, 1933, was the enactment of "the economy act," restricting drastically the payment of veterans' pensions and benefits, and reducing the salaries of the employees of the federal government. Even such modest service to economy was however not permitted by Congress to continue more than a year. In March, 1934, over the veto of President Roosevelt, Congress restored three-fourths of the reduction in veterans' compensation made by the act of 1933 and also dictated gradual restoration within the coming year of the 15 per cent pay cut which had been made in federal salaries. In 1936 Congress granted another extraordinary veterans' bonus, which added $1\frac{1}{3}$ billions to the expenditures of the fiscal year 1936 and another half billion in 1937.

The double budget. Inconsistency between the claims of economy and the lavish expenditures of the "New Deal" was officially disavowed through the device of the double budget, by which means the administration sought to place special emphasis upon the "emergency" character of the "New Deal" expenditures, in distinction from those expenditures which might be presumed sufficient to meet the ordinary needs of the government. The enormous deficit in government operations was after the fiscal year 1933 mostly relegated to the "emergency budget," while the administration continued to proclaim its interest in economy and the preservation of the public credit. Of course the double budget has no magic power to banish a deficit. The deficit, as disclosed by the total account of revenues and cost payments or by the mounting public debt, goes on unhindered. In its basic implication moreover the "New Deal" budget was from the start misleading, through inclusion in the emergency budget of considerable expenditure which was destined —

and even intended — to be permanent, thus making the balance between revenues and "general" expenditures essentially fictitious.

Expenditures, revenues, and deficits. In order to visualize clearly the net effect of depression and "New Deal" upon the income account of the United States government, we may profitably make our start at the beginning of the fiscal year 1929, thus including for the sake of comparison the last two years in which prosperity still smiled upon the National Treasury. The table opposite shows the revenue receipts and the cost payments (*i.e.*, not including the proceeds of borrowing or sums expended in redemption of debt) for thirteen years, the figures for 1940 and 1941 being the official budget estimates. The total costs have been separated into "general" costs and expenditures for "recovery and relief." [1] These figures, especially those for "recovery and relief" costs, have been subject to frequent official revisions, which are retroactive. The table on page 527 agrees with the Finance Reports of 1938 and 1939 and the Budget Message of 1941. Owing to changes in the official classifications, it is not possible to give "recovery and relief" estimates for 1940 and 1941 corresponding to the figures for the years 1932–39.

Attention should be called to the fact that the expenditures here shown are less than they usually appear in the official statements (general and special accounts), owing to the exclusion of sums (averaging a little less than half a billion dollars a year through 1936 and about 90 millions in the last five years) expended for reduction of debt; such items are not true cost payments. It should also be noted that the receipts and payments for 1934, as here shown, do not take account of $2,811,375,757 "increment resulting from reduction in the weight of the gold dollar," of which 2 billions were devoted to the "exchange stabilization fund." Transactions

[1] For facts such as these relating to the finances of the United States government, the reader is referred to *Annual Report of the Secretary of the Treasury on the State of the Finances* (known as the "Finance Report"), *Budget of the United States Government* (annual), and *Daily Statement of the United States Treasury*.

REVENUE RECEIPTS AND COST PAYMENTS, UNITED STATES GOVERNMENT
(In millions of dollars)

Year	Revenue receipts	Cost payments			Surplus (+) or Deficit (−)
		General	Recovery and Relief	Total	
1929	4,033			3,299	+ 734
1930	4,178			3,440	+ 738
1931	3,190			3,652	− 462
1932	2,006	3,910	625	4,535	−2,529
1933	2,080	3,478	385	3,864	−1,784
1934	3,116	3,248	2,763	6,011	−2,896
1935	3,800	3,641	3,369	7,010	−3,209
1936	4,116	5,224	3,441	8,666	−4,550
1937	5,294	5,369	3,073	8,442	−3,149
1938	6,242	5,388	2,238	7,626	−1,384
1939	5,668	6,105	3,105	9,210	−3,542
1940	5,704			9,637	−3,933
1941	6,151			9,027	−2,876*

* In the Budget Message of 1941 this deficit is reduced to 2,176 millions by subtracting 700 million of surplus funds of government corporations, although this item is not added to the revenue estimate.

in the gold increment and the stabilization fund (comparatively small in amount) are excluded likewise from the figures for later years.[1]

Efforts to fight depression by public spending were undertaken by the Hoover Administration, with consequences as shown by the figures of 1931, 1932, and 1933. Late in the fiscal year 1933 the expenditures of the "New Deal" proper began to have their effect. It will be noted that total expenditures rose to about 8½ billions in 1936 and then declined slightly in the next two years. The heavy increases for 1939, 1940, and 1941 are consequent upon the "pump-priming" policy employed to meet the return of depression conditions in the fall of 1937.

On the revenue side of the account, the blighting effect of depression is clearly discernible in the decline of receipts in 1932 and 1933 to only about half the amounts received in

[1] Trust fund transactions are, properly, excluded from the figures for the years 1931–41. Owing to the form of the official reports, it is not possible to exclude these transactions from the figures for 1929 and 1930. The differences are not sufficient to affect the comparisons here made.

1929 and 1930. Thereafter revenue legislation was enacted in each year from 1934 to 1939. These acts increased heavily the rates of the income tax, both individual and corporate, and the estate tax. The excise system was strengthened by broadening its base and raising rates. These measures, together with improvement in general economic conditions, brought about a progressive increase in revenues through the fiscal year 1938. The decline in revenues thereafter reflects the depressed condition of national industry which prevailed from the fall of 1937 to the middle of the calendar year 1939.

In spite of the enactment of such revenue legislation, the fact is that no determined effort was made to meet expenditures by corresponding increase in taxation. Income taxes, especially with rates as high as they were, offer little elasticity in time of depression. And the excise system was never called upon for a contribution commensurate with the increase in expenditures. In fact the revenue acts of the period were devised with as much regard to certain other ends as to their revenue producing effects. The revenue act of May, 1934, for example, was announced as aimed chiefly at "tax avoidance" and was estimated as capable of adding only 417 million dollars to the annual income when fully in operation. Similarly the act of 1935 was intended primarily as a means toward the "redistribution of wealth"; the President in January, 1936, estimated its yield at only 222 million dollars a year. These sums were obviously only small fractions of the current increases in expenditure. The revenue acts of 1937, 1938, and 1939 sought chiefly amendment in details and improvement of administration and disclaimed any intention to increase materially the demands of taxation. It was clearly the policy of the administration generally to avoid any drastic increase in taxation and to rely chiefly upon credit for meeting the "emergency" expenditures.

The handsome surpluses of 1929 and 1930 are typical of the period following the World War, during which the National Treasury enjoyed great prosperity. In the fiscal year

1931 there began a series of heavy deficits, due for the first three years in part to anti-depression expenditures but principally to declining revenues, and mainly chargeable to "New Deal" costs in 1934 and thereafter. The fiscal year 1938 came nearer to a balance than any previous year since 1931, but 1939 shows a deficit greater than in any year except 1936, and the estimated deficit for 1940 is even higher.

The gross debt of the United States stood at its then maximum of 26,597 million dollars on August 31, 1919. It reached its low point since the World War on December 31, 1930, when the debt amounted to 16,026 million dollars. The deficits of the depression years reversed the progress of debt payment, and by June 30, 1934, all of the gain since the World War had been lost. Succeeding deficits brought the debt to the unprecedented total of 42,365 million dollars on March 15, 1940. Thus have been broken all records for the United States, with the national debt exceeding by almost 16 billions its maximum height in 1919. It appears certain that the debt will continue to increase for at least another year. Official estimates indicate a total of 44,939 millions on June 30, 1941. Indeed it is probable that the debt, as technically calculated, will before that date reach the legal debt limit of 45 billions.

Progressive increase in cost of government. After every period of extraordinarily swollen public expenditures, the United States has found itself upon a new permanent level of governmental costs far above that which formerly prevailed. To the time of the Civil War, the annual per capita cost of the United States government was never much above two dollars; following the Civil War it never fell below four dollars. In the fiscal year 1916, the total ordinary expenditures of the national government were 734 million dollars; since the World War they have never been below 3 billions, except in 1927 (when the amount was 2,974 millions). The charge for interest alone in 1920 (1,020 million dollars) exceeded the entire cost of the government in 1916 or any previous year with the single exception of 1865.

That the aftermath of the existing régime of swollen expenditures and deficit financing will run counter to past experience is hardly to be expected. With a debt greater than ever before, the government will be bearing a charge for interest alone well over a billion dollars. Assuming that the American people adhere to their traditional policy of debt reduction, there will be an appropriation for debt payment each year which may be as much as half a billion (on the precedent of the years following the World War). That the extraordinarily expanded governmental machine which now exists can ever be reduced to its pre-depression dimensions appears most unlikely in the light of past experience. The year 1939 showed the highest level of expenditure ever experienced by the federal government in time of peace, and at the present writing (March, 1940) the year 1940 seems destined to go half a billion higher.

CHAPTER XXVIII

The Business Cycle

Business fluctuations: In general. In a developing social order economic forces do not remain for long in a state of stable equilibrium. Changes are constantly occurring in some branches of activity which do not instantly harmonize with corresponding changes in other branches. The tendency for capital and labor to be distributed among the different productive enterprises so that their marginal productivities are everywhere the same is thwarted by unforeseen changes of fashion, by invention, by political disturbances and similar events. The price level may be shifted by the forces of change, and the individual prices within the system which were formerly in approximate balance with each other may move erratically, sometimes even in opposite directions. These are random examples of economic alterations which tend to keep the productive system in a state of disequilibrium and uncertainty.

It is impossible of course to analyze each of the countless changes in economic activity which characterize our social order, but there are some major aspects of this problem which do repay study. When we examine the course of economic activity over a period of years we find that the changes are not altogether sporadic and erratic. On the contrary, there are certain well-marked fluctuations affecting the productive system in its entirety. Business in general is more active at certain times than at others. The total volume of production is seen to increase during one period of years, whether this be measured in absolute quantities or in value terms, and to fall during a subsequent period. Sometimes there are jobs aplenty; later there is widespread unemployment. Boom times alternate with times of business depression. These general fluctuations

of economic activity must be the result of forces which play upon the productive system in a comprehensive manner. It is our purpose in the present chapter to analyze these fluctuations and so far as possible to ascertain their causes.

Types of economic fluctuations. When we speak of an economic fluctuation we mean a change of business activity on a fairly broad scale which can be represented in statistical terms. The statistical index chosen to represent the change may be selected from among a number available. For example we may chart the volume of agricultural and factory production over a period of time, or the amount of employment of wage earners, or the average price level, or the rate of return on invested funds, and find definite evidence of a general movement upwards or downwards as the case may be. In the following pages we shall use such indexes as these as illustrations to aid us in the analysis of the problem of economic fluctuations.

Economic fluctuations may be classified in three main categories: (1) independent irregularities; (2) cycles; (3) the secular trend. Of these types of fluctuation, the second or cyclical type is by far the most important, both as a matter of practical concern to mankind and as a problem of economic theory. In this chapter therefore we shall give our attention almost exclusively to the cycle, pausing only to note the existence of the other two forms of fluctuation.

Independent irregularities. This term is used to designate such disturbances in the normal development of economic enterprise as result from isolated events, such as widespread strikes, national disasters, or popular reactions to changes in governmental policy. These disturbances are usually short-lived and self-correcting, but their effects appear on any long-run chart of business activity as irregularities not to be accounted for by reference to the major forces of change which are at work.

Cyclical fluctuations. As the term indicates, a cyclical fluctuation is one which discloses a recurring, rhythmical

character. During a complete cycle business activity will pass through a phase of acceleration followed by a second phase of decline. In a series of cycles this process will repeat itself over and over again, though not of course in exactly the same statistical terms. Long-range studies of business trends have brought to light not one but many different cyclical movements which seem to be characteristic of our economic order. These cycles may be classified in four important groups: (1) seasonal cycles, (2) long-wave movements, (3) minor non-seasonal cycles, (4) major non-seasonal cycles.

Seasonal cycles. These are the simplest and most readily identified of the business cycles. They result in part from market response to the changing climatic seasons, in part from the conventions of our calendar. The increased business activity which corresponds with seed time and harvest in an agrarian country is an example. So also is the business rush during the Christmas season as contrasted with the slack markets of midsummer.

Figure 27 (on page 536) has been prepared primarily to illustrate other types of cycles. However reference to it will provide the student with evidence of the seasonal fluctuation of one index to business activity; *i.e.*, the trend of factory employment and payrolls. Disregarding changes which occur over a period of several years, it will be seen that each year considered by itself records a definite seasonal variation. The line indicating payrolls, for example, tends to reach a high point at two different times within the year and again to fall to a low point at two other times. This is more clearly shown in some years than in others — in the year 1926 for example. Still the tendency can be detected in every twelve-month period covered by the chart.

Seasonal fluctuations do not present a major social problem. It is true that certain specific industries — such as the manufacture of clothing, coal mining, the building trades — may be seriously affected by them. In such industries the problem is one of stabilization through improved business or-

ganization and management. In society generally the existence of seasonal fluctuations is clearly recognized, and economic activity as a whole is fairly well adjusted to them.

Long-wave movements. There is general recognition of an important group of cycles, somewhere between forty and sixty years in duration, which are usually designated as long waves. Their presence has been detected in the trend of

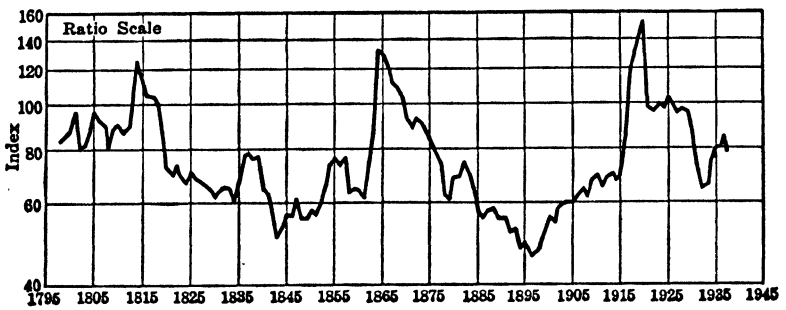


FIG. 26. Index of Wholesale Prices, All Commodities, United States, 1795–1938.[1] (1926 = 100)

price data and production data for the United States, England, and France. Figure 26 reveals the character of the long waves in wholesale prices in the United States, with low points in 1843 and 1896 and high points in 1814, 1864, and 1920. It is important to note that this chart is drawn to a *ratio scale* and hence shows proportional changes rather than absolute rates of change.[2] An analysis of long-wave movements falls outside the scope of the present discussion, but we may note

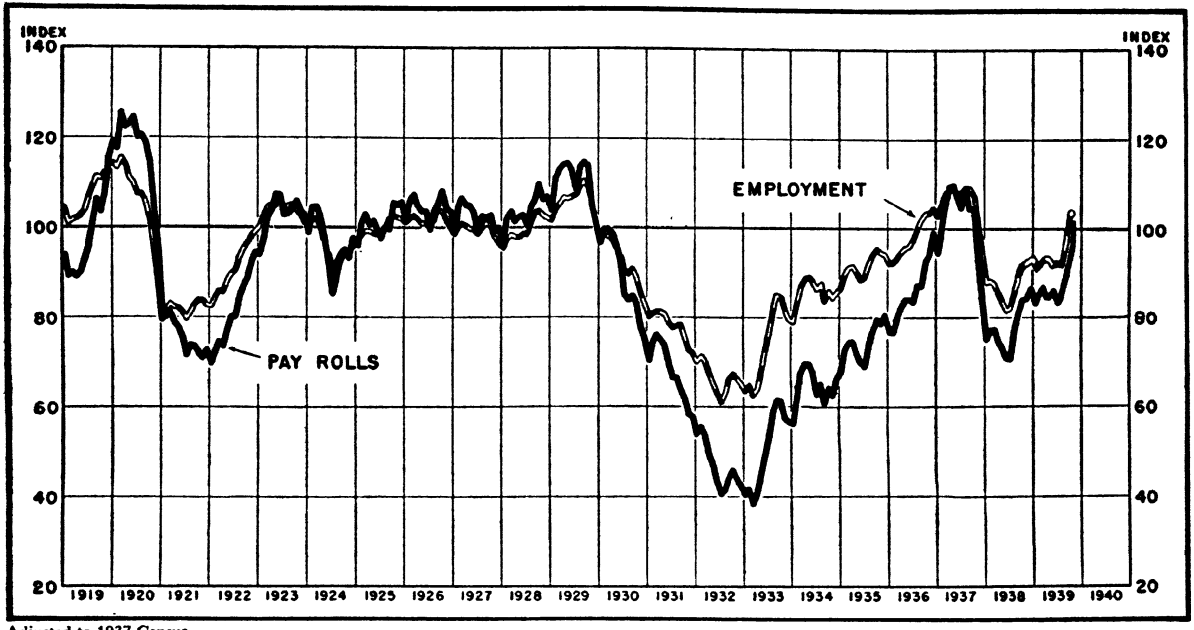
[1] Warren and Pearson index to 1889; United States Bureau of Labor Statistics from 1890.

[2] The ratio scale is drawn so that a given vertical distance always represents the same proportional change, rather than the same absolute amount of change (160 is the same distance above 80 as 80 is above 40 and as 120 is above 60, etc.). Lines drawn against a ratio scale show directly rates of change; a line continuing straight showing a constant rate of change; a line concave to the horizontal axis shows a declining rate of increase or an acceleration of rate of decrease, depending on the direction of movement, while a line convex to the horizontal axis shows the reverse; a line rising or falling more steeply than another shows a faster rate of increase or decrease respectively.

that the underlying causes of these long waves have much in common with those of business cycles proper and that much of the ensuing discussion of the causes of the business cycle may be extended to explain this slower fluctuation in business activity.

Major and minor cycles. The term *business cycle* is ordinarily used to refer to fluctuations in employment and output which occur with an average duration of approximately twelve years, but this description is not quite adequate because the experience of the past twenty or thirty years reveals that there are in fact two business cycles — a major cycle and a minor cycle. Figure 27 (on page 536) indicates that factory employment and factory payrolls rose from a low point in 1919 to a high point in 1920 and then dropped to a new low point at the end of 1921. Beginning here the process repeats itself, another high point being reached in 1923 and another low point in 1924 — then a slow rise to a peak in 1929 and an abrupt fall to a very low point in 1933. This is a record of a minor cycle, in this case varying in duration from two and one-half to five years. Another glance at the chart will expose a longer rhythmical swing, particularly with regard to the points of deepest depression. An extremely low point was reached in 1921, and another in 1933; this is a major cycle embracing a twelve-year period. The high points of 1920 and 1929 illustrate the opposite phase of the major cycle, in this case a period of nine years. The chart we are using covers too brief a period to show these major cycles to advantage; but practically all long-term indexes of economic activity reveal major cycles, recurring in periods of from seven to twelve years, and running concomitantly with briefer minor cycles.

It is impossible to say which of these two types of cycle will be the more important in the future, but in all probability the longer business cycle of seven to twelve years' duration is the one that deserves closer study. The reader should realize however that business cycle analysis and any discussion of



Adjusted to 1937 Census

FIG. 27. Employment and Payrolls, All Manufacturing Industries.[1] (1923–25 = 100)[1]

[1] United States Bureau of Labor Statistics, *Employment and Payrolls*, April, 1939, p. 17.

the causes of these fluctuations in employment may be applied to the explanation of either the shorter or the longer variation.

The secular trend. One other type of variation must be mentioned before we discuss in detail the business cycle proper. Even in the absence of cyclical fluctuations our productive system would not remain entirely static. There are forces of

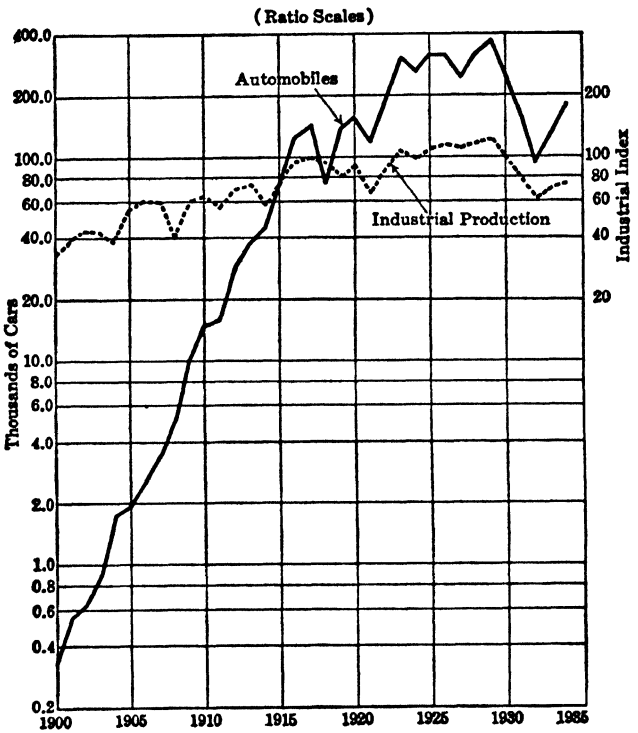


FIG. 28. Average Monthly Production of Passenger Automobiles and General Index of Industrial Production, United States, 1900–34.

growth which would carry economic development in a given direction, even though the different branches remained in equilibrium with each other during the process. The term secular trend refers to this normal line of balanced movement. In the broadest meaning of the term this trend of development may be either upwards or on a uniform level or downwards,

and specific industries can be found which illustrate each of these trends during any period of time of reasonable length.

If students of cyclical change do not take account of secular trend, their statistical image of the cycle may be badly distorted. This may be illustrated by reference to Figure 28. It will be seen that the trend of automobile production during the period 1900–19 was almost steadily upward, although the chart of industrial production as a whole shows clearly defined cyclical movements during this period. Obviously a statistical study of any group of industries in which automobiles formed an important part would be strongly affected during these years by the secular trend of this specific industry. The important point for our purpose is to note that if the figures are to be a faithful measure of cyclical change alone, they must be corrected for secular trend; otherwise they will reflect not only the effects of the cycle but other forces of growth as well. It is our intention merely to make this clear, not to discuss the methods of measuring secular trend and eliminating its influence on statistical series. In the discussion which follows and in Figure 29, which illustrates the business cycle, it may be assumed that the necessary correction for secular trend has been made.

The business cycle: General description. Each of the various types of economic fluctuation which we have mentioned exerts its own influence at any given time upon business activity. Furthermore none of these variations affects all departments of business activity in precisely the same manner. Different strands of business activity may vary during a given cycle by as much as a year in the times at which they reach the peak of the cycle and the trough of a given depression. When dealing with these cycles therefore we have to do with very complex phenomena, a large number of movements in economic life, which are partly independent of each other and partly interacting.

Despite this complexity, it has become customary to speak of "the business cycle" as if a single general movement took

place over a term of years affecting in the same direction the whole range of economic activity. Such movements, if they exist, must result from the interaction of all the major and minor cyclical tendencies operating in the period. That the "business cycle" in this comprehensive sense is an actual phenomenon will become clear upon examination of Figure 29 (on page 540). This is an index of general business activity during the past sixty years. It will be seen that, apart from interim disturbances, the trend of business activity as a whole, resulting from complex and diverse cyclical movements, describes a definite pattern within certain well-marked periods of fluctuation: 1878–85; 1885–97; 1897–1908; 1908–15; 1915–21; 1921–33. Although the existence of different cyclical movements within the business cycles must be kept in mind, there is practical justification for taking a simplified view of the problem of fluctuations such as is implied in this generalized concept of the cycle. For these are the variations of business activity which are really felt. They contain the major features of economic disturbance which present business men as a group with weighty problems and throw on society the strain of widespread maladjustment.

Phases of the business cycle. When we compare the series of cycles pictured in Figure 29, we find that they were in many respects dissimilar. For one thing, they were not of the same duration, as one can see by comparing the period 1878–85 with the period 1885–97 and so on throughout the chart. In other words, no standard period of years can be designated as the normal duration of a business cycle. Nor did these cycles correspond in degree of fluctuation; there was greater fluctuation in the 1885–97 period, for example, than in the 1908–15 period. But despite these diversities there still remains a recognizable pattern of similarity. They all passed through four rather well-marked phases, to which it is customary to give the names, revival or recovery, prosperity, recession, depression. The revival phase is taken as the starting point of the cycle; for example, the 1878–85 period and

all subsequent periods begin when the indexes of business activity turn upwards. The prosperity phase is marked by the later stages of this upward trend. Recession is defined by the downward slump of the curve. Depression is indicated by the subsequent low points on the curve, continuing until another revival sets in. With minor differences in detail this general pattern is seen to apply to all business cycles. It is now generally believed that all these successive phases of the cycle, except the first, are self-generating. The revival phase develops through its own momentum into the prosperity phase;

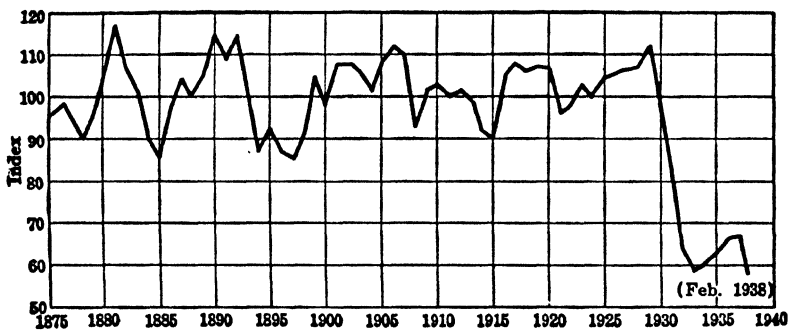


FIG. 29. Index of General Business Activity, United States, 1875–1938.[1]

the prosperity phase contains within itself forces which initiate recession; recession in turn develops through its own characteristics into depression. The phase of revival is an exception to this rule. Depressions are not self-limiting; they seem incapable of moving into a new revival without the impact of some external force.

The conditions prevailing in the successive phases of a cycle may be roughly summarized. In revival we find rising prices of securities and commodities, a rising volume and value of industrial production, increasing sales and forward orders in most lines of trade, increasing employment, mounting loans or debts, and a rising rate of interest. In prosperity we find

[1] Clearings index of the Federal Reserve Bank of New York, with adjustment for secular trend, seasonal variation, and changes in the general level of prices.

prices continuing to rise but gradually levelling off, large accumulated stocks of goods, high but stabilized employment, an intensified demand for credit accompanied by high rates of interest, and the production of capital goods beginning to weaken. In recession the business picture is one of decline in sales, orders, production, employment, and prices, an increasing real burden of debt as the amount of debt is reduced more slowly than the decline in the value of assets, and a falling rate of interest. In the depression phase debts and stocks of goods have been reduced; prices run at low levels, though security prices may show some slight return of strength; the rate of production is low; unemployment is widespread; interest rates are low. The method of development of these conditions is the important problem for our consideration.

Revival. The most significant feature of the revival, which we take as the starting point of the business cycle, is an increase in spending. Ordinarily we find that business spending increases first and is followed subsequently by an enlargement of consumer spending. Economic science has as yet no complete answer to the question as to what the forces are which induce this increase in spending at a time when business is depressed and unemployment is widespread. There are however some points on which there is general agreement.

In the first place, depression itself to some extent serves to prepare the way for a revival. Weak firms and weak banks which have survived the first months of the depression will eventually be forced through bankruptcy and dissolution or reorganization, and the surviving firms, with assets written down and capital structures reorganized, will be better credit risks and so in better position to secure funds for spending whenever business spending appears to be warranted.

Furthermore the continued deflationary pressure of the depression itself is almost sure to bring about a reduction in costs. During a depression wage rates, interest rates, and many of the prices of producers' goods will have been reduced. Not only are costs, both of current production and of new invest-

ment, lowered in this fashion, but the less efficient members of any working force and the less efficient producers in any given industry are likely to be the ones first thrown out of employment. The favorable effect of these reductions in cost should not of course be exaggerated. For one thing, it is quite clear that most reductions in costs bring about equivalent reductions in the incomes of other members of the economic system, so that unfavorable effects on purchasing power may tend to counteract the favorable effects of cost reductions. Nevertheless if the depression has the effect of decreasing certain sorts of costs that have been out of line with the general level of prices and incomes of the community, it may make the prospects of future profits more favorable and thus encourage increased business spending.

But more important is the fact that, if business is to maintain even a low rate of production, the time will come when it will require new machinery and equipment to replace that which has worn out. During a period of depression it is the practice of business firms to reduce expenditures for maintenance of plant to the minimum by postponing replacements of worn-out equipment, but this process of using up capital goods cannot continue indefinitely; at some time increased spending for this purpose will be inevitable.

All three of the causes of revival so far discussed are in some degree automatic in their operation, and they have considerable significance. But any history of the business cycle makes it clear that the most powerful forces working to initiate revival are ones which do not develop automatically out of the preceding phase of the business cycle but are in a sense fortuitous. Of these the most important is the appearance of new investment opportunities.

Historically the classic example during the nineteenth century was the development of railroads throughout the world. In the United States, up to 1890 at least, this provided the largest single outlet for investment funds. The existence of such investment outlets affected not only the industries directly

and indirectly concerned with the railroads themselves; the high rate of investment in railroads made for a high national income and thus stimulated investment in many other lines. A more recent example is the automobile industry in the United States, which during the decade of 1920–29 grew at a prodigious rate. Though founded many years previously, the industry had evolved by 1921 to the point where it presented an ideal opportunity for the investment of capital on a large scale. This investment opportunity and the shortage of housing that had developed during the war were largely responsible for the quick ending of the 1921 depression.

Investment opportunities may take several forms. Though perhaps the most important ones occur when a new industry or a new area begins to develop, much investment is undertaken by existing firms in existing industries and without reference to the opening up of new territories anywhere in the world. Such investment frequently occurs as the result of new and cheaper methods of production, such as are being discovered almost continuously. After a depression has been in progress for several years, even if no major new investment opportunity appears, replacement of old equipment and expansion of plant are eventually forced upon business men both by the wearing out of what they already had and by the availability of new kinds of machinery and new productive methods.

Revival breeds prosperity. Once a revival is under way it tends to become cumulative, feeding on itself for a considerable period at least. It is not difficult to see why this is so. As new construction projects are launched and carried through, men will be employed in building factories, installing machinery, and other such operations. Part of the increase in expenditure will thus be matched by increases in wage payments. Many of the new expenditures will go toward the purchase of building materials, machinery, tools, etc., and the business firms which supply capital equipment will thus find the demand for their goods increased and their incomes

enlarged. They in turn will find it necessary to place orders for raw materials, to engage larger laboring forces, and to revise their production schedules upward.

Consumers who have benefited from the primary increases in income will of course increase their expenditures for consumers' goods, and the demand for these goods will be swelled still further as the larger consumer incomes generated by renewed activity in the capital goods industries manifest themselves. As wage earners come to feel that their jobs are secure and their incomes likely to increase, they incline to expand their expenditures beyond their immediate means and so resort to borrowing in one form or another. This factor is of great importance for those industries which produce durable consumers' goods, such as automobiles, refrigerators, washing machines, and the like, and especially in the field of housing. During the decade of 1920–29 consumer spending for the purchase of new houses was of very large dimensions and was largely financed by borrowing from the banks. This substantial increase in the demand for consumers' goods of course gives rise to increased production and employment in the consumer goods industries and still further reinforces the trend toward increased business activity.

Furthermore investment in stocks, both of raw materials and of finished goods, is certain to be stimulated. During a depression business men almost invariably attempt to reduce their inventories to the lowest possible level, but once revival is under way this entire picture changes. The expectation that the volume of sales will continue to rise and the fear that a shortage may develop and that deliveries of goods may be slow will induce business men all along the line to build up much larger inventories than they have been carrying. Rising prices of goods of all sorts give them an added incentive, for even though they may be most anxious to avoid speculation in inventories, the expectation of a continued rise in prices will make it seem prudent to add substantially to the volume of goods carried in stock. The initial increase in the demand

for both producers' and consumers' goods tends thus to be magnified.

As recovery proceeds, and especially if it proceeds rapidly, the increasing demand for the use of the factors of production and for both consumers' and producers' goods will bring about higher prices. But it is characteristic of a period in which the general level of prices is rising that many of the prices which determine costs of production — such as wages of labor and prices of capital goods — lag behind the sale prices which determine business income. The result is a wider margin of profit, which serves as an added stimulus to expansion and employment and therefore to income.

In summary we may say that the initial increase in spending gives rise to a primary increase in income. But this in turn stimulates an increase in production over wide areas of the economic system, and the increase in production in turn gives rise to a secondary increase in income. Even there of course the process does not stop, for the enlarged output of capital goods gives rise to an increased demand for the raw materials which they require and hence to further increases in income. Thus the initial increase in spending spreads throughout the economic system, and with its spread the level of production rises, consumers' and business incomes are increased, further spending follows, and production is again stimulated.

Limiting factors in prosperity. What are the limits to this process? There would of course be none if each expansion of spending and income gave rise to a further expansion in spending of exactly the same size. For in that case the assumed initial increase in business spending would give rise to an infinite series of further expansions, each of the same size. But actually each successive expansion will generally be smaller than the one preceding it. It is to be presumed that consumers will plan to save at least some portion of the increments to their incomes, and it is highly probable that each increase in incomes will give rise to some stable increase in the rate of saving. This means that consumer spending will

normally increase less than consumers' incomes. Now this would not in itself argue a declining rate of increase in total spending if consumer savings were, as is normal, made available to business and spent for investment in new plant and the like. Similar considerations however apply to the expansion of business incomes and business spending. If a firm receives a rush of new orders, its disbursements in the form of wage payments and payments for raw materials will of course increase, but it is most unlikely to begin paying out higher dividends the moment higher profits begin to be earned. Business men, like consumers, will plan to save some part of the increment of their incomes in the form of liquid assets. This implies that, although total spending in any such period will be increasing, it will be increasing at a decreasing rate.

This fact will be reflected in the relative movement of prices and costs. After the initial upward surge, industrial prices tend to rise for a period at a decelerating rate. In the meantime some of the more sluggishly moving prices which are costs to industry, such as wages, interest, and the like, begin to catch up and to decrease the margin of profit. Increasing costs weaken the motive for further expansion of capital equipment, and the same effect comes from the tendency of increased capital resources to show a declining physical productivity at the margins of application. The expansion of the capital goods industries and of their disbursements is retarded, tending to a decrease in consumer incomes and a slowing down of the rate of increase in purchases by consumers and retail dealers. Prices of consumers' goods then show fluctuations around a temporarily stabilized level or even show a slight tendency to decline.

Recession and depression. At the culmination of a period of prosperity the developments we have traced finally interact so as to bring a reduction in the rate of prospective profit. The profit margin is pinched between a retarded rise of selling prices in general and an accelerated rise of costs. As the rate

of interest finally rises above the rate of prospective profit, the incentive to borrow for expansion is replaced by an incentive to curtail business spending and pay off loans. As soon as the industries producing capital goods find a definitely declining market, they add an absolute reduction of output, disbursements, and bank loans to the now rapidly changing picture of business. With consumer spending then further curtailed, a definite decline in industrial activity becomes general, and profit margins are further narrowed as business costs tend to fall less rapidly than other prices. Thus begins a spiral of falling values, forced sale of goods to permit required debt liquidation, consequent reduction of credit currency, further depression of values, increased unemployment, and a general decline in incomes and trade. A psychological depression results, and a reduced velocity in the turnover of currency augments the other factors. Declining prices of securities add further to business curtailments, partly by their psychological influence, partly because of their effect on the credit standing of business enterprises.

As trade approaches a minimum level and surplus stocks of goods are exhausted, as the burden of old debts is slowly removed, as the rate of interest falls, as costs of production fall into line with the general decline of prices, profit margins tend to increase enough to permit the use of funds more profitably than in paying off loans. Such new funds might be used for slight expansions here and there in the volume of operations or for meeting the requirements of postponed maintenance of capital equipment. A period of uneven but general equilibrium between costs and selling prices, between production and consumption, thus tends to prevail until events external to the conditions of depression occur to stimulate another revival.

Forces initiating the business cycle. Apparently the conditions internal to the business situation in a period of depression can at most set the stage for a possible revival, leaving it to external forces to set the cycle in motion. The normal

operation of the credit system exerts no original impulse upon the cycle; low interest rates, excess bank reserves, and superabundant supplies of credit proved quite ineffective in 1932 and 1933 to bring revival out of depression. Similarly the equilibrium of costs and income which obtains throughout the business world after recession has worked out its effects is only an invitation to accelerated activity. The spur of self-interest under such conditions certainly moves enterprisers constantly to seek opportunities to expand their businesses. These impulses may result in a degree of growth on a somewhat uniform level, but they do not appear to be effective for starting the upward accelerated surge of activity which is characteristic of a cycle in its first phase.

As a rule depressions have not ended until substantial new investment outlets have been discovered. These may take the form of government borrowing for war or for a program of armament, or the growth of a new industry, or the opening up of new territory, or a general advance on a number of fronts simultaneously. The necessity of making up arrears of depreciation is often a powerful contributing cause of the upturn and may even have a good deal to do with its exact timing, but as a rule we must look to some cause external to the operation of the economic system itself in order to explain the initiation and continuation of a prolonged upward movement.

Consequences of cyclical fluctuation. Generalizing from past experience of cycles, it may be said that business activity fluctuates as a rule between twenty-five per cent above and twenty-five per cent below its average level. This fluctuation is in addition to the changes produced by long waves from one generation to another, these long-range changes being interpreted as shifts in the average level about which the cycle oscillates. The percentage of fluctuation given does not apply of course to all industries and all occupational groups alike, for in any business cycle some interests are much more seriously affected than others.

These continued fluctuations of the economic system have many injurious social consequences. Most obvious is the general impoverishment which occurs at the depression points, a temporary phase for society as a whole, but permanently disastrous for multitudes of individuals. Over and above this the alternations between general plenty and general want are seriously disruptive to the social order. They upset the morale of the people and develop a fear psychology by introducing uncertainty and insecurity into their lives. Experience of recurring business depression, with its reduction of wage income and widespread unemployment, has its effect on the efficiency of wage labor. These wide fluctuations upset business leadership by thwarting the best laid plans for future economic development. They draw the various branches of industry out of adjustment to each other, thus diminishing the efficiency of individual enterprises and the balance of the entire economic structure. They cause positive waste of economic power, by inducing misapplication of resources during the speculative phases of prosperity periods and precipitating idleness of resources during the subsequent depression. Nothing need be said of the incalculable human misery caused by prolonged depression; this is too well known to require emphasis. The broader impersonal consequences of the cycle are sufficiently serious to justify the statement that cyclical fluctuations are the chief cause of many of our most grievous social problems.

These cyclical fluctuations of business are features of an economic system centred around the getting and spending of money incomes, a system in which groups of specialized productive processes serve uncertain markets under the guidance of prospective pecuniary profits. The conclusion does not follow that an economic system of this type — a business economy — should be replaced by some other form of organization. Historically the vicissitudes of non-business economies have been even more violent and humanly disastrous, although they have been episodic in character rather than

cyclical. The conclusion is not to destroy the business system with its cyclical disruptions but to seek measures that may control so far as possible the cyclical disturbances.

Control of cyclical changes: Conditions of the problem. Apparently an individualistic, competitive economic system, organized around prices, cannot automatically overcome the forces causing cyclical changes and adjust itself to these changes rapidly enough to avoid cumulative disequilibrium. On the other hand, the problem of applying external controls is by no means simple. There are two reasons for this.

In the first place, there is the question of what the objective of control should be. Since the crisis of 1929 the attention of most economists has been focused on the problem of bridging the gap between depression and revival, of finding methods of giving an impetus to business activity which would initiate the upward spiral of increased employment, output, and income. But there is also the problem of trying to restrain the fluctuations of business activity within narrower bounds. This implies the application of controls in the period of revival or during the early stages of prosperity in the hope of avoiding such an accumulation of internal strains as would make inevitable a lengthy and disastrous period of liquidation.

In the second place, there is not substantial agreement among economists as to the measures which are appropriate to achieve either of these objectives. This is not surprising in view of the fact that the two points in the analysis of the business cycle on which economists differ most widely have to do with the forces which bring about revival and those which turn prosperity into recession.

The measures of control which have been proposed are so numerous that we can only touch briefly on a few which seem promising. These proposals all encounter one common problem, since all require intelligent, disinterested, and courageous administration by some central authority. One's judgment as to the practical virtue of any proposal depends therefore not alone on the verdict as to the abstract soundness of the measure

but quite as much on his confidence in the governmental body which is to administer it.

Flexible prices. The suggestion has frequently been made that the control of cyclical fluctuations might properly take the negative form of securing greater flexibility in the economic system, especially with respect to the price structure. Complete flexibility in any system of human relations and humanly operated institutions is obviously impossible. But with a greater degree of flexibility of prices than now exists, cyclical variations would be less extreme; for it is the failure of prices to move in harmony which causes most of the trouble. When the price level is being shifted by uncontrollable forces, rigidity in any group of prices will throw an increased burden of readjustment upon the more flexible areas. Even with more flexible prices cyclical fluctuations would still be serious, and there would always be the tendency for the economic system to move back again toward greater rigidity. The self-interest of groups and individuals leads them to resist price changes, including changes in wages, which are immediately unfavorable to themselves. These resistance policies take very many forms: commodity price fixing, coöperative marketing, industrial concentration, agreements within trade associations, the gentlemanly following of a leader so as not to "spoil markets," a widening of the scope of regulation of public utilities. But despite the difficulties which confront any government in its attempts to prevent monopoly and restraint on trade and to maintain truly competitive and flexible prices, there is every reason for making strenuous effort to preserve those conditions which promote price flexibility all along the line.

Control of wage rates. Another suggestion which has frequently been made is that some authority exercise direct control over wage rates. If in the early stages of revival wages could be raised sufficiently to neutralize that particular tendency toward increasing the margin of profit, some of the incentive toward rapid and speculative expansion of business would be removed. In further defense of this policy it is also

argued that it would place increased purchasing power in the hands of labor and thus tend to maintain a better balance between consumption and investment. Such a plan has serious practical obstacles to face, for a rapid rise in wages in a period of revival would necessarily carry with it the stipulation that wages should be decreased in a period of falling prices when the profit margin was declining, and this would invite the determined hostility of labor. It is similarly certain that employers would not submit without a struggle to an arrangement which would thus curtail their profits. This struggle might take the form of a reclassification of the more highly paid labor, in order to place them in a lower wage group, or a greater resort to the use of capital. Despite these practical difficulties, the proposal to secure greater flexibility in the price of labor by positive measures of control is deserving of serious consideration.

The rate of interest. For many years economists have been interested in the possibility of controlling the rate of investment and by this means of exercising some measure of control over the business cycle. This is an instrument of control which lies ready to hand, for in all capitalistic societies the rate of interest is regulated to a considerable degree by action of the central bank. A policy of high discount rates designed to restrict lending by banks is recognized as frequently effective to discourage investment. It has seemed only natural to inquire whether or not a corresponding easy money policy, in the United States on the part of the federal reserve system let us say, would stimulate investment.

The rediscount rate, being the rate at which member banks may borrow from the federal reserve banks, has its effect upon the rates of interest charged by member banks to their customers. The direct effect, to be sure, is only upon a narrow group of interest rates. But there is available also the instrument of open-market purchases of government bonds, which either increase the excess reserves or diminish the indebtedness of the member banks to the federal reserve banks, thus putting

strong pressure upon member banks to find other outlets for the investment of their funds. A central bank therefore, if it is sufficiently resolute, can exercise not only a rather powerful and direct control over bank rates of interest, but can also exert pressure upon the banks which will affect long-term interest rates.

In view of these powers possessed by the central bank, it might seem that monetary control of the rate of investment should work effectively. The history of the last fifteen years however does not confirm this expectation, for reasons that are by now quite generally accepted. Although the interest rate does undoubtedly tend to affect the rate of investment in the manner just described, the influence exerted by the rate of interest does not appear to be sufficiently powerful to achieve the results expected of it. Particularly is it true that when a recession has once started, a lowering of the rate of interest is virtually powerless to prevent the recession from running its course. Low rates of interest, as they spread through the economic system, may do much to promote an up-turn if other factors are favorable. And the high rates of interest that prevail during the late stages of a boom are frequently sufficient of themselves to dampen the rate of investment and thus to precipitate a depression. It may be concluded then that control of the rate of interest might be used with some success to prevent a revival from progressing too rapidly or a period of prosperity from turning into an inflationary boom but that such control certainly will not be effective to halt a depression when once started, though it might later assist in bringing about a revival.

Public works. In recent years many governments have intervened directly in the economic lives of their communities in the effort to increase the rate of spending. This has been done in the democratic countries, not by enforcing rules of conduct upon individual and private enterprise, but rather by entering directly into the function of investment through the medium of public works and relief financed by borrowing.

When a government spends money in excess of its receipts, the effect upon the economic system as a whole, so far as concerns the relationships which we are here investigating, is similar to that of an equal volume of private investment financed by borrowing, and the effects of such deficit spending may be traced out in the same manner as the effects of private investment. An increase in government spending will lead to a simultaneous increase in the incomes of individuals or business men who receive the benefit of the spending. They in turn will be in position to increase their expenditures and thus increase the incomes of a second set of individuals. Thus the expenditure of, let us say, a billion dollars of borrowed money by government may produce an increase in aggregate gross income of several times that amount. The proponents of government spending argue that, if it is carried out on a sufficiently large scale, it may initiate a healthy recovery of private investment. They reason that an increase in aggregate income always stimulates investment and that deficit spending by the government, when its full effects upon income have worked themselves out, will initiate a normal recovery and thus permit the deficit spending itself to be halted.

It must be noted however that the similarity between private and government spending applies only to the act of spending the money. When private enterprise borrows funds for investment there is normally the expectation that the undertaking will produce an income sufficient to cover (1) all operating costs, (2) interest on the funds invested, and (3) amortization of the loan; *i.e.*, to accumulate by the time the loan matures a reserve fund equal to the amount borrowed. Public expenditure on the contrary often involves the government in debt without leading to the creation of income producing assets capable of yielding a surplus over costs for the payment of interest and amortization. In this respect government deficit spending contrasts sharply with the investment of borrowed funds by a private enterprise. It is for this reason that private investment may continue indefinitely in an ex-

panding economic system. Although it usually involves a steadily increasing total of private indebtedness, it also involves the creation of a larger and larger volume of capital capable of earning a net operating income for its owners and thus enabling them to carry the interest charges they have incurred. In contrast public deficit spending, if continued indefinitely, has the effect simply of increasing the interest burden that must be borne by the government which finances it.

Furthermore the favorable effects of government borrowing upon income may be offset by an unfavorable effect on business, if it results in raising interest rates and making capital difficult to obtain. Therefore the policy of deficit spending must, to be effective, be accompanied by an easy money policy on the part of the central bank. Again if government spending is carried out in such a way as to increase costs of production, and above all costs of investment for the private business man, then its net effect upon private investment may be unfavorable instead of favorable. Similarly if government funds are spent in the creation of agencies and facilities which compete with private business, the net effect of the spending may again be unfavorable to private investment. A final point, perhaps more important than any of the preceding, is that the continuation of a large government deficit over a period of years may have depressing effects upon business confidence. Business men envisage the prospect of currency inflation, which in the past has generally followed in the wake of deficit financing by the government; such inflation injects uncertainty into all business conditions and gives to investment the character of gambling. Business men take alarm from the possibility of actual collapse of government finance and the unpredictable actions which government may take in that event. More obviously the business world is disturbed by the prospect of further additions to the burden of taxation, thus adding to business costs elements which cannot be predicted or intelligently provided for.

We may conclude that a program of deficit spending will for the time being increase the aggregate income and reduce the volume of unemployment, but that if it is to have the desired secondary effects upon private investment it must be carefully engineered so as to avoid unfavorable effects upon interest rates, costs, or the demand for the services of private business. Above all it must be financed in a manner that commends itself to the business world as sound. Because the results of government spending depend upon so many intangible factors, its ultimate effect will depend upon the way in which the program is carried out. Even though a public spending program be carried out in such way that it does not impair the confidence of business men, it will not suffice by itself to produce a period of prosperity unless substantial private investment opportunities appear. Public spending may be successful if it merely initiates a revival for which the stage is already set. But if no substantial private investment opportunities appear, then the revival is not likely to last when government deficit spending is stopped. These conclusions are strongly supported by the experience of the United States during the last few years.

CHAPTER XXIX

Monopolistic Industry and Government Industry

The nature of monopoly. Monopoly, as the term is used in economics, is the exclusive control of the supply of an economic good. The essence of monopoly is therefore the power lodged in the hands of someone to raise the price of the good which he offers for sale by restricting the quantity placed on the market. In the absence of legal regulation, monopoly price is usually higher than a competitive price would be, and monopoly profits are normally larger than the normal profits necessary to induce the production and sale of these goods.[1] In this sense therefore monopoly means the exploitation of consumers for the benefit of certain producers.

Capitalistic monopoly. Monopoly appears in many different forms. A city may create a monopoly by granting a railway company an exclusive franchise to lay railway tracks in the city streets, or a government may reserve to itself the exclusive right to manufacture and sell tobacco or liquor. The owner of the sole source of supply of some commodity or service has what is called a natural monopoly in contrast to the legal monopolies created by the state. In this chapter however we shall be concerned with the so-called *capitalistic monopoly*, for it is this type which gives rise to the most serious public problems. Capitalistic monopoly rests on no exclusive control of the raw materials of industry, nor yet on any special privilege conferred upon it, although these may be present and may facilitate its growth. It arises rather out of the highly capitalistic method of production which characterizes modern industry and which makes it possible for a concern both to acquire and to maintain its position of monopoly, not by virtue of superior productive ability, but because of more ample financial resources.

[1] See Chapter XII.

The capitalistic monopoly is a by-product both of large-scale production and of the corporate form of organization. When production is on a small scale, the number of actual or potential competitors is so great as to impede the formation of monopoly. Even if a monopoly be formed, the capital outlay necessary to enable anyone to challenge the position of the monopoly is so small that its life is likely to be short. In certain lines of manufacturing however the superiority of the large producing unit is such that the small producers have to a considerable extent been eliminated, thus leaving the field mainly to a smaller number of large-scale units; this condition simplifies the problem of effecting a monopolistic combination.

In a still more significant way large-scale production has tended in the direction of monopoly. Such production generally means a relatively heavy permanent investment. This not only tends to discourage potential competitors, but it has a determining influence on the type of competition among those already in the field. Cutthroat competition is apt to develop in any industry where the fixed costs are heavy in relation to the variable costs, and the effects of competition of this type are so ruinous that even the threat of a price war may cause a scramble of competitors to seek shelter in combination.

It is difficult to see how a monopolistic control in any large industry could be built up on the basis of individual proprietorships or partnerships. The mere size of the capital investment and the hazards of unlimited liability would generally be prohibitive. There are also the difficulties which would ensue on the death of the individual owner or of a partner. It is indeed doubtful if monopoly would present so serious a problem without the device of the holding company, owing to the difficulty and expense of consolidating competing companies by outright purchase and amalgamation. It is in this sense that the monopoly problem is also a problem of corporate organization.

The reader will doubtless recognize that complete monopoly — the exclusive control of supply — is seldom if ever attained by mere capitalistic monopoly without the aid of some other contributing cause. It is virtually certain that some other sellers will always be present in the market. However the serious public problems of monopoly do not wait upon the complete consummation of capitalistic monopoly but appear as soon as one concern has progressed so far in that direction as to be able to shape its selling policy without regard to its small rivals and so to wield a predominant influence in the control of price. Indeed the problem may arise even earlier than this; that is, as soon as there is evident a clear tendency toward monopoly of the capitalistic sort. Actually there are, outside the public utility field, very few real monopolies in the sense in which we define that term. The subject matter of the first part of this chapter is accordingly, not industrial monopoly in the strict sense, but monopolistic industry, or industry that approaches a state of monopoly.

Monopolistic industry: The problem. Popular support for governmental regulation of monopolies in the United States undoubtedly comes from a well-justified feeling that monopoly profits are not representative of any social service for which payment is due.

Valid as this reason is for public suspicion of monopolies, there is an even more powerful argument for regulation in the fact that the extension of monopoly control over any considerable portion of economic life would destroy the basis of private capitalism. We have observed in an earlier chapter the importance of competitive price as a regulator of economic life. We look to competition to determine the distribution of the social income, to direct the flow of capital, to select the goods to be produced, to determine the quantities of each which shall be produced, and in other ways to promote the effective functioning of the diverse elements of our economic structure. At best it is not a perfect regulator, for adjustments do not take place immediately or automatically, and many

maladjustments may require conscious control, as for example in the case of the public utilities. But should monopoly conditions ever hold sway over industry as a whole or be proceeding to that end, there could be no alternative to complete governmental control over economic life, for there is no reason to believe that monopoly price has any virtue as a regulator of economic life. It cannot be emphasized too strongly that the maintenance of truly competitive conditions is the prerequisite of economic liberty. Regulation to secure this should be viewed, not as an infringement of individual liberty, but as a preserver of it.

Historical: The trust movement. The term "trust," although it has another and technical meaning, has in common language come to be identified with the large industrial enterprise of monopolistic power or tendency, and we shall find it convenient to use the term with that significance. In the United States the movement toward combination for the sake of monopoly power — "the trust movement" — dates from the late seventies; in the eighties it gathered momentum, and in the nineties it attained its greatest proportions. Some indication of the extent of the movement can be gathered from the statistics of combinations formed. In the period from 1887 to 1897 eighty-six combinations were formed, in 1898–1900 one hundred and forty-nine, and in the years 1901 to 1903, one hundred and twenty-seven. Relatively few of the combinations formed have been "trusts," but from 1898 to 1900 a larger portion of the combinations succeeded in acquiring monopoly powers in their respective fields than in any other period before or since. After 1903 the movement slowed down, partly because of the difficulty experienced in marketing the huge volume of securities issued at the time of the formation of each new combination, partly on account of adverse court decisions in anti-trust cases, and partly because of the fact that trusts had already been formed in most of the lines where monopoly was at all possible and indeed in some lines where monopoly power could not be retained.

With the close of the World War the formation of combinations again proceeded at a rapid pace. During the years 1919–28 there were 1,268 combinations formed in manufacturing and mining, involving the union of 4,135 and the disappearance of 5,991 separate concerns, while in the field of public utilities 3,744 public utility companies disappeared in the years 1919–27 through acquisition by other companies. With the onset of the depression in 1929 the formation of combinations decreased abruptly.[1]

Anti-trust legislation in the United States. American public opinion has always looked askance upon monopoly and indeed upon all large business enterprises and combinations which bear any odor of a tendency toward monopoly. In the course of the rapid movement toward combination, whose outline we have sketched, there arose fear lest the entire industrial fabric become tainted with the virus of monopoly and the control of industry become concentrated in the hands of a relatively small number of men. One result was the enactment of legislation, both by the states and by the federal government, designed to break up trusts and to prevent the existence of conditions favorable to the development of monopoly. Our attention will be given to the activities of the federal government, which, because of its jurisdiction over interstate commerce, has been most directly concerned in the political attacks upon the trusts.

The Sherman anti-trust act of 1890. With the exception of the interstate commerce act, no legislation designed to control trusts was enacted until 1890, when the act commonly called the Sherman anti-trust act was passed.

The Sherman act is brief and exceedingly general in its phraseology. In the first section "every contract, combination in the form of a trust or otherwise, or conspiracy, in restraint of trade or commerce among the several States or with foreign nations" is declared illegal, and according to the

[1] *Big Business: Its Growth and Its Place.* Twentieth Century Fund, Inc., New York, 1937, p. 32.

second section "every person who shall monopolize, or attempt to monopolize, or combine or conspire with any other person or persons, to monopolize any part of the trade or commerce among the several States, or with foreign nations, shall be deemed guilty of a misdemeanor. . . ." Under the act the government may bring criminal prosecutions against persons guilty of violating its provisions, or it may apply to the circuit courts, by means of suits in equity, to prevent and restrain violations of the act. Upon conviction in the courts of a violation of the criminal sections of the act, a penalty, consisting of fine or imprisonment or both, may be imposed. Any person injured in his business or property by any act declared unlawful or forbidden by the Sherman act is authorized to start suit in a circuit court of the United States, and if his claim of damage be upheld in the court he can recover three-fold the damages sustained and the costs of the suit, including a reasonable attorney's fee.

Operation of the Sherman act. The criminal features of the Sherman act have not proved especially useful in preventing the growth of trusts, so difficult has it been to secure a conviction by jury on the rather vague charge of restraint of trade. On the other hand, in the field of forcing the dissolution of trusts, the government has met with greater, though by no means unvarying, success.

During the first ten years after its enactment the Sherman act seemed ineffective in preventing the formation of trusts or in breaking up trusts already in existence, and it was not until President Theodore Roosevelt began his energetic campaign against the trusts that there was convincing proof that the Sherman act was still alive.

One of the first of the significant cases to come up at this time was the Northern Securities case.[1] The Northern Securities Company had been organized as a New Jersey corporation to acquire control, by stock ownership, of two parallel

[1] The government had already (1897) established the point that railroads were amenable to the Sherman act.

and presumably competing railroads, the Northern Pacific and the Great Northern Railway Companies. The government instituted a suit in equity to have the company dissolved, and in 1904 the Supreme Court rendered its decision ordering that this be done. Of still greater importance was the dissolution of the Standard Oil Company, which was ordered in a decision rendered in 1911. This case marks a new interpretation of the terms of the act. According to the language of the earlier decisions all combinations or contracts in restraint of trade were held to be unlawful, but in this decision the Supreme Court enunciated the so-called "rule of reason." As it is understood today, the rule of reason means that a restraint of trade which is not undue or unreasonable and which is incidental to a lawful purpose is not unlawful.

We may note one other case, that involving the Tobacco Trust. The Tobacco Trust was in part formed by consolidation, and it remained to be seen whether outright consolidation of the competing businesses would stand the legal test. The court however went behind the form to the motives and results of the Trust's activities and ordered its dissolution.

The trade commission act of 1914. While considerable success had attended the efforts to dissolve trusts under President Theodore Roosevelt — and President Taft continued the program with equal energy — the feeling that more rigorous legislation was desirable was clearly reflected in the campaign pledges of the three parties actively in the field in the election of 1912. Mr. Wilson came into office pledged to "the proposition that private monopoly is indefensible and intolerable," and early in 1914 he called upon Congress to enact legislation to supplement and amend the Sherman act. As a result of his recommendations two new laws, the trade commission act and the Clayton act, were passed by Congress in 1914.

Investigation of corporations. The first of the above-named acts of 1914 provided for a new administrative agency in the form of the Federal Trade Commission, su-

perceding the former Bureau of Corporations. The act gave the Federal Trade Commission comprehensive powers of investigation and report, and granted it the authority to call upon business men to produce records of all sorts and to give testimony when required. The commission was authorized to investigate the practices of any corporation guilty or alleged to be guilty of violating the anti-trust laws; furthermore it was permitted, and in some circumstances required, to investigate the way in which a corporation was carrying out the decree of a court in an anti-trust case. But the act did not limit the commission to the investigation of infractions of the anti-trust laws, for it could call upon any corporation for information as to its business organization, methods, and relations with other business units. This provision was inserted in order to enable the commission to accumulate a fund of knowledge with regard to the conduct of business which would be of value to Congress and to the public.

Advisory functions. Certain advisory functions were also provided by the trade commission act. In the first place, it was made the duty of the commission, upon the application of the Attorney General, to recommend a method of readjustment which would bring within the law a corporation alleged to be guilty of violating the anti-trust laws. An amicable settlement was believed in some cases to offer advantages over a long-drawn-out and costly lawsuit. In the second place, the commission was required, at the request of the court, to prepare an appropriate decree in any suit in equity brought by the Attorney General under the anti-trust laws. The court might accept the decree, modify it, or reject it entirely, but at least it would have the benefit of expert advice. Finally the commission was given the right to recommend legislation to Congress on the subjects which lie in its field.

Unfair competition. It is also the duty of the Federal Trade Commission to prevent unfair methods of competition. In cases adjudicated under the Sherman act, the courts had

often been influenced in their decisions by acts which, while perhaps innocent in themselves, could not be considered lawful when practiced by a great combination in circumstances which seemed to indicate an intent to destroy competition rather than merely to gain business. Among the many methods which have been held to be unfair, we may mention: (1) price discrimination, which means that prices are cut, below the cost of production if necessary, at the competitive points while being maintained at a higher level elsewhere; (2) the exclusive dealers' or factors' agreement, which compels a wholesaler to handle the goods of one concern exclusively or penalizes him, by charging him higher prices, if he carries competitive goods; (3) the use of "fighting" brands, especially designed to be sold at a price ruinous to competitors; and (4) brand imitation.

The trade commission act created a method whereby relief might be obtained without the cumbersome recourse to a damage suit, which alone was possible previously. A concern which believes that it is being injured by an unfair method of competition may complain to the Federal Trade Commission, which, after a hearing, may grant relief through a restraining order to the offender, with later appeal if necessary to the courts. The courts have the final decision as to whether practices condemned by the commission's restraining orders are actually "unfair" in each particular case.

The amendment of 1938. The expression "unfair methods of competition" in the original act was interpreted by the courts as referring only to methods which were unfair to some other competitor. An amendment to the Federal Trade Commission act, passed on March 21, 1938, broadened the legal prohibition so as to include not simply "unfair methods of competition in commerce" but also "unfair and deceptive acts or practices in commerce." Thus the commission is now able to proceed against practices inimical to consumers and the general public even though no direct injury to competitors can be shown.

This amendment also corrected an administrative weakness. Under the original act the orders of the commission were not self-enforcing, since no penalty attached to non-compliance. Hereafter the cease and desist orders of the commission become final after sixty days if no appeal to the courts is begun within that time, and after that period penalties for non-compliance apply automatically.

In the past much of the commission's work has been directed against false advertising as an unfair method of competition. The amendment of 1938 enlarged the powers of the commission in this field by prohibiting the dissemination of any false advertisement to induce the purchase of food, drugs, devices, or cosmetics. An advertisement is "false" if it is misleading in any material respect or if it fails to reveal material facts.

The Clayton act (1914) and the Robinson-Patman act (1936). The Clayton act of 1914 prohibits, with certain limitations, local price discrimination, tying contracts, holding companies, and interlocking directorates, the intention being, by making these practices specifically unlawful, to stifle the creation of trusts, conspiracies, and monopolies in their incipiency and before their consummation.

In 1936 the prohibition against price discrimination was elaborated by the passage of the Robinson-Patman act, which provides that it shall be illegal either to grant or to receive, directly or indirectly, any discrimination in price or service in any sale of commodities in interstate commerce. Price differentials may make due allowance only for differences in the costs of manufacture or sale to different purchasers. Furthermore the Federal Trade Commission may fix and establish quantity limits if price differentials on purchases in larger quantities would be discriminatory or promotive of monopoly. Whenever price discrimination is found, the commission is to order its removal, and the burden of proof in defense of discrimination in prices, services, or facilities rests upon the person charged with the discrimination. The difficulties and complexities inherent in the application of these statutory

provisions have raised serious doubts as to the possibility of effective administration.

Remedies provided by the anti-trust acts. In addition to proceedings before the administrative boards under the amendments of 1914, the statutes have provided other methods of proceeding against violations of the anti-trust laws. Criminal prosecutions may be started by the government, charging monopoly, attempt to monopolize, or conspiracy in restraint of trade. The government may initiate suits in equity, seeking the dissolution of combinations existing in violation of the law, or injunctions against specific acts or agreements in restraint of trade. The government may demand the forfeiture of goods being carried in violation of the law when moving in interstate commerce. Individuals and corporations may sue for recovery of three-fold damages for injuries resulting from violations of the law, and under the Clayton act individuals and corporations are permitted to seek injunctions against losses threatened through violation of the anti-trust laws.[1] The Clayton act provides that if, in a suit brought by the United States, it is decided that the defendant has violated the anti-trust laws, this fact must be accepted as *prima facie* evidence in any suit brought by another party under these laws, thus making it easier and cheaper for the injured person of small means to sue a trust.

Achievements of the United States anti-trust program. Though little effort at enforcement was made in the years immediately following its enactment, the Sherman act and its amendments have been a continuing threat to monopolistic practices since the turn of the century. The government has won a number of spectacular victories in the courts, securing the dissolution of some of the most conspicuous trusts and combinations. In many industries of national importance the threat of anti-trust proceedings has sufficed to secure voluntary

[1] It may also be noted that violation of the statutes may be cited as a defense in a private suit brought to enforce a contract alleged to be in violation of the law, and criminal proceedings may follow a violation of an injunction issued by a court in an anti-trust case.

agreements, or "consent decrees," whereby the alleged offenders have accepted court orders binding them to desist from specified practices criticized by the government. The large and powerful business units, being more exposed to the critical appraisal of the public, have found themselves compelled to forego the ruthless tactics of the early Standard Oil and American Tobacco trusts, and policies of coöperation, "live-and-let-live," and persuasion have become the usual methods of rendering competition less distasteful to industry. The Federal Trade Commission has been successful in eliminating many unfair competitive practices and raising the level of business ethics.

At the same time it must be recognized that such results as have been achieved have not sufficed to stop entirely the trend toward combination and concentration of control in American business, and there is evidence pointing toward a continued decline in competition in certain lines of industry. In view of the avowed objective of the federal legislation to prevent monopoly and preserve competitive conditions, it must be admitted that the regulatory efforts have at best attained only partial success, since the problem is still with us. Nevertheless comparison of existing conditions with those which would presumably have prevailed in the absence of governmental interference and regulation is sufficient to indicate a real accomplishment to the credit of the anti-trust program.

Is monopoly inevitable? The many forces which appear to be tending toward monopoly and the moderate degree of success achieved by the government in opposing such forces by restrictive legislation raise the interesting and far-reaching question whether there is after all a fundamental tendency toward monopoly which it is idle to resist — whether in short monopoly is inevitable. If it be true that natural laws, operating in a social system based as ours is upon personal liberty, private property, and individual control of enterprise, tend inevitably to bring about a condition of monopoly in the production of a substantial part of the basic products which

mankind requires, then we are faced with a human problem of the first magnitude, a problem which may well put upon the defensive the very fundamentals of the existing social order.

Forces favoring monopoly. At first glance the forces favoring monopoly appear very strong. There is first of all the well-known economy of large-scale production in many lines of industry. There are the economies which in certain cases appear to accompany the combination of separate business units under a single management. There is the occasional appearance of cutthroat competition, an unstable situation leading almost inevitably to monopoly. There is the frequent effort of the captains of industry, when competition becomes especially keen and irksome, to find escape through some form of combination, making use of every possible device which human ingenuity with millions at stake can devise. The corporate form of organization is a helpful instrument. The protective tariff, while probably not the "mother of trusts" as it has sometimes been called, has unquestionably reduced the competition which American producers have had to face and in that respect has coöperated with the other forces tending toward monopoly. And besides the desire to secure monopoly profits there is also the urge to secure promoter's profits by the organization of large combinations which may have monopoly power.

Acquaintance with economic history will convince one that the enactments of legislative bodies are futile if they are in opposition to powerful economic forces. If cutthroat competition is an inevitable feature of large-scale production, it would be the part of wisdom for society to seek stabilization in monopoly and to direct its efforts to regulation calculated to enforce good service and reasonable prices. If a combination of producing units under one management is more efficient than when each of the units is under separate management, if a combination of combinations in exclusive control of the field is the most efficient of all, it is folly to try to run counter to the natural tendency of business men to seek the

most effective form of organization. Society cannot permanently reject the most economical type of productive organization. This is virtually the conclusion which we have reached as regards the railroads, and our railroad policy and legislation are now based largely on the assumption of the inevitability of monopoly.

Limits to monopoly. But the conditions which prevail in the realm of transportation, communication, and the municipal public utilities, which confessedly point the way to ultimate monopoly, are not universal throughout the whole field of production.

At the forefront stands agriculture, most basic of all basic industries, where economy is not to be found in large-scale production or combination, and where fundamental economic laws seem destined to preserve indefinitely the régime of small productive units and free competition.

In the fields of manufacturing and marketing it is doubtful if the monopolistic combination is in many cases the most efficient form of organization. We have already learned that the efficiency of large-scale production and combination has very real limits. There are few important lines of industry in which this limit would not be reached long before the would-be monopolist had become great enough to absorb the whole. Even assuming the monopoly to become established, lethargy in management and lack of progress in the technique of production and the methods of organization will tend to appear, so that the monopoly which at the moment of its formation is representative of the most efficient methods of production carries with it no assurance of continued efficiency.

There is very little reason, from the standpoint either of theory or of history, for believing that the economies of combination alone are sufficient in volume to warrant the growth of monopolies in many fields of production. Nor is there evidence to show that competition between large units must necessarily be of the cutthroat variety. This is essentially the result of an unbalanced condition in production, where the

total productive capacity in a given industry is greater than the market requires. This situation is more or less normal in certain lines, such as the railroad business and the public service industries in general; here monopoly may frequently be inevitable and, under adequate regulation, in the public interest. As to industry in general, apart from occasional excess capacity resulting from mistaken judgment or depression conditions, there is little reason to anticipate that such a condition will present a constant menace to industry. It is more characteristic of the general situation that, as each period of intense competition weeds out the weaker producers, industry settles down to a normal condition of competition between a few strong rivals, no one of which has reason to hope for monopoly control. While reduction in the number of competitors may produce conditions of imperfect competition, it must be remembered that imperfect competition is still competition and often of the most vigorous sort.

Reasons for moderate results of the anti-trust program: Standards set by legislation. Having concluded that, powerful as the forces are which tend toward monopoly in industry, they are not so compelling as to make monopoly inevitable, we must look elsewhere for an explanation of the moderate results of the anti-trust program.

One of the basic causes is to be found in the fact that present legislation, as it is interpreted by the courts, places great stress upon the question of intent on the part of the corporations affected. Thus when a case comes before the courts it is not enough to consider merely the results of the combination; the intent to achieve an illegal combination or restraint must be shown. The demonstrable economic facts — whether the combination realizes important economies and whether it creates opportunities for monopolistic restraints — the only objective criteria by which the law can be administered, are not conclusive in proving a violation of the law. The absence of objective criteria fosters protracted and expensive litigation and imposes upon the Department of Justice the impossible

task of ferreting out evidence as to the unexpressed intent of those who formulate and control corporate policies.

The task of administration. Another important cause of the meagre accomplishment of the federal anti-trust legislation is to be found in the very magnitude of the task in relation to the administrative machinery that Congress has provided. Initiative in administration centres in the Anti-Trust Division of the Department of Justice, which is responsible for suits arising under some thirty-one important federal statutes. With its present staff and multitudinous responsibilities, the division finds it impossible to prosecute, or even to investigate, all of the alleged violations of the anti-trust laws that are reported to it.

Influence of patent laws. The patent laws have in practice sometimes served the ends of the monopoly builders. When the owner of a patent licenses others to manufacture under his patent, he may specify the quantity of the patented article to be produced, the markets in which the article may be sold, and the price to be charged. The patent laws have sometimes stifled new competition at its inception, by limiting the market for patented improvements on existing processes or mechanisms to the owners of the basic patent. Newly patented articles which might provide competition with products in current use have on occasion been bought and suppressed by corporations financially interested in established products or technical processes. Thus the patent laws, whose basic purpose is frankly to grant certain monopoly privileges, have in some cases promoted the development of monopolistic industry beyond what the spirit of patent legislation contemplated.

Contradictory laws. A loss of faith in the efficacy of competition is reflected in the enactment of laws by the federal and state governments in conflict with the basic assumptions of the anti-trust legislation. At the present time serious threat to the continuance of competitive conditions is to be found in a recent epidemic of state "fair trade laws." These laws permit manufacturers of branded or trade-marked articles

which are in "fair and open competition with commodities of the same general class" to enter into contracts with dealers for the maintenance of designated prices in the resale of such articles. Similarly in conflict with the spirit and the prevailing interpretation of the federal anti-trust statutes is the Miller-Tydings act, passed by Congress on August 17, 1937, which exempts such price maintenance contracts from the prohibitions of the anti-trust laws. Of similar effect is the already large group of state "unfair sales practices laws," which seek to limit the power of manufacturers and dealers to reduce prices in competition. By such legislation government itself is working against competition.

Conclusions. To assert that the survival of the system of private capitalism — perhaps even of political democracy itself — requires the control of monopoly and such regulation of trade and industry as will maintain competitive conditions in a reasonably large portion of industry is not an overstatement of the case. The efficient regulation of trade and industry demands legislation soundly conceived and drafted, administration by competent and honest public servants, and the support of an alert and intelligent public opinion.

The problem is not a simple one. Conditions vary from industry to industry, and an effective program of regulation must have sufficient flexibility to adjust itself to many situations. There is even reason to expect that experience may find instances, perhaps not now suspected, of industries to which competitive controls are not appropriate. Here the extension of regulation such as is now applied to the public utilities may be the order of the future. Past experience also suggests that laborers, consumers, and investors may require additional protection from some of the undesirable results of competition. But despite exceptions and difficulties, there is reasonable hope for the success of a revised and vigorous program. Certainly the importance of the goal warrants the attempt.

The attitude toward government industry. Every government is inevitably engaged in industry to some extent. A

people may feel that the postal service is too important to the general welfare to leave in private hands. Military considerations may lead a government to control the railroads, the telephone, the telegraph and wireless service, as is not uncommon in Europe. Protection against disease sometimes requires city governments to take a hand in such enterprises as milk stations, bath houses, housing projects, etc. To protect the country against flood and drought, as well as to provide for future timber requirements, a government may establish national forests and build reservoirs and dams to regulate the flow of water. Enterprises such as the Panama Canal or Boulder Dam may be too vast or the returns too uncertain to interest private capital; government however may decide that regardless of the financial return such undertakings are worth while in the public interest.

These examples — and many more will at once occur to the reader — illustrate the fact that there is a province for government industry which is little disputed. Few question the propriety of government industry undertaken in the realm of its primary functions, nor is there much criticism of government projects in fields where the risk is so great as to deter private capital from entering, when once it is settled that the project would truly serve the public interest. Serious disagreement regarding government industry arises when government invades fields which tradition has reserved for private capital.

Those who advocate government industry in order to protect consumers from the exactions of private monopoly are not to be confused with those irreconcilable critics of the present order who are persuaded that the weaknesses of private capitalism are so fundamental that the only hope of stable prosperity lies in the direction of socialism. On the contrary the former advocates of government industry are men who believe in the present economic system and want to see it perpetuated. They are however so impressed with the weaknesses of government regulation that they feel that the system

can best be preserved by removing from private hands those parts of it which have become monopolistic in character and marked by conspicuous abuses.

Opposed to them is a large group by no means convinced either that government regulation is futile or that the abuses of private business are sufficiently serious to warrant so drastic a step. Moreover many of these are convinced that government industry tends usually to be less efficient than private business, and they are genuinely fearful lest an enlargement of the scope of government industry bring consequences even more unwholesome than those which it is intended to cure.

Financial results of government industry. An investigation into the efficiency of government enterprises gives a picture that is far from clear. In one country state-owned railways may give excellent service and show satisfactory financial results, while in another country poor service and continued losses follow year after year. The United States Post Office, the nation's most conspicuous example of government industry, has operated for years without any real accounting system and with a very large and unnecessary deficit. These results are chargeable not only to legislation by Congress but also to serious lack of sound business judgment in the management of the enterprise.

In the field of municipally owned public utilities the evidence as to financial results is often incomplete, conflicting, and even unreliable. Frequently the accounting systems of municipal enterprises are not adequate or reliable; the United States Census Bureau complains regularly of the untrustworthy accounts upon which it has to base its statistics. Apparent net operating revenue must in many cases be reduced because of failure to make proper allowances for interest, depreciation, or amortization. Services, such as street lighting, may be provided free to a city without an appropriate charge on the books, and similarly many services may be provided by other departments of the city without appearing on the accounts. In the face of these facts any sweeping generalities

would be unwarranted and unsound. On the whole the financial record of public industry is unfavorable; there are conspicuous cases of financial success, but the number and magnitude of the industries operated at heavy deficits is more impressive; and in most cases — in America at any rate — the actual facts are far less favorable than the published figures on account of the prevalence of faulty accounting methods.

The problem of government industry: Interest and profits. Attention is directed frequently to the large sums paid for interest on private capital and the large profits made by private enterprise, and it is claimed that these might as well accrue to the whole people in the form of lower rates. It is evident that such claims cannot be substantiated. So long as saving is left to private initiative, interest is a charge which must be met whether the industry be public or private. It is a payment essential to the accumulation of capital, and the government cannot get capital without the payment of interest any more than can a private entrepreneur. If attention be called to the fact that government bonds bear a lower rate of interest than industrial investments, the answer is that the financial difficulties which are the heritage of the World War have already deprived many governments of this advantage, and the whole advantage would tend to disappear in proportion as government became involved in the risks of industry. If the taxing power be relied upon to make good the losses of unsuccessful government enterprises, this is merely robbing Peter to pay Paul. Even in a different social order in which capital were not privately owned, there could be no capital without previous saving by someone. Government might, it is true, get its capital without paying interest, but the ultimate cost of saving would not be thereby avoided.

The same general considerations apply to the alleged saving of profits. Profits, as we know, are paid for the service of risk bearing, and we have concluded that the payment, considering losses and gains together and taking industry as a whole, is distinctly moderate. If the risks of industry were transferred

from the shoulders of private entrepreneurs to the government, they would still have to be paid for. Governments cannot take the profits and avoid the losses. The government industry would have to charge enough to enable it to make experiments, assume risks, and stand the inevitable losses. There is no reason to believe that the cost of risk bearing would be any less if undertaken by the government, unless it were to avoid risk by permitting industry to stagnate.

Profits and local public utilities. While the above considerations apply to industry in general, it is possible that in the special fields into which the municipalities particularly have ventured there may be a saving to the consumer by reason of a reduction in the volume of profits. The purveying of water or electricity is generally a monopoly whether the plant be publicly or privately owned. As compared with competitive manufacturing, for example, the risk is slight and in many cases perhaps non-existent. The stockholders of a privately owned public utility might not be content with profits only sufficient to compensate for the risk involved. With private ownership pressure is constant to maintain existing rates or secure higher ones. Quite commonly a rate which at one time yielded reasonable profits has, with a thickening of the population, given very great profits. Under private management the action of the public service commission in reducing rates is slow and halting; under public ownership, assuming for the moment that public management be as efficient as private, or even slightly less efficient, either the consumers would get the benefit at once in the form of lower rates or the taxpayers would benefit from lower taxes.

Corporation abuses. In their advocacy of government industry, some lay greater stress on the necessity of eliminating certain undesirable aspects of corporate control than on the actual savings which government industry might secure. They point out that the problem of watered stock, which frequently affects the rate base and therefore the cost of service to consumers, would be done away with, that the looting

of utilities by insiders would be impossible, and that speculation in securities of utilities would no longer prevail.

But more fundamental is their belief that regulation of public utility rates has been halting and ineffective and that it is idle to place reliance upon regulatory commissions to secure rates which are only reasonably remunerative. It is not within our province to discuss the problem of regulation of public utilities, but it is perhaps not out of place to note that in many states appointments to regulatory commissions have been on a political basis, that the task of proving that a given rate is unwarranted involves long-drawn-out and very expensive litigation, and that the problem of regulation has been exceedingly complicated because of the interstate character of many of our holding companies and the absence of federal regulation.

It is however possible to admit the existence of these conditions, as do most authorities, without reaching the conclusion that government industry is the only solution, for the maximum possible effectiveness of regulation has certainly not been reached.

Politics. There are certain positive weaknesses which seem inherent in government management. In the first place, there is weakness in personnel and discipline. Politics controls more or less the selection of employees, both managers and subordinates. Politics also interferes with discipline; those in charge hesitate to enforce adequate control over their subordinates or to discharge the incompetent, for fear of political protest. For the same reason the managers have often been prevented from installing approved labor-saving machinery, cost accounting systems, or up-to-date business methods. Thus is efficiency sacrificed and industrial progress hindered. Political sentiment also tends to prevent the payment of salaries sufficient to attract into the public service the most capable managers and technical experts.

On the other hand, the utilities are frequently in politics and exerting pressure on legislation through lobbies and or-

ganized and "educated" public opinion. Society would gain something by removing this danger to democratic government, although it would probably stand to lose far more from the development of bureaucracy and the political control of government industry.

The motive in business. Public industry is weak in comparison with private industry as to the motive of the managing head. The private business man puts the utmost of his skill and energy into his business because he has a personal stake in it. Even if he be only the hired manager of a corporate business, he finds the same standard held up to him by the stockholders, who look to him to protect their capital and produce profits for them. His position and his fortune depend on making a success of the business under his charge. Self-interest is the spur which drives the private business man to make his business efficient.

To the public official this motive is weak or absent. He has not risked his own capital in the government industry. If profits are made they will not accrue to him, and he will not have to bear the financial consequence of failure. The continuance of his position and his further advancement do not depend upon the judgment of a critical body of stockholders as to the efficiency with which he has managed the business. His position is more often dependent upon pleasing the general body of voters or the leaders of his political party. This approval may be won by easy-going treatment of his subordinates and showy service to the public, rather than by the hard test of business efficiency.

This is of course not to assert that there never is nor can be efficient management in public business. There are capable men animated with a keen sense of public service, and there are communities in which the citizens insist on able and honest management. There have been conspicuous cases of high-grade leadership impelled by motives of patriotism and public service, especially in times of national emergency such as war. On the other hand, it is not to be denied that there are

instances in private business where other considerations complicate or interfere with the simple relationship between efficiency and the profit motive. This is most likely to be found in the very large corporations or combinations where ownership and management are less closely identified and other interests than those of the owners are more likely to appear. Indeed as private business units become exceedingly large there is sometimes observable a tendency to take on certain of the characteristics of government industry. Considerations such as these however are the exceptions to the general rule as to the important difference in motive between the managers of private and of public industries.

Conclusion. The problem of public *versus* private industry must still remain undetermined, in the sense that it is impossible to arrive at any single broad generalization as to the relative merits of the two forms of ownership. It is not at all likely that the problem ever will be solved in the form of such general conclusion. In no two countries and with respect to no two classes of industry are the conditions of the problem the same. The decision as to any industry may go one way in Germany, the opposite way in America; in any given country the decision may be for municipal water works and against municipal gas works. The defects of private industry may compel public ownership in one case, as of the forests on the water sheds; while the weaknesses of public ownership may lead to substitution of private industry, as in the case of the Philadelphia gas works some years ago. In one state effective regulation may give satisfactory rates and service, while in another a weak commission may lead municipalities to assume the risks of government industry. The mass of evidence and the body of general principles bearing on the question are such as to place the burden of proof upon the one who advocates government ownership. From this position the public will undoubtedly feel its way along the road of economic progress, determining each case upon its merits in the light of such knowledge as may be at hand.

One general conclusion may perhaps be stated with some confidence. The public is not likely to tolerate unlimited private monopoly in the necessities of life upon any large scale. In any basic line of industry where monopoly appears inevitable, the choice is between private industry with government regulation and government industry. Attempts at government regulation have thus far been generally short-sighted and bungling. We may now be upon the threshold of the first real test of this program; should it fail, the alternative of government ownership would still be open. In the meantime there may well be no definite conclusion — no decided swing either to public or private ownership — but rather a balance between the two, with private ownership the fate of the inefficient public plant, and public ownership a threat and a spur to the privately owned public utility.

CHAPTER XXX

Railroad Transportation

The railroads and the public. In some countries railroad transportation is a government function. In the United States and many other countries however the railroads are owned and operated by corporations, whose stocks and bonds are the property of thousands of individual and corporate investors; they are private wealth, owned and operated for private gain.

The fact of private ownership and operation however does not prevent the railroads standing in a class apart from the ordinary types of business enterprise, with public relations of the utmost importance. On the one side, the railroads occupy a peculiar position of dependence upon the government. Generous land grants were made by the United States government to several of the railroads, and in many cases state and local aid was forthcoming to assist in their building. What is of far greater importance, the government has given the railroads the advantage of its *right of eminent domain*, by virtue of which private property can be condemned for public purposes at a price determined by the courts or other authority. Without this aid from government, the railroads would generally find it impossible to acquire the necessary right of way. The dependence is not all on one side. The reason back of government aid is the supreme dependence of the people upon the railroads. This dependence in the United States has reached the point where a stoppage of service might assume the proportions of a national calamity and where inadequate service or extortionate rates might neutralize the advantages of large-scale production or at least impose a great handicap upon producers and consumers. In return for its coöperation the government imposes upon the railroads the

obligation to give to the public adequate service at reasonable rates. This double relation of dependence and obligation between the railroads and the public is what puts them, along with certain other public service enterprises, such as the street railways, the telephone, the telegraph, water supply, etc., in a special class apart from private business in general, to which has been given the appropriate title of *public utilities* or *public service corporations.*

The railway problem. Experience has shown that the operation of economic forces, such for example as determine the quantity, the quality, and the prices of the products of competitive industry in general, has failed to bring correspondingly satisfactory results in the field of railway transportation. When competition has prevailed, it has generally been of the cutthroat variety and has brought in its wake disturbing discriminations in rates, disorganization of service, and other abuses. On the other hand, competition has frequently broken down altogether, introducing serious problems of monopoly. The investment in the railroads is so great that their methods of financing have attained public importance. Differences between the railroads and their employees are matters of vital public interest, since railway strikes or lockouts may easily play havoc with the nation's commerce. There is thus a "railroad problem," whose public importance can scarcely be exaggerated. Though, as we have seen, it has many sides, the heart of the problem is in the determination of the rates charged by the railroads.

In our investigation of railway rates we shall find it convenient to posit the first steps of the inquiry on the assumption of free competition and the entire absence of governmental regulation, although the reader will doubtless recognize that this is not the situation existing today in the United States, whose railroad conditions are the result of more than sixty years of regulation both by state and federal authorities.

Railway expenses. The railroad business furnishes a most striking example of an industry subject to the law of decreas-

ing costs. The manner in which this condition emerges in an industry where there is economy in large capital investment and in which the plant of maximum efficiency is a huge organization has been made sufficiently clear in an earlier chapter.

Fixed costs. Interest. To begin with, the railroads require a large investment of capital. The right of way must be secured, tracks must be laid, rolling stock, signalling apparatus, etc., must be purchased or leased. Funds for these purposes are secured through the sale of railway bonds or other securities; and the interest on bonds and notes, which must be paid at definite rates, as well as the funds which the railroad may have agreed to set aside each year to provide for the eventual redemption of such securities, taxes, and the rentals of leased lines, form the greater part of the railroad's fixed charges. In harmony with the nature of fixed charges,[1] these expenses bear practically no relation to the volume of traffic. Given the facilities which are prerequisite to the satisfactory handling of any traffic, an increasing amount of goods and passengers may be hauled over the road without increasing the fixed charges, up to that point where it is necessary to enlarge the existing facilities in order to take care of additional traffic. When this point is reached additions to the fixed charges may be expected, but again, once they have been incurred, they will not vary until it is again necessary to enlarge the plant.

Other fixed costs and variable costs. Other fixed costs are costs which would cease if the railroad were abandoned but whose amount bears no close relation to the volume of business done. Costs which do so vary are called variable costs. These are general distinctions of importance, although in actual practice costs can seldom be thus separated by a hard and fast line.

The maintenance of the right of way includes rail and tie replacement and supervision. While these expenses vary somewhat with the volume of traffic, their major part is a necessary charge so long as any traffic at all is carried. Sim-

[1] See Chapter VIII for an analysis of fixed costs and variable costs.

ilarly about half of the expense of maintaining equipment and terminals goes on irrespective of the volume of traffic. The cost of conducting transportation consists chiefly of the wages of trainmen, fuel for locomotives, and terminal charges. These expenses are most responsive to changes in the volume of traffic, but even they do not change in exact proportion. Most of the general expenses, of which the chief item is the cost of maintaining general offices for the railroad system, bear but little relation to the size of the operations and can be assigned to the group of other fixed costs.

Both interest charges and other fixed costs are thus especially heavy in the railroad business. It has been estimated that about two-thirds of the total costs can be considered as constant,[1] leaving only about one-third in the class of variable costs.

Decreasing costs. The presence of this large ratio of fixed expense indicates that a railway operates under conditions of decreasing costs. Up to the point of maximum utilization, an increase in the number of passengers to be carried or in the amount of freight to be hauled does not cause a proportional increase in the total expenses. Hence the unit expense will decrease. If we assume that the rates remain constant while the traffic increases, profits must show an increase which is proportionally greater than the increase in the volume of business, and the reverse would be true if the rates remained constant while the volume of traffic declined. This explains the fact, which has frequently been noticed, that a moderate growth in the volume of traffic, with no increase in the rates, is often enough to raise a railroad from financial embarrassment to the position of a highly profitable business enterprise.

The unusually bitter competition between railroads, such as was found in the United States prior to federal regulation, finds its origin in this fact of sharply decreasing costs. So long as a railroad has unused capacity there is a constant temptation to cut rates or to give secret rebates to large shippers, in order to divert traffic from competing roads.

[1] Eliot Jones, *Principles of Railway Transportation*, New York, 1924, p. 78.

There is however always the likelihood that this may initiate a period of cutthroat competition, with all competing roads recklessly slashing rates until financial weakness forces them to seek peace in combination or consolidation.

Joint costs. Suppose it is desired to know the exact cost of shipping a ton of coal from Scranton to New York. We will assume that two freight trains, one consisting of twenty cars, the other of thirty cars of coal, are sent from Scranton to New York. It is quite possible to determine with a fair degree of accuracy the variable expenses of each train. There will be charges for fuel, wages for the trainmen, some wear and tear on the tracks and the rolling stock, and a few other expenses which might have been avoided entirely had the particular train not been run. These expenses will be greater for the second train than for the first, but they will not be 50 per cent greater, and if we were to determine the proportion of the variable expenses which should be borne by each ton of coal we would find the amount somewhat greater in the first case than in the second.

A third train of thirty cars is sent from Scranton to New York, but in this case it carries mixed freight — coal, cement, hardware, textiles, and groceries — and furthermore some of the cars are incompletely loaded. Again we can determine the variable costs for the train as a whole, but to distribute them among the different goods carried according to the expense caused by each individual shipment is another matter. The easy solution would be to calculate a flat rate according to weight or bulk, but it is quite clear that this does not accord with our desire to find out the expense of shipping each individual package. It costs more, for example, to handle a ton of small packages than a ton of coal. A car fully loaded with textiles costs little more to haul than one only half loaded, and a meticulous effort to allocate the variable costs according to the cost of rendering the service might lead to the surprising result that it cost more to carry some of the textiles than others on the same train.

An accurate apportionment of the variable costs presents enormous difficulties, but when we take the other expenses into consideration the case becomes hopeless. The fixed costs are incurred to provide equipment and personal services which are available, not only for the passenger and freight services, but for specialized services within these two great groups, and which cannot be apportioned on any scientific basis. *Costs which are incurred in the simultaneous rendering of two or more separate services and which can be apportioned only arbitrarily between the separate services or products are called common or joint costs.*

With regard to the greater part of a railroad's expenses it is impossible to tell what portion should be assigned to passenger service and what portion to freight service, to say nothing of further assignment of costs within each of these services. It is just as impossible as to say, for example, what part of the cost of raising cotton should be attributed to the cotton fibre and what part to the cotton seed.

Railway rates: Peculiarity. In competitive business generally, cost of production has an important if not a controlling voice in the determination of price, and even where monopoly prevails cost generally at least sets a minimum below which price will not fall. In railroading it is possible to estimate rather closely the additional out-of-pocket expense for affording an additional increment of service, but because of the presence of common costs it is impossible to secure more than a rough estimate of the total cost of rendering this service. While railroad managers can and do use this rough estimate of cost as a guide in setting rates, it is quite evident that specific cost of rendering service does not provide a satisfactory basis for rate determination. In practice the railroad managers know what their total costs are, and they seek to meet these total costs and earn what profit they may from the traffic as a whole.

Charging what the traffic will bear. In the absence of any scientific method of determining railway rates, the officials charged with the duty of rate making have recourse to a

practical method, which has been given the somewhat repellent name of "charging what the traffic will bear." Some goods can stand a higher rate per ton than others. A ton of pig iron may be worth from twenty-five to thirty dollars, while a ton of silk cloth would be worth thousands of dollars. A freight rate per ton or per pound which would be prohibitively high if applied to pig iron might have little effect on the selling price of silks and consequently on the volume of traffic in silk. A high rate therefore is charged for carrying goods which combine high value and small bulk, while rates on cheap bulky goods are relatively low. No freight of course will be carried at a rate insufficient to cover the additional cost which it would occasion, but so long as it pays enough to cover its own variable costs (estimated as closely as the traffic officials are able) and ever so little above, it is good business to take it. So the cheap bulky goods such as coal, trap rock, etc., are carried for what is actually less than the average ton-mile cost, while other parts of the service are counted upon to make up the interest and other fixed costs for the whole. In this way it is planned that the total revenue shall cover the total expenses and yield as large a net revenue as possible.

Discrimination between goods. Discrimination between goods is thus frankly an essential part of the prevailing method of determining railway rates. The word "discrimination" has an ugly sound, and a connotation of injustice is apt to attach to it, but a moment's reflection will show that such discrimination is not only inevitable but is in the interest of the people as a whole.

It will be granted that no railroad will be built unless the promoters of the venture can, with some degree of confidence, count on securing a gross revenue sufficiently large to cover all costs and yield a profit. If they were forbidden to charge different rates for transporting different classes of goods, it might not be worth their while to build the road, for a rate high enough to secure the maximum contribution toward the constant expenses from the high-grade freight would preclude

the possibility of shipping bulky freight, while a rate low enough to attract the low-grade freight would materially reduce the profitableness of carrying the other. In other words, discrimination between goods may be the only condition under which a railroad can be built.

Suppose now that a railroad, already built and in operation, sees an opportunity to add a new class of traffic, but only at a rate considerably below the scale paid on its present business. The new business will add little or nothing to the fixed expenses of the road, while the revenue obtained will contribute something, though perhaps not much, over and above the variable costs added by the new traffic. This will be profitable to the railroad, and it does no injury to the former patrons of the road. In fact it is possible that the contribution of the low-grade freight toward the fixed costs may make it possible to charge lower rates on the high-grade freight than would otherwise be imperative.

That taking on additional low-grade traffic at rates less than the prevailing scale is to the interest of the railroad with unused capacity is quite clear. On the other hand, a railroad which has reached the point of maximum utilization does not have the same inducement to increase the volume of traffic. In fact it may find greater profit in raising the rates on the least profitable goods, thereby reducing the volume shipped, in order to have greater carrying space for the more profitable freight, at least until such time as it appears wise to enlarge its facilities.

Local discrimination: Description. Two other kinds of discrimination are virtually sure to appear where there is competition between railroads. The first of these is local discrimination, which, in the broadest sense of the term, occurs whenever the rates between different points for a given class of goods are not proportional to the respective costs incurred. If the New York, New Haven & Hartford Railroad should charge the same rate for shipping a case of firearms from New Haven to Hartford, a distance of thirty-seven miles, as

it charges from New Haven to New London, fifty-one miles, this would be a local discrimination. If it should charge more to Hartford than to New London, its action would be even more clearly discriminatory. Finally if the rate to Springfield were equal to or less than the rate to Hartford, which must be passed through on the way to Springfield, we should have the extreme case, in which the shorter haul is included in the longer.

All variations from a flat rate per ton-mile, or other unit, are not evidences of local discrimination. The goods must be of the same kind; a different rate for coal and shoes is not local discrimination. They must be shipped in the same volume; it is quite natural to make allowance for carload shipments as compared with less than carload shipments. Finally the conditions of transportation must be essentially the same; we should expect the rate charged on trains running through mountainous country to be higher than rates charged on similar shipments through level stretches.

The reason for local discrimination. The chief reason and the sole justification for local discrimination are found in the situation created by competition with another railroad or with a carrier by water. Let us suppose that two roads run between cities A and B, as shown in Figure 30. The X & Y railroad has a more direct route between A and B than does the M & N railroad. The former charges, let us say, three cents a ton-mile on freight from A to C, C to D, D to B, and also on the stretch from A to B. The M & N railroad likewise charges three cents a ton-mile on its intermediate traffic, but if it were to charge the same rate from A to B, the total charge per ton would be considerably greater than that from A to B via the X & Y. It therefore charges the same amount as the X & Y railroad, even though this charge may be less than that from A to F. Shippers and consumers at F complain against this discrimination; they say, with truth, that more is charged

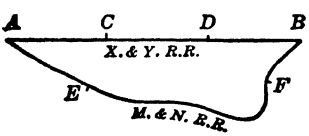


FIG. 30

for the short haul than for the long haul, even though the short haul is included within the long haul, and they further maintain that this practice is unjust and should be stopped.

The defense of the railroad rests on the assumption that some of the through traffic is essential, either to justify the existence of a railroad through the intermediate towns or to enable the intermediate rates to be as moderate as they are. The reasoning underlying this is essentially similar to that used to explain discrimination between goods. The rate on the through traffic is remunerative in the sense that it covers the variable expenses involved and makes some contribution toward the constant expenses. Without the through traffic this amount would have to come from the intermediate traffic alone, and the rates would therefore be higher than they now are, or in anticipation of such a situation the road would not have been built at all. The discrimination, though real, is nevertheless essential to providing service for the intermediate points, and while the producers, shippers, and consumers at the intermediate points may be handicapped as compared with those at A and B, they are better off than they would be without the service or with still higher rates.

When two railroads are competing for the through traffic, there will almost certainly be local discrimination on at least one road, and there may possibly be discrimination against local points on both roads. Discrimination between places is a serious thing. Towns against which there is discrimination are injured, their industries may be destroyed, and their population driven elsewhere. The towns which enjoy the low competitive rates are given artificial advantages over the other towns which may offset or completely eliminate real natural advantages possessed by the latter. Population is attracted to the favored towns, and the problems of city life are intensified. Nor do the railroads themselves as a whole gain by the practice, for while a certain increase in traffic is likely to be stimulated by the lower rates, it is questionable whether it will be sufficient to compensate for the rate reduction. Nevertheless

local discrimination is an inevitable feature of railroad competition, and it is probable that there would be a certain amount even under monopoly.

Personal discrimination. We have now to consider a third form of discrimination in railroad rates; namely, discrimination between persons. Such discrimination is peculiarly the result of railroad competition. Each road is forced to meet the cuts in the rates initiated by its competitors in order to get any business; if it can secure additional business by a secret contract with a particularly prominent shipper it is to its advantage, temporarily at least, to do so.

The essential feature of personal discrimination is to select some one shipper and attach him so firmly to the railroad by favors of one sort or another that he will patronize that railroad to the exclusion of all others. The tendency is quite naturally to favor the growth of industrial monopoly by restricting the area and scope of competition. Quite aside from its general social bearing, the growth of monopoly power among its shippers is not without danger to the railroad itself. When the fierceness of competition has died down and the railroad wants to readjust its rates, it may find itself no longer the master of the situation. The monopoly, ungrateful for past bounties, is in position to demand special rates by threatening to throw all of its business to the competing railroads.

Discrimination between shippers is totally at variance with the ideal of the public utility. It is quite contrary to all our notions of fair play that one man should be singled out and favored to such an extent that others are ruined and driven out of business, particularly by an agency upon which the whole public is dependent and which has received important privileges from the public. Nor can the encouragement to monopoly which results be viewed with complacence. That the public guards its own best interest in demanding that public authority put an end to these abuses is scarcely open to question.

The breakdown of competition. Complete competition between railroads has probably never existed in actual practice; even under what would be called competitive conditions, a partial monopoly usually is found. If one examines the routes taken by the roads running between New York and Chicago he will realize that, while competitive rates may exist on the through traffic, each of the roads has a more or less complete monopoly of its local traffic. In fixing the rates on local traffic the zeal of a railroad to secure monopoly profits is tempered only by the realization of the fact that the producers of its region have to meet in the common markets the competition of producers of other regions and that therefore the road cannot charge rates so high as to make it impossible for the local shippers to meet this competition.

The most important reason for the failure of competition however is that competition is self-destructive; it degenerates into cutthroat competition and ends in combination or monopoly. As we have seen, the presence of so large a percentage of expenses incurred for all services makes it impossible to equate the service to the expense of rendering it, and the fact that so large a part of the expenses are constant in the face of either a decrease or an increase in the volume of traffic places a premium on securing a large volume of traffic, up to the point of complete utilization of the plant. The result is that, when keen competition starts, the prime object of traffic managers is to secure traffic almost regardless of the rate. Rates will be cut, and they may fall so low as to cover only the variable expenses; secret rebates are given to attract shippers, while at the same time the shippers at the non-competitive points are charged monopoly prices. The end of this is the elimination of competition through consolidation or combination.

When once a road has gained a monopoly position, it is easy to hold. Even though the road may be securing monopoly profits which would ordinarily be more than ample to attract the investment of capital, it may be impossible to interest capitalists in the building of a competing railroad. Investors

want to be certain that the traffic is sufficiently large to assure both roads of an adequate rate of profit at the competitive rate on freight and passengers. The investment in a railroad is large, and when once made it cannot be transferred to other uses. Competition therefore gets under way very slowly and cautiously even in the presence of monopoly profits.

Monopoly inevitable and economical. There can be no escaping the conclusion that, on the whole and in the long run, monopoly in the railroad business is inevitable. And an equally important truth is that railroad monopoly is more economical than competition. Not only does a lower unit cost result from the more complete utilization of existing facilities, but when additional facilities must be provided it is better to enlarge the existing plant than to build a new railroad. The laying of additional rails, the expansion of terminals, etc., entail, it is true, a considerable increase in the fixed charges, but the additional cost is much smaller than that necessary to develop the right of way, lay the rails, and provide the equipment necessary for a new railroad. In other words, there are economies to be obtained from large-scale operation. A fully utilized two-track road operates at a lower unit cost than a similar one-track road, and the same holds true of a four-track road as compared with a two-track road. Monopoly allows us therefore to reap the benefits to be derived from complete utilization of the plant and to realize the economies which result from large-scale operation. It represents a more economical use of the limited supplies of labor and capital.

Regulated monopoly. While it is pretty clear that monopoly of rail facilities is the condition best adapted to the public interests from the standpoint of the most efficient utilization of labor and capital, unregulated monopoly is the cause, not merely of fears which may be somewhat irrational, but of misgivings which are well grounded on facts and principles. Unregulated monopoly carries with it no guaranty of adequate service at rates which are only reasonably remunerative

or of impartial treatment of all shippers. In fact we may expect the railroads to consider profits before service, to be influenced by private motives rather than motives of public service, to forget that they have, by reason of their quasi-partnership with the public, obligations which should be fulfilled.

From a practical standpoint unregulated monopoly of the railroads is inconceivable in the United States. The choice lies therefore between (1) government ownership and operation and (2) government regulation of the activities of private companies. The problems arising out of this conclusion form the subject matter of the remainder of this chapter.

State regulation of rates: Early history. The need for governmental regulation of the railroads was not clearly recognized in the pioneer days of railroad building and indeed did not find expression on the statute books until the years around 1870. In the preceding decades the primary need was felt to be the building of roads; competition could be relied upon to regulate them. Gradually however abuses sprang up which the public refused longer to tolerate, and legislation was passed in many states to control, to a greater or less extent, the activities of the railroads. It is significant of the spirit of the times that the individual states made the first move in this direction. The first state to create a railroad commission was Massachusetts, in 1869. But the movement for strict control of railroads found its origin in the agricultural states of the Middle West, which in the early seventies were suffering from an overexpansion of agriculture, as well as from the universal depression in business. The chief complaint against the railroads was that exorbitant rates were charged on shipments of agricultural produce, and when the competitive wars of the seventies somewhat reduced the force of this argument, the charge of local discrimination took its place. There was moreover a profound conviction that the common attitude of the railroad executives, who appeared to regard themselves solely as the representatives of private interests without obligation

to the public and not amenable to the legislative power, was dangerous and needed curbing. The corruption evidenced in giving passes to those whose official position might enable them to lend support, in bribing legislatures, and in the skilful juggling of the finances of the roads for the benefit of insiders added to the force of the movement.

The "Granger laws." Early in the seventies many of the states of the Middle West passed the Granger laws (so called because of the support which the National Grange of the Patrons of Husbandry lent to the movement) forbidding discriminations and prescribing maximum rates which might be charged. Naturally enough the railroads at once contested the constitutionality of laws which threatened to curb their arbitrary powers. We cannot enter into the details of these cases, but we may note that the Supreme Court laid down the doctrine that the railroads were common carriers affected with a public interest which it was the inherent right of the state to safeguard and further upheld the contention of the states that they had the right to prescribe maximum rates. In the course of time this absolute right to prescribe rates was modified to the extent of giving the railroads the right to appeal to the courts for a review of rates which they considered unjust.

Many of the Granger laws were ephemeral. Conceived in a spirit of bitterness and even of hostility to the railroads, it cannot be questioned that they overshot the mark and reacted adversely on the interests of the railroads and of the public. Their chief services lay in bringing about an enunciation of first principles, in blazing a path in the unknown and untried field of railway legislation, and in opening the eyes of the railroad executives to the possibilities of the future.

The conflict of jurisdiction. Conflict between state and federal control was inevitable as soon as the railroads ceased to be short, disconnected lines only and became the carriers of commerce between the several states. The United States Constitution gives to the federal government control over

foreign and interstate commerce, while each individual state has authority over the commerce originating and having its entire course and destination within its boundaries. The difficulty of fixing rates for intrastate trade without such action having an influence on the interstate rates, or the reverse, is quite apparent.

In the Granger cases it was decided that, in the absence of federal legislation on the subject, the state had the power to prescribe rates to regulate intrastate traffic, even though such action might indirectly affect interstate commerce. But the Wabash case, decided in 1886, established the principle that, even though the federal government did not choose to exercise its power, it nevertheless possessed sole authority to regulate interstate commerce. Since it was inconceivable that interstate commerce should be left to the arbitrary control of the railroads, particularly since about three-quarters of the traffic of the roads in 1886 was between the several states, Congress was forced to enact its first comprehensive regulatory legislation — the act to regulate commerce of 1887.

With two authorities determining rates within a state — the federal authorities in control of interstate rates and the state authorities in control of intrastate rates — the question of the dominant authority had to be settled. In case of conflict should intrastate or interstate rates rule? After some years of uncertainty and some tentative steps in the direction of greater control by federal authority, the final step, short of complete control over intrastate rates, was taken in the transportation act of 1920, which states that where a state-made rate discriminates against interstate traffic, the Interstate Commerce Commission shall prescribe a rate to remove this discrimination, which rate shall be binding, "the law of any State or the decision or order of any State authority to the contrary notwithstanding."

Federal regulation. Commencing with the interstate commerce act of 1887 the federal government has given constant attention to the problem of railroad regulation. Space does

not permit us to give in detail the provisions of this act, nor of the many acts which have been passed from time to time modifying the original act and extending the scope of government regulation; we shall have to confine ourselves to a very brief description of the more important of the regulatory powers at present exercised by the federal government.

In the first place, the Interstate Commerce Commission was created as a supervisory and regulatory body with authority over interstate and foreign commerce by rail or by rail and water. The commission consists of eleven members appointed by the President with the advice and consent of the Senate for terms of seven years. It has the right of investigation and report, either on its own motion or on the complaint of a shipper. It can compel witnesses to testify and to produce books and papers; it can order unlawful practices to cease, and if necessary it can apply to the federal courts to issue an enforcing order to the offending roads.

The control of the commission over interstate rates is, subject to review by the courts, complete. It may prescribe both minimum and maximum rates for interstate traffic and, as pointed out above, it may where necessary set rates which apply to intrastate traffic as well. The rates so laid down are to be just and reasonable, and in its determination of rates the commission, among other considerations, is required to take into account the effect of the rates on the movement of traffic, the need of adequate service at the lowest cost consistent with the furnishing of such service, and the need of revenues sufficient to enable the carriers under honest and efficient management to provide this service.

By the transportation act of 1920 the exclusive right of controlling new issues of rail securities was given to the commission. State control had proved itself in many cases to be arbitrary and discriminatory, and control by the federal authorities was adopted to ensure uniformity between the different railroads and to prevent the financial manipulation which was possible under the lax laws of some states. Power was given to

determine whether the purpose to which it was proposed to apply the proceeds of the issue was "reasonable, necessary, and appropriate" for the proper performance of public service. If an application is approved, the commission may make such terms or conditions governing the issue as it deems wise.

Exceedingly broad powers with respect to car service were at the same time conferred on the commission. It may regulate the movement of cars and locomotives, compel the purchase of all classes of equipment, determine the compensation which one road shall pay for the use of the equipment of another, compel the common use of cars, locomotives, and terminals, and in cases of emergency issue priority orders or embargoes on traffic. The extension or abandonment of service facilities is likewise under its control; no new railroad may be built, no existing road may be extended, nor may any line be abandoned without the consent of the commission.

The railroad crisis. In the decade of general prosperity following the close of the World War the railroads were conspicuous because of their failure as a whole to show satisfactory financial returns, and in the period of depression after 1929 the financial condition of many railroads became desperate, so desperate in fact that many were saved from bankruptcy and reorganization only by large loans from the Reconstruction Finance Corporation. In 1932 only a small minority of the Class I railroads in the United States earned a net income, a large majority of them did not earn fixed charges, and a considerable minority even failed to cover operating expenses. The rate of return earned on property used by the railroads as a whole, which had run between 4 and 5 per cent from 1926 to 1929, was down to 1¼ per cent in 1932. A number of railroads were in the hands of receivers, and the railroads had sought and secured 280 million dollars in short-term loans from the Reconstruction Finance Corporation.

The causes of the distress of the railroads are many and can be mentioned only in general terms. In part the situation was

a continuation of highly unsatisfactory financial conditions in the pre-war period — the result both of financial mismanagement of some of the important railroads and of the commission's failure to sanction rate increases sufficient to offset the increase in costs which a rising price level was bringing. Government operation of the railroads during the war also brought about certain inevitable elements of disorganization which were not readily eliminated when the roads were returned to private management in 1920. Among the more immediate causes we may mention lack of efficiency in management and failure to effect available economies in operation and to modernize the services offered — a failure ascribed by some to the paralyzing effects of government regulation. There can be little doubt that the financial structure of some of the lines reflected unwise practices in the past and that the interest charges were beyond the normal capacity of the roads to meet. In such cases receivership and reorganization would seem the most prudent path.

More specifically the railroads have complained of new types of competition in transportation. Since 1920 there has been a considerable extension of pipe lines for the transmission of gas and oil, and the rapid development in the use of electric power has caused a decline in the tonnage of coal carried. Some lines have had to face what they consider unfair competition from river lines operating on inland waterways improved and maintained by large federal expenditures. Recent developments on the Ohio and Mississippi rivers have enabled river steamers to carry freight at rates which the railroads have claimed would be impossible without this indirect subsidy and which rail transportation cannot meet profitably.

But the type of competition which most seriously affects all railroads is that offered by the motor truck and the passenger bus, both in local and in long-distance services. Until 1936 the motor traffic, both passenger and freight, was under such regulation only as the individual states saw fit to apply to intrastate traffic, and it has moved on roads built by the states

from general funds raised by taxation. Hindered by no federal requirements as to service or safety and subsidized to the extent of having their right of way built and maintained for them, the motor vehicles offered a more flexible and cheaper service than the railroads and in many instances were able to divert to themselves the cream of the traffic.

The emergency railroad transportation act of 1933. The serious financial condition of the railroads as well as the obvious need of better coördination of the various means of transportation led Congress to enact in 1933 emergency legislation, intended to be in force for one year but extended by subsequent legislation to June 17, 1936.

The act provided for the appointment of a Federal Coördinator of Transportation and charged him with the immediate task of attempting to improve the condition of the railroads by devising measures of more economical operation. It was his duty also to attempt to reduce the fixed charges of the railroads by promoting the financial reorganization of their capital structures and to make studies of other methods of improving transportation as a basis for recommendations for legislation to be transmitted to Congress.

The staff of the Federal Coördinator was chiefly occupied with the investigation of the problem of waste in railroad operation and with recommending, for study and consideration by the railroads, means of achieving efficiency and economy by standardization of practices. Congress gave the coördinator the power to issue orders to the railroads having the full authority of orders of the commission, unless revoked by the commission on appeal by the railroads, but, by laying down the principle that economies should not be at the expense of labor, the law handicapped him seriously in his efforts to achieve economy. It was not permitted to reduce the number of laborers employed below that existing in May, 1933, except by the operation of natural forces such as death, retirement, or resignation, nor could a laborer's pay be cut below what he was receiving at that time. If in the interest of economy it

became necessary to shift labor from one point to another, the financial burden was to be borne by the carrier concerned.

The office of the Federal Coördinator ceased to exist on June 16, 1936, and the railroads again were at liberty to effect economies by reducing the number or the pay of railroad laborers, but it is interesting to note that in the spring of 1936 an agreement was reached between the carriers and the representatives of labor whereby laborers adversely affected by such economies would be given appropriate compensation.

The emergency act amplified somewhat the powers of the Interstate Commerce Commission. The commission was given jurisdiction over railroad holding companies, with power to require that the stock holdings of such a company should be deposited with a trustee. Its authority to approve or to disapprove of mergers and consolidations was enlarged. It was instructed not to approve any loan to a carrier by the Reconstruction Finance Corporation if in its opinion the public interest would better be served by the financial reorganization of the carrier.

The motor carriers act. The Federal Coördinator of Transportation recommended to Congress the enactment of legislation placing the motor carriers and water carriers under federal control. Thus far Congress has not taken action on the bill for the regulation of water carriers, but the motor carriers act became a law on August 9, 1935. By this act all motor buses and trucks engaged in interstate commerce, with a few exceptions, as well as brokers who arrange for transportation but do not themselves provide it, became subject to the regulation of the Interstate Commerce Commission.

The powers of the commission over motor carriers are similar to those which they have been exercising over railroads. To operate in interstate commerce a common carrier must secure a certificate of public convenience and necessity from the commission, while a contract carrier must secure a permit, and a broker a license. Rates charged must be just and reasonable and must not be discriminatory, and to enforce

this provision the commission is given power to prescribe both maximum and minimum rates. The commission may require proof of financial stability and financial reports; it may enforce uniform accounting methods; it may assume control of consolidations of companies and issue of securities. It has the right to determine the qualifications to be demanded of operators, to specify maximum hours of work, and to prescribe such minimum standards of safety as it thinks the public interest requires.

The act does not contemplate exclusive federal control, for provision is made for the creation of joint state and federal boards. Furthermore complete control of intrastate traffic is left in the hands of state commissions, although it is highly probable that this cannot remain unaffected by standards set up for interstate traffic.

The regulation thus provided does not touch upon one of the objections of the railroads to the motor traffic — the use of highways supported by general taxation — but this is scarcely within the province of the federal government. Nor is this act intended to bolster up rail finances by crushing motor transportation facilities, for there is no intention of bringing the rates up to the level of the railroad rates. In determining rates which are just and reasonable the commission is to give due regard to the inherent advantages of motor transportation. It does however aim to achieve a greater degree of stability in transportation by eliminating the competition of the financially irresponsible, fly-by-night concerns and by forbidding disruptive price cutting, and it aims to make the highways safer by insisting on uniform standards of equipment and operation.

Labor legislation. The far-reaching effects on the economic welfare of the nation of a tie-up of railroad service resulting from labor disputes have impelled Congress from time to time to pass legislation setting up machinery for the settlement of labor disputes or for improving the condition of the worker. It is interesting to note that as far back as 1888

Congress provided for voluntary arbitration in interstate railway labor disputes. From that time on various plans of conciliation, mediation, or arbitration have been put into effect. The most recent of these acts was passed in 1934 and is known as the Dill-Crosser railway labor act. This creates a National Railway Adjustment Board with four divisions, each equally representative of the roads and the workers and having also a neutral member. In addition there may be established regional or system boards of adjustment if the railroads and the employees desire to set them up. If these voluntary boards or other agents of conciliation fail to secure a settlement of a dispute over the interpretation or application of existing agreements as to rates of pay, rules, or working conditions, the case goes to a division of the National Railway Adjustment Board for consideration and award. The board of mediation of five members, established by the act of 1926, is supplanted by a new National Mediation Board of three members to handle disputes involving changes in agreements as to rates of pay, rules, or working conditions.

This act marks also an important step in the process of unionization of railroad labor. It specifically prohibits any carrier from providing financial assistance from funds of the carrier to any union whatever. The carrier is also prohibited from interfering in any manner whatsoever with employees in joining or in refusing to join any organization or union, and the act further provides that the choice of representatives of any craft for the purpose of concluding a wage or other contract shall be determined by a majority of the employees voting. In other words, the right of collective bargaining by majority representation is clearly guaranteed, and financial contribution by the company to a union of its employees is forbidden. The result has been a rapid movement in the direction of organizing national unions where these previously have not been found.

The same session of Congress enacted a law making retirement allowances for railroad employees mandatory, and re-

quiring compulsory contributions from the railroads and from the workers to be pooled in a fund deposited in the United States Treasury. In May, 1935, the Supreme Court declared this act unconstitutional. New legislation, passed in 1935, proved ineffective, and Congress finally passed the railroad retirement act of 1937, which is now in effect. This provides annuities for laborers, engaged on the railroads and in certain auxiliary services, who have reached the age of 65, or who have become totally and permanently disabled and have completed 30 years of service, or who, being at least 60 years of age, have had 30 years of service or have become totally and permanently disabled. The annuity for any individual is determined both by his years of service and by his average monthly compensation and is paid from the railroad retirement account set up in the Treasury by virtue of another act also passed in 1937. This act provides for an income tax on employees and an excise tax on employers. The two taxes have the same base; *i.e.*, the compensation received by or paid to employees, and the same rate, which begins at $2\frac{3}{4}$ per cent of the monthly compensation not exceeding $300, remains at this figure through 1939, and thereafter rises gradually to $3\frac{3}{4}$ per cent after December, 1948. Both taxes are collected in the first instance from the employer.

Conclusions. In the course of sixty years the policy of the federal government has changed from one of indifference to one of minute regulation of most phases of railroad management. Early regulation stressed the necessity of maintaining competition between railroads; later policy has recognized both the futility and the undesirability of trying to maintain competition in rates and has emphasized competition in service. Regulation however has not as yet brought about a satisfactory solution of the railroad problem.

The outcome of the present situation is not clear. Railway revenues have improved somewhat since 1932, but the financial condition of the railroads as a whole is still unsatisfactory. Many railroads are in the hands of the courts, and others are

in default in payment of interest and taxes. This deficiency in income has inevitably brought about a curtailment both of railroad employment and of expenditures for maintenance of ways and structures and for replacement of equipment. Although existing equipment is adequate to handle the present volume of traffic, a revival of business activity would find the roads suffering from a very serious shortage. Further it is authoritatively stated that large investments are needed for modernizing equipment and shops, not simply to place the roads in position to handle more traffic, but to enable them to handle present traffic at lower cost and with greater expedition.

A special committee of the Interstate Commerce Commission made a report in 1938 on measures of immediate relief for railroads, but despite continued discussion in Congress of the urgency of an immediate and comprehensive attempt to solve the railroad crisis, no program to effect this has as yet been adopted.

The public has a vital interest in a successful solution of this problem. As of December 31, 1936, the total railway capitalization was $21,961,035,000, of which about 12 billions were funded debt and nearly 10 billions stock. The annual reports of the railroads showed that there were 855,261 stockholders in 1936 and that approximately 65 per cent — $6,617,000,000 — of the railway bonds were held by insurance companies, banks, endowed educational institutions, and foundations. The holdings of life-insurance companies alone amounted to $3,267,000,000. As investors therefore the public is immediately concerned with the financial problem of the railroads.

As shippers the public is interested in the maintenance of an adequate, efficient, and economical transportation system. Inasmuch as the railroads now are, and so far as one can see will continue to be, the backbone of the transportation system, anything which threatens their continuing effectiveness is a matter of grave concern. The railroads are also large purchasers of equipment, particularly from the heavy in-

dustries, and large employers of labor, so that any decline of railway purchases in the goods and labor markets must have serious repercussions on related industries.

It is in consideration of the far-reaching consequences of distressed conditions in the railroad industry that agitation for government ownership and operation of railroads secures much of its support. Even though it be true that the present plight of the railroads is in part the result of so much regulation that initiative in management has been weakened, unregulated railroad operation is a practical impossibility in the United States. Politically the only alternative to continued regulation is government ownership and operation. Some time ago the Federal Coördinator expressed his conviction that this was an inevitable step in the future. Railroad executives and many others disagree emphatically. They fear that government operation would be even more rigid and unprogressive than has been private operation under government regulation; they fear the political power which such a step would give to railway labor, as well as the precedent which might be set up for still further extensions in the scope of government activity. Nevertheless it must be recognized that continued failure to find a solution to the present problems of the railroads is likely to keep alive and strengthen the demand for government ownership and operation.

EXERCISES

1. It has been estimated that about 60 per cent of a railroad's expenses are independent of the volume of traffic, while 40 per cent are dependent on the volume of traffic. Assume the following conditions of cost and return for a railroad concern:

Operating income	$102,000,000
Fixed costs	60,000,000
Variable costs	40,000,000
Net income	2,000,000

(*a*) How would a 10 per cent increase in traffic affect the railroad's net income?

(*b*) How would a 10 per cent decrease in traffic affect the railroad's net income?

2. Of what importance is the distinction between fixed and variable costs to the problem of railroad rates? If a railroad between Pittsburgh and Chicago is already in existence and trains are running, what is the lowest rate at which the railroad could afford to accept additional traffic? Would it be feasible for the railroad to accept all traffic at such rates?

3. How does the presence of joint cost affect the determination of railroad rates? Why do not railroads charge uniform ton-mile rates?

4. A railroad runs from New Orleans through El Paso to Los Angeles. Is there any justification for charging a lower rate per ton for shipments from New Orleans to Los Angeles than for shipments from El Paso to Los Angeles?

CHAPTEI

Interregio

The bases of interregional
Trade, in its simplest form, cor
goods or services in the same ı
volves more; namely, the exch
places. It is this sort of trade w
ing and important economic
ciples of the simpler form of tr
more complex, we may confi
ject of interregional and inter

So many popular fallacies h
of interregional trade that it i
care into the nature of the
division of labor and interre
articles is readily enough expl
way could the community
goods. Diamonds, emeralds,
could not be enjoyed by Am
in through trade. The resi
could not have fresh fruits a
of the agricultural hinterlan

The real problem of trade l
commodities which apparen
and which are nevertheless ı
regions. The people of New
automobiles instead of leanin
United States could produce
only a part. In fact the liter
article at home can seldom b
port. The real reason is that l
industry of other communities

l in their production at home.
· things it needs to those parts of
ere each commodity is produced
nds the rest of the country and
r those particular commodities
st. So each region, because of
resources and energies to the
t is best adapted as evidenced
cost.

r causes. Many factors com-
ıe region can produce a good
The rich soil and favorable cli-
to grow sugar at a lower cost
e exceptional quality of the
Mountains, India, and Brazil
vantage in the extraction of
manganese used in the steel
imported, in spite of the fact
rves of low-grade manganif-
he low cost of production is
from large-scale production,
and the minute subdivision
s of iron and coal of England,
industrious population, large
estment, and the application
tion, have given the British
of many articles. Numerous
which the products of a region
ority in natural resources, by
laborers, by relatively great
economies in the organization
to the reader. The essential
ctive advantages are enjoyed
se lead to regional specializa-
ıange of the surplus products

The gain from specialization and trade. It will doubtless be evident by now that specialization by each region in the production of the goods for which it is best fitted must increase enormously the total of the world's production and correspondingly add to the goods available for the satisfaction of human wants, but the situation will be clearer and the causal relation brought more out into the open if we inquire somewhat minutely into the economic effects of interregional trade. For this purpose we may set up a hypothetical case which, in spite of its obvious artificiality, will serve to clarify the principle in which we are interested.

Let us suppose that country A and country B can produce either wheat or corn, but that A has an advantage in the production of wheat, while B has an advantage in the production of corn. To make the example concrete, let us say that with the expenditure of a unit of labor and capital (composed of so many hours of labor, so much machinery, fertilizer, etc.) A can produce either two bushels of wheat or one bushel of corn, while B with the same expenditure of effort (one unit of labor and capital) can produce either one bushel of wheat or two bushels of corn. Then if A and B each spends two units of labor and capital and each produces both corn and wheat the result will be:

A	2 bu. of wheat + 1 bu. of corn
B	1 bu. of wheat + 2 bu. of corn
Total	3 bu. of wheat + 3 bu. of corn

If each country spends its two units in the production of the good in which it has a real advantage the result will be as follows:

A	4 bu. of wheat
B	4 bu. of corn
Total	4 bu. of wheat + 4 bu. of corn

In this case there is a gain in total product of one bushel of wheat and one bushel of corn, but neither country wants to be without the other grain. Rather than do without corn

entirely, A would be willing to spend one unit in producing a bushel. Let us then further assume that A exchanges wheat for corn with B at the rate of one bushel of wheat for one bushel of corn.[1] If two bushels are sold, each country would have wheat and corn, as follows:

A	2 bushels of wheat + 2 bushels of corn
B	2 bushels of wheat + 2 bushels of corn

Comparing this table with the first one, we see that A has gained a bushel of corn and B has gained a bushel of wheat as the result of specializing in production and exchanging the surplus.

It may however happen that a country finds itself in the unhappy position of having no natural advantages. Its soil is sterile, nature has not endowed it with any mineral resources of prime importance, and its population is lethargic; yet we find the country has an export trade. On the other hand, there are countries which are so favorably situated that they have advantages in many different lines of production; yet they regularly import certain goods from other countries which are actually inferior to them in those lines. This situation, as to the practical truth of which there can be no possible doubt, may seem anomalous if our previous example has led us to the erroneous conclusion that each country will necessarily produce for itself everything in the production of which it has an advantage, with the corollary that the country which has no superiority in any line can have no production at all.

To explain the situation let us assume two other countries, C and D. Country C with one unit of labor and capital can produce either fifteen yards of cloth or one ton of steel. Country D with one unit of labor and capital can produce ten yards of cloth or half a ton of steel. C has an advantage in the production of both cloth and steel, but its advantage in the production of steel is greater. Compared with C, D has

[1] In these illustrations we ignore the transportation charges.

no advantage in the production of either commodity, but its "least comparative disadvantage" lies in the production of cloth. If each country devotes one unit of labor and capital to the production of cloth and two units of labor and capital to the production of steel, the amounts produced by each will be as follows:

C with 3 units	15 yards of cloth + 2 tons of steel
D with 3 units	10 yards of cloth + 1 ton of steel
Total	25 yards of cloth + 3 tons of steel

If however with the same expenditure of effort as before C specializes in the production of steel, in which its comparative advantage is greater, and D specializes in the production of cloth, in which its comparative disadvantage is less, the amounts will be as follows:

C with 3 units	3 tons of steel
D with 3 units	30 yards of cloth
Total	30 yards of cloth + 3 tons of steel

From the standpoint of total production this specialization has brought about a gain of five yards of cloth, but this is of no particular interest to either country unless the terms of exchange are such that each country can gain thereby. C would be unwilling to sell steel for less than fifteen yards of cloth a ton, since it would be more profitable to manufacture the cloth at home; D, for the same reason, would be unwilling to pay more than twenty yards of cloth for a ton of steel. A rate of eighteen yards of cloth for one ton of steel falls within these limits, and we will assume that this is fixed and that C sells D a ton of steel for eighteen yards of cloth. Then we find, after the exchange has been effected, the following situation:

C	18 yards of cloth + 2 tons of steel
D	12 yards of cloth + 1 ton of steel

A comparison of this with the first table will show that C has gained three yards of cloth and D has gained two yards

of cloth. In spite of C's greater productive advantage in both industries, it is worth its while to specialize and exchange products with D. Of course the people of D still have less to show for their efforts than the people of C, but they would have even less if they were unable to specialize on cloth and get their steel by exchange.

The law of comparative advantage. From this example may be drawn an important generalization. It is to the advantage of a country or region to devote its energies, not to all the lines of production in which it may have superiority, but to those in which its superiority is greatest, provided trade gives it the opportunity to secure other goods from abroad. Likewise the country which has no line of superiority will find it advantageous to devote its energies to those lines in which its inferiority is least marked, provided the opportunity of trade with other regions is open to it. This generalization is called the *law of comparative advantage* and is stated in formal language as follows: *A country gains by concentrating its energies in the production of those goods in which it has the greatest comparative advantages or in which it has the least comparative disadvantages.*

Our conclusion is that regional specialization and interregional trade promote economy in the use of the productive forces of society. By utilizing the most fertile soils and the richest mines, by turning labor and capital into the industries for which each region is best suited, society as a whole has a greater quantity of consumable goods to share among its members, and each region is enabled to secure a greater quantity of consumable goods than otherwise would be possible.

Since the gain from trade consists in the increased command over consumable goods, it follows that imports are the aim and end of commerce. We export in order that we may import. This conclusion is somewhat at variance with popular belief, but if it be true that commerce exists to serve men's wants, its benefits are to be measured by the command it gives over consumable goods. Our exports are of importance only in so far as they provide an easier means of securing

the goods imported than by making them at home, and the greater the quantity of imports which we can obtain with a given quantity of exports, the more fortunate our position and the greater the national income. A falling off in the export trade is to be considered just cause for public alarm only because it may indicate that our buying power in foreign markets has decreased and therefore our command over foreign goods is less than formerly.

These conclusions were derived without reference to political boundaries, and they apply as well to regional specialization within the confines of a single country as to specialization between countries which are politically distinct. The gain which comes to the United States from purchasing the fine woollens of England with wheat is of exactly the same character as that which the people of New England experience when they buy western foodstuffs with their manufactured goods.

Diminishing returns and the law of comparative advantage. In every country hundreds and thousands of different articles are produced. Why, if the law of comparative advantage really holds good, should not each country, or each section of each country, concentrate on the production of one article only, and secure all the other goods it needs by exchange?

In partial answer to this question it may be said that the existence of a comparative advantage in the production of any commodity depends to some extent upon the quantity produced. There is a limit to the economies of large-scale production; sooner or later all industry operates under conditions of increasing cost. A rising marginal cost of production is an indication that the comparative advantage in the production of additional units of the good is becoming less. The United States, for example, has such decided advantages in the production of wheat that it regularly has a surplus of several hundred millions of bushels for export, but if it were to attempt to supply the entire world with wheat the cost per bushel would rise far above that which would obtain in

other wheat producing countries or even in some at least of the countries which now import wheat. Rather than pay the price which in such circumstances would be necessary to remunerate the American farmers for their efforts, foreigners would buy wheat from other sources or grow wheat at home. The law of diminishing returns therefore sets limits to specialization, and we find few products of any consequence in the world's commerce which are produced in one region exclusively, just as we find few important regions with only one industry.

Transportation and the law of comparative advantage. To complete the answer to the question raised above we must also take into account the factor of transportation. In our hypothetical examples we ignored the costs of transportation, but in actual life they offer one of the greatest obstacles to the development of regional specialization and trade. English manufacturers may be able to produce certain grades of brick at a lower cost than can American manufacturers, but if the cost of shipping them to New York is so great as to raise their price above that of similar domestic bricks, the advantage of the English manufacturers so far as the New York market is concerned is without significance. There are many goods of which this is true, goods which are commonly produced in almost every locality, because the cost of transporting them from one place to another is so great as to wipe out entirely the advantage which any one locality may have. Nearness to market therefore gives an advantage in production which may outweigh positive disadvantages, as is shown for example by the fact that common bricks are almost everywhere a local product, seldom brought in from any great distance, even though other places may enjoy advantages in their manufacture.

The control of trade. Every producer of goods naturally seeks to sell his product in the market where he obtains the highest price, be it near at hand or far away. The consumers just as naturally endeavor to make their purchases where

goods may be obtained most cheaply. All the agencies of trade and transportation are enlisted in the effort to aid producers and consumers in the accomplishment of their respective purposes; every class of middlemen is actuated by the same motives as the other producers and the consumers and strives to buy in the cheap market and sell in the dear market. There is a tendency therefore for goods to move from places where their prices are low to places where their prices are high. By thus reducing the supply in one market and increasing it in another, the price in the one is raised and in the other is lowered, till the prices in different markets are brought into equilibrium.

Equilibrium of prices does not mean equality of prices. The movement of any commodity from the low-priced to the high-priced market reaches a limit when, as the result of such movement, the difference between the two prices is brought down to the cost of shipping the commodity from the first market to the second. Thereupon the movement ceases, price equilibrium having been reached. If transportation could take place instantaneously and without cost we would find the price of each commodity the same in all markets. As it is, equilibrium is attained when the prices of a commodity do not differ by more than the cost of transportation. This principle may be stated in the following law: *In the absence of restraints on trade, the prices of any commodity in two markets tend to differ by an amount not greater than the cost of transporting the commodity from the market of lower price to that of higher price.*

Thus price and the cost of transportation control the flow of goods from one place to another. There will be trade so long as there would otherwise be a difference in the prices of the same commodity in two markets greater than the cost of transportation between them. Where the difference is less than this there is no interregional trade, as is the case of many commodities of which the supplies are relatively local; bricks for example. The ability to specialize to advantage is evidenced by a relatively low cost of production, which enables

the producers of one region to undersell those of other regions, and the opportunities for the extension of regional specialization and trade are brought to light by differences in the prices of the commodity in the various markets. Under competitive conditions price serves as an automatic control over regional division of labor and interregional trade.

The restraints which may prevent trade and so prevent the attainment of price equilibrium are numerous. Frequent examples have already been encountered in our study, and we shall later give attention to one of the most important, the customs tariff.

The foreign trade of the United States. This is not the place for an extensive survey of the foreign commerce of the United States, but the reader will doubtless be interested in some broad knowledge of its size, its general character, and its development. The table below gives the total value of exports and imports for a number of years and will serve to illustrate the growth of American foreign commerce.

This table is interesting in showing the tremendous increase

FOREIGN COMMERCE OF THE UNITED STATES

(In millions of dollars)

	Exports (domestic and foreign)	*General imports*	*Excess of exports (+) or imports (−)*
1800	70.9	91.0	−20.1
1820	69.7	74.4	−4.7
1840	123.7	98.2	+25.5
1860	333.6	353.6	−20.0
1880	835.6	667.9	+167.7
1900	1,394.5	849.9	+544.6
1910	1,744.9	1,556.9	+188.0
1920	8,228.0	5,278.5	+2,949.5
1925	4,910.0	4,227.0	+683.0
1928	5,128.0	4,091.0	+1,037.0
1930	3,843.0	3,061.0	+782.0
1932	1,611.0	1,323.0	+288.0
1934	2,133.0	1,655.0	+478.0
1936	2,456.0	2,423.0	+33.0
1938	3,094.0	1,961.0	+1,134.0
1939	3,177.0	2,318.0	+859.0

in exports and imports during the period as a whole, together with the very material decrease in foreign trade during the recent years of depression. The reader will however avoid drawing hasty conclusions with respect to the growth of trade from figures which give money values only, for there have been marked changes in price levels during the period. It is quite possible for the physical volume of trade to decline from one year to another, while the value of the trade increases. Professor Fisher gives the following interesting example of this: [1] "The money value of car wheels exported from the United States in one month was $12,000 and in a later month $15,000, from which fact we might infer that the quantity of these exports had increased. But the *number* of car wheels exported in the first of these two months was 2,200, and in the second only 2,100, showing a decrease. The price had increased faster than the number had decreased."

INDEXES OF CHANGES IN QUANTITY, AVERAGE UNIT VALUES, AND TOTAL VALUE OF GENERAL IMPORTS AND EXPORTS OF UNITED STATES MERCHANDISE [2]

(1923–25 average = 100)

Year	*Exports of United States merchandise*			*General imports*		
	Quantity Index	Average Value Index	Value Index	Quantity Index	Average Value Index	Value Index
1913	84	65	55	66	70	46
1920	116	156	181	88	155	136
1925	107	100	108	104	105	109
1928	128	88	113	115	92	106
1930	109	78	85	111	71	79
1932	69	51	35	79	43	34
1934	74	63	47	86	50	43
1936	82	66	54	118	54	63
1938	105	65	68	94	54	51
1939	110	64	70	108	55	59

[1] I. Fisher, *Elementary Principles of Economics*, New York, 1912, p. 19.

[2] For a description of the methods by which these index numbers are derived the reader is referred to pp. 24 and 38 of *Foreign Trade of the United States, Calendar Year 1934*, Trade Promotion Series No. 162, Bureau of Foreign and Domestic Commerce, Washington, 1935.

The Bureau of Foreign and Domestic Commerce has for a number of years prepared index numbers showing the changes in quantity, average unit values, and total value of the exports and imports of the United States. The reader will find it interesting to compare the changes in these index numbers, which are given for certain years in the table on page 619, with the value figures given in the preceding table.

CHAPTER XXXII

Foreign Exchange and the International Balance of Payments

The problem. In our first approach to the study of foreign exchange it will be convenient to eliminate for the time being certain complicating factors of recent years. The difficulties which confront the student of contemporary foreign exchange arise partly from the abandonment of the gold standard by most of the countries of the world and partly because of a variety of controls, exercised by governments or by central banks in pursuance of governmental policy, affecting exchange operations or transactions which would involve the exchanges. We shall consider some of these problems in the next chapter. but for the present we will assume (1) that the countries with which we are concerned are on the gold standard, implying thereby that all forms of money within each country are ultimately redeemable in the country's standard gold coins and that the right of coinage, export, and import of gold may be freely exercised by private persons, and (2) that there are no formal exchange controls which interfere with the operations of a competitive market in foreign exchange. Such were the conditions which existed with respect to the exchanges of all the leading industrial nations in the pre-war period, and it is quite possible that they may again be attained.

Financing exports. Our first problem — one which is much the same whatever the monetary systems on which the exchanges are based — is to investigate the mechanism which is available for financing exports and imports of goods. The simplest approach to this is by way of the study of typical examples of export and import transactions. Let us therefore assume that Richards & Green of New York has sold Harris & Company, Ltd., of London a certain number of bales of

cotton for £1,000. Now it is evident that the ordinary forms of currency used in domestic trade are not available for payment in this case. The universal currency for international payments is the bill of exchange, and we may assume, what would almost always be true in such a case, that the terms of the sale have called for a bill of exchange to be drawn by Richards & Green on Harris & Company, Ltd. The bill will be in the following form:

No. 726 New York, December 29, 1929

Pay at sight to the order of the New York Bank
.....One Thousand Pounds Sterling.....

(signed) Richards & Green

To Harris & Company, Ltd.
London, England

The bill of exchange is an instrument of such importance that we must pause here to formulate its precise definition, which is as follows: *A bill of exchange is an unconditional order in writing, signed by one person (called the drawer) directing a second person (called the drawee) to pay a third person (called the payee) or his order a certain sum of money at some definite time in the future.* It will be noted that the example given above conforms with this definition.

When Richards & Green ships the cotton it secures from the steamship company, or possibly from a railroad company if the shipment is to be made from an inland warehouse, a bill of lading, which not only shows that the company has received a certain number of bales of cotton and has agreed to transport them to England, but also serves as the evidence of ownership in the cotton. Anyone in possession of a properly endorsed bill of lading can secure the cotton from the steamship company, and it can be secured in no other way. Marine insurance must be taken out on the shipment in order

that the owner may be protected if the cotton suffers loss or damage during the sea voyage.

Richards & Green next takes the bill of exchange, with these two documents (bill of lading and insurance policy) and possibly others attached, to its bank, the New York Bank, for sale. The bank knows nothing about the reliability of Harris & Company, Ltd., but it knows that Richards & Green is an old, well-established house with a reputation for financial straightforwardness. If it buys the bill of exchange it will acquire a credit instrument bearing the signature of the responsible officer in the concern and having the same standing as a promissory note signed by the same officer. If Harris & Company, Ltd., fails to meet the bill when presented for payment the bank can enforce payment from Richards & Green. Furthermore the bank acquires possession of the bill of lading, which will give it the title to the cotton. Now cotton is a staple article, not perishable, and commanding at once a definite price in all the important markets of the world. If the worst should happen and the bank could not collect from Richards & Green, it could sell the cotton and still not lose much. Relying therefore on the responsibility of Richards & Green and on the cotton as collateral security, it purchases the bill of exchange at the rate prevailing in the market at the moment for bills of that character. The rate, let us say, happens to be $4.84[1] per pound sterling, and Richards & Green therefore receives $4,840 in cash. The exporting firm has thus received payment for the cotton in American dollars immediately after shipment and has no further interest in the transaction, save for its contingent liability.

The bank now sends the bill, with the documents still attached, to the London Bank, with which it has correspondent relations, with instructions that it be presented at once to Harris & Company, Ltd., for payment and that the proceeds be credited to the account of the New York Bank. Harris & Company, Ltd., thereupon pays the bill in British currency

[1] The student will realize that this rate reflects conditions of 1929 when both Great Britain and the United States were on the gold standard.

and receives the bill of lading, which entitles it to claim the cotton from the agents of the steamship company.

We may summarize these stages in the financing of the export of cotton as follows: (1) Richards & Green draws a bill of exchange on Harris & Company, Ltd., for £1,000. (2) Richards & Green sells the bill of exchange (with documents attached) to the New York Bank for $4,840. (3) The New York Bank sends the bill and the documents to its correspondent in London, the London Bank. (4) The London Bank collects £1,000 from Harris & Company, Ltd., and surrenders the bill of lading. (5) The New York Bank's deposit account with the London Bank is credited by £1,000.

Financing imports. A similar example will aid us to visualize the several steps involved in the financing of an import. We assume that White & Company of New York has purchased £800 worth of cotton goods from R. J. Littlefield of London. It would be quite in accord with commercial practice for Littlefield to draw a bill of exchange on White & Company, thus reversing the process described in the previous section, but trade practice frequently places the responsibility of making payment, as well as of securing payment, on the American merchant. We may assume then that in this instance White & Company is required to place Littlefield in possession of £800 in cash or its equivalent in London. This it can do by purchasing from its banker a bill of exchange such as the following, often called a demand or sight draft.

No. 918 New York, January 5, 1930

At sight pay to the order of R. J. Littlefield
.....Eight Hundred Pounds Sterling.....
and charge the same to our account

(signed)
New York Bank

To
The London Bank

The New York Bank sells this bill of exchange to White & Company at the prevailing rate for bankers' demand drafts, which we will assume to be $4.87. The draft therefore costs White & Company $3,896. White & Company sends the draft by mail to Littlefield, who presents it at the London Bank and receives £800 in cash, which amount is debited to the account of the New York Bank.

The rôle of the banks in exchange. The part which the bankers play is that of middlemen. It would indeed be possible for the importers to buy bills of exchange from exporters and to remit these to their creditors. In the past this has been done, but today virtually all bills are, in the first instance, either sold to or bought from the bankers. The banker has a wide circle of customers importing and exporting every known ware and dealing with all the markets of the world, and he is able to buy and sell drafts in all currencies and in all amounts at a moment's notice. An importer of goods from China can secure a draft payable in Shanghai or Hongkong with no difficulty, and furthermore he knows and the Chinese exporter knows that the risk of loss inherent in a banker's draft is much smaller than in a draft drawn by one commercial house on another. Because of these advantages it is not surprising to find the banker at present the focal point of the foreign exchange market.

Middlemen always expect to sell what they buy and to make their profit by charging a higher price for the good than they paid for it. Bankers do not differ from other middlemen in this respect. Their willingness to buy bills of exchange and thereby to acquire bank credits abroad arises from the fact that they expect to be able to utilize these credits by selling their own drafts at a higher rate. It is the margin between the banks' buying rate and their selling rate which gives them their profit in exchange transactions.

The exchange market. The exchange rate responds in so sensitive a manner to the play of the forces of demand and supply, and there is such a vast quantity of transactions in

foreign exchange in New York and other centres, that it is frequently assumed that there must be an organized market similar to the stock and produce exchanges where the rate of exchange is determined. This is an error; there is in New York no formal organized exchange for dealing in foreign exchange. The bulk of the business is concentrated in the hands of a relatively few large banks and exchange houses. Each bank fixes the rates at which it will buy and sell exchange on all the financial centres with which it has dealings, but each house keeps so closely in touch with the activities of the other houses that there is seldom any material discrepancy between the rates charged by the various dealers on the same class of bills.

There is no official rate of exchange. Each bank posts the rates at which it is willing to buy and sell bills of exchange, but it frequently departs from its posted rates both on bills of large amounts and unquestioned character and on bills of dubious standing. In fact some bills will not be bought at all, but accepted for collection only. Since 1921 the Federal Reserve Bank of New York has been required to issue a daily report of the buying rate in New York for cable transfers on the important financial centres, and this possibly is the nearest approach to an official quotation. But it should be borne in mind that this rate is not a rate at which the Federal Reserve Bank itself is offering to buy cable transfers, but is one obtained from the regular dealers in exchange.

The rate of exchange: General principles. Thus far we have been concerned chiefly with the technical facilities which modern business has at its disposal for the transmission of funds from one country to another, but we are equally interested in the forces which determine the *rate of exchange*. This may be defined as *the price paid in the home country for one unit of the money of the foreign country payable in that country*. For example the rate of exchange in New York on London is the amount of American money that must be paid in New York for a bill

of exchange calling for the payment of one pound sterling in London.[1]

The first step in the study of the rate of exchange is to determine what is known as the *mint par of exchange*. This is a simple ratio derived by dividing the fine gold content of the standard unit of one country by the fine gold content of the unit of another. Italy and Rumania, for example, used to have gold coins containing the same number of grains of fine gold; the mint par of exchange between them was therefore one; *i.e.*, one lira was equivalent to one leu. Formerly the sovereign of Great Britain contained 113.00116⅔ grains of fine gold, and the gold dollar of the United States contained 23.22 grains of fine gold. The ratio was $\frac{113.00116\frac{2}{3}}{23.22}$ or 4.8665+. Hence the British sovereign contained 4.8665+ times as many grains of fine gold as the dollar, and the mint par of exchange was $4.8665+ per pound sterling. This means that if ten thousand gold sovereigns of full weight were to have been taken to the United States mint they would have contained sufficient fine gold to have furnished the gold content of $48,665+.

The rate of exchange however is not the same thing as the mint par of exchange. Occasionally the rate of exchange quoted in the market stands at the mint par, and when it does so we say that exchange is "at par." But this rarely occurs, and then only by chance. The mint par of exchange furthermore does not determine the rate of exchange; its significance lies, as will appear, in the fact that it helps to set limits to the fluctuations of the actual rates.

The rate of exchange is a price and, in common with competitive prices in general, is determined by demand and supply. The demand for bills of exchange comes primarily from

[1] Rates of exchange may equally well be expressed the other way around, as the amount of the foreign money payable in the foreign country which may be obtained in the home country for one unit of its money. Thus before the World War the rate of exchange on France was often stated as so many francs per dollar. This is not the usual method however.

the importers and others who are under the necessity of making payments in foreign countries. The supply comes primarily from the exporters and others who have money available to them in foreign countries. The bankers are the middlemen who, buying and selling, bring together the forces of demand and supply. In accordance with the general laws of price, the greater the demand for bills (other things being equal), the higher the rate of exchange; the greater the supply (other things being equal), the lower the rate. In terms of international trade, when imports are large compared with exports, the rate of exchange tends to be high; when exports are relatively large, the rate tends to be low.

The gold points. But the rate of exchange differs from most competitive prices in one important respect; between countries on the gold standard it can fluctuate only within definite limits. Having risen to the upper limit, the rate will stand there, no matter how great the demand, until changed conditions send it down. Having fallen to the lower limit, it can go no lower and will eventually rise. To make this clear let us assume that, because of the transactions of a certain short period, business men in America owe £50,000,000 for imports received and are owed only £25,000,000 for exports. By buying the bills of the exporters, American bankers can acquire credits in London of £25,000,000 and be in position to sell their own bills to this amount to the importers. But this will leave £25,000,000 of the debts of the American importers unpaid. As this discrepancy between the volume of bills which importers and others want and the volume offered to the bankers for sale begins to be manifest, bankers will tend to charge a higher price for their own bills of exchange in order to protect their foreign balances; and individual bankers, each one of whom will want to get as large a share of the possible credits abroad as he can, will bid against each other for the bills of exchange which are offered for sale by exporters. The rate of exchange will be forced up gradually, and the higher rate will tend to discourage some who want to

make payments and perhaps encourage others to anticipate the drawing of bills of exchange. But if the basic forces at work are strong enough the bankers will soon find their foreign balances exhausted, with importers still trying to buy bills of exchange. In this situation the bankers in America, if they are not to cease operations entirely, are forced to resort to other expedients to build up their foreign bank balances. They might acquire balances against which they could draw by borrowing from a bank in England or by exporting securities for which there is an international market, but, granted our assumption of a free gold market, they would in all probability export gold from the United States for this purpose.

At a mint par of exchange of $4.866 per pound sterling, gold equivalent to one thousand pounds sterling can be purchased in the United States for $4,866. It will obviously cost something to cover the shipment of gold — to take care of crating, freight, insurance, and the like. Assuming that these expenses amount to two and one-half cents per pound sterling, then the total cost of placing one thousand pounds in London would be $4,891. The bankers could therefore offer to sell bills of exchange drawn on banks in London at $4.891 per pound, and the importers who were under obligations to make immediate payment would have to pay this rate. True they could ship the gold themselves, but the banks can do this much more cheaply, and competition among different banks forces them to do so. This rate we call the *gold exporting point*, determined by the cost in New York of transferring one gold sovereign to London. It is the highest price which will be paid in one country for one unit of a foreign currency payable in the foreign country and, no matter how great the demand, bankers will continue to furnish bills of exchange to all comers at that rate so long as gold is available for shipment.

In like manner we may determine the *gold importing point.* Let us suppose that the supply of sterling bills of exchange— that is, bills drawn on London in pounds sterling— increases greatly, without any corresponding change in the demand.

As we have seen, a supply of bills of exchange arises when exporters and others having debts to collect draw bills of exchange against the foreign debtors and attempt to sell them to bankers, and a demand arises when importers and others having debts to pay to foreigners go to the bankers to buy bankers' drafts. In this case therefore bankers are likely to have large credit balances abroad for which there is no call, and if the situation continues they will refuse to buy bills except at rates which will permit them to transfer their foreign balances to replenish their dwindling domestic cash resources. The cost of shipping gold from England to America is slightly higher than in the other case, for in this instance the bankers lose interest during the time of shipment; whereas, when gold is shipped from New York, the bankers can sell drafts to go by the same steamer and so avoid tying up their funds. If we assume the cost of shipment to be three cents per pound sterling, bankers will be willing to buy bills of exchange at the rate of $4.836. With a draft of £1,000 a banker can obtain a thousand gold sovereigns, bring them to the United States, and sell them for $4,866. The cost of doing so being $30, he will recover $4,836, the amount he paid for the bill. This rate, $4.836 per pound sterling, is known as the gold importing point, and at this rate bankers will be willing to purchase an unlimited quantity of bills of exchange, no matter how great the supply.

The gold points then are determined with respect to the mint par of exchange and fix the limits within which the rate of exchange may fluctuate. If the rate rises to the upper gold point, gold will be *exported;* if it falls to the lower gold point, gold will be *imported.* The rise or fall in the rate is determined by the relation of demand and supply to each other, which in turn rests on the relation between our obligations to foreigners and the obligations of foreigners to us. If these were just equal, the rate would presumably be at or very near the mint par of exchange, and there would be no tendency for gold to flow at all; if at any time one greatly outbalances the other,

gold will flow out to create the necessary foreign bank balances, or gold will flow in as bankers find this the only way to utilize their foreign balances.

Indirect exchange. It is scarcely necessary to remind the reader that the demand for and supply of bills of exchange, which determine exchange rates and gold movements, relate to the total of all international transactions of the country concerned and not to the account between just two countries. If the latter were the case, there would be almost continual shipping in and out of large quantities of gold, since it is scarcely conceivable that the debits and credits of any two countries should exactly offset each other except as a rare coincidence. As a matter of fact, when the majority of the countries of the world were on the gold standard, international trade was always conducted on a clearing house basis, in accordance with which all the obligations of the residents of one country to residents of all other countries were offset against all claims against foreigners. There was no international clearing house, but the result was accomplished with remarkable certainty and smoothness through the foreign exchange markets of the world. To a great extent this is still true, although many barriers have been erected which impede the smooth operation of the system.

In this system London and sterling exchange played an important rôle. The leading position of Great Britain in world trade, the solidity and flexibility of her monetary and banking system, and the supremacy of London in international short-term and long-term lending as well as in other financial services gave to London a dominating position in international finance which only recently she has had to share with New York and Paris. These circumstances gave to sterling exchange the character of an international currency. For many years bills on London were in all markets a most acceptable form of exchange.

An illustration of the workings of indirect exchange, utilizing sterling exchange as an international currency, can be seen

in the trade between Great Britain, Brazil, and the United States. The American people normally import more goods and services from Brazil than they sell to her, they export more to Great Britain than they purchase from her, and the people of Great Britain sell more to the Brazilians than they purchase from them. If each country were to settle with each separately and without regard to its financial obligations to other countries, the United States would be required to send a large sum in gold annually to Brazil, Brazil in turn would have to settle in gold with Great Britain, and a third shipment of gold would be necessary from Great Britain to the United States. Or if gold could not be shipped, Americans would cut down their imports from Brazil until they balanced the exports to her, Brazil in turn would decrease her imports from Great Britain until they were in balance with her exports to Great Britain, and American exports to Great Britain would likewise be decreased. The second alternative would mean obviously a tremendous decrease in international trade, with unfortunate reactions on international division of labor. The obligations created by this three-cornered trade can however be settled much more simply. America's excess of exports to Great Britain gives bankers in the United States balances in London banks which can be utilized in drawing bills of exchange in favor of Brazilian exporters. In due course these drafts will be sold by the Brazilian exporters to bankers in Brazil and sent to London for collection. Brazilian bankers therefore have acquired balances in London banks against which drafts can be sold to Brazilian importers of goods from Great Britain for transmission to the British exporters. It is possible, though very unlikely, that the surplus balances acquired from American exports to Great Britain might be exactly the amount necessary to pay for the excess of imports from Brazil; it is more likely that Americans would still have a balance in London, which could be utilized to pay for an excess of imports from Colombia, let us say.

The existence of an exchange such as sterling exchange which is everywhere preferred is not an indispensable prerequisite to international clearing arrangements of the character illustrated above; American importers could remit drafts expressed in terms of dollars or in terms of Brazilian currency and drawn on the London banks if these were preferred in Brazil. The existence of a stable and generally acceptable form of exchange does however make for smoothness in international financing.

The balance of payments. We have seen that gold will move into or out of a country only when demand for and supply of bills of exchange are in such relation as to cause the rate of exchange to move to the gold import point or to the gold export point. Any transaction therefore which gives rise to a demand for or a supply of bills of exchange will have influence on the determination of the rate of exchange and consequently on the international movement of gold. There are innumerable transactions between individuals, corporations, and governments in one country and individuals, corporations, and governments in other countries which require the transfer of funds from one country to another and thus give rise to a demand for or a supply of bills of exchange.

As an example we may assume that a New York investment house has been authorized to sell an issue of bonds for the British government. In all probability the investment house would use the proceeds from the sale of the bonds to purchase from a New York banker a bill of exchange on a London banker drawn to the order of the British Treasury. It would thus add to the demand in New York for bills of exchange on London. When the interest on these bonds became due the British Treasury might authorize its agent in New York to draw a draft on the Treasury for the amount of the interest payments. To secure dollars for actual payment to the bondholders the agent would offer this draft for sale to a New York banker and thus add to the supply of bills of exchange in New York on London. An immigrant in the United States

might add to the demand for foreign bills of exchange by purchasing from a New York banker a draft payable to someone in the homeland.

It is unnecessary for us to examine each of the many international transactions — shipping, banking, insurance, personal services, and the like — performed for or by persons in a given country, but it is essential to realize that we cannot discover whether gold will be exported or imported by bankers until we have a complete picture of all such transactions. These may be arranged in the form of an international balance sheet. In the credit column would be found, for the period under consideration, exports of merchandise as well as all other transactions, such as the sale of securities abroad, services performed for foreigners, and the like, which could give rise to a supply of bills of exchange in the country in question and hence give its bankers credits in foreign banks. In the debit column we would include imports of merchandise and all other transactions, such as purchases of foreign securities, tourist expenditures abroad, services performed by foreigners, and the like, which could give rise to a demand for bills of exchange in this country and so permit its bankers to utilize its credits in foreign banks. Since it is the effect on gold movements that we are now studying, we do not include among the credit and debit items the export or import of gold itself. If these total credits exceed the total debits, it means that the volume of bills offered to bankers exceeds the volume demanded of them, the rate of exchange will be at the gold import point, and gold will be imported to settle the balance. This situation is known as a *favorable balance of payments.* On the other hand, if total debits exceed total credits, conditions of demand and supply will push the rate of exchange to the gold export point, and gold will flow out of the country. This is known as an *unfavorable balance of payments.*[1]

[1] The terms "active" and "passive" balance of payments are frequently used nowadays in place of the terms, "favorable" and "unfavorable," which tradition has handed down.

The international balance sheet of the United States. The picture of the international financial relations of a great nation can be made clearer by a glance at the following table, derived from figures compiled by the United States Department of Commerce, showing the international balance sheet of the United States for the year 1928.[1]

ESTIMATED BALANCE OF INTERNATIONAL PAYMENTS OF THE UNITED STATES, 1928

(In millions of dollars)

Commodity trade	*Credits*	*Debits*
Merchandise exports and imports	5,129	4,091
Silver	87	68
Ship chandlering, sales of bunker coal and oil, etc.	98	62
Unrecorded parcel-post shipments	20	20
Adjustments in above items	....	256
Total commodity trade (as adjusted)	5,334	4,497
Miscellaneous invisible items		
Freight payments and receipts	143	227
Tourist expenditures	168	782
Ocean-borne passenger traffic	89	....
Earnings of long-term private investments	817	252
Earnings of short-term interest and commissions	65	107
Immigrant remittances	28	217
War-debt receipts of United States Treasury	210	....
Other United States government receipts, payments, foreign representations here	53	110
Missionary and charitable contributions, etc.	...	52
Motion-picture royalties	70	6
Insurance transactions	80	70
Miscellaneous	46	53
Total miscellaneous items	1,769	1,876
Movement of private long-term capital		
New American investments abroad (net)	....	2,070
Reduction of previous American investments abroad	853	....
New foreign investments in the United States	1,704	....
Reduction of previous foreign investments in the United States	....	1,223
Total private, long-term capital items	2,557	3,293
Movement of short-term capital		
Net change in international banking accounts	....	226

[1] The balance sheet for 1928, rather than one of a later year, is considered more useful for purposes of illustration since it antedates the months in 1929 when speculation was at its height, the years of depression, and the abandonment of the gold standard by various nations.

ESTIMATED BALANCE OF INTERNATIONAL PAYMENTS OF THE UNITED STATES, 1928 (*Continued*)

(In millions of dollars)

Pure cash items		
Gold shipments	561	169
Changes in ear-marked gold*	68	188
Total	629	357
Grand total, all items †	10,289	10,249

* Gold held in the vaults of a bank, but segregated and marked in such a way as to show it is the property of another bank.

† It is apparent that the credits and debits do not exactly balance, but it should be realized that many of the figures represent estimates only, and necessarily so, since there is no law compelling tourists to inform the government of the sums they spend abroad or compelling American shippers to certify as to the amounts they pay foreigners for ocean freight. It must also be remembered that there is no settlement date upon which a country's account for international payments must be balanced. The balance of payments is a continuing account, and even if our statistics were exact to the last dollar we should not expect an exact balance.

In this table every transaction which creates in the United States a claim against some foreigner is considered a credit. Such items furnish the supply of bills of exchange on foreign countries, or a demand in foreign countries for bills on the United States. Thus the item "Earnings of long-term investments" means that individuals or corporations in the United States acquired credits abroad amounting to $817,000,000 because of interest and dividends from their holdings of foreign securities. Similarly every transaction which gives rise to a claim of some foreigner against an American is placed in the debit column of our account. Thus shipping services performed by foreigners cost $227,000,000, and the cashing of the drafts to which the payment of this amount gave rise decreased the foreign bank balances of Americans by this amount. Such items furnish the demand for foreign bills of exchange, or the supply of bills on the United States available to foreigners.

Limits to the flow of gold. As a first approach to our next problem let us consider two countries, A and B, dealing exclusively with each other, and let us suppose that A in a given year imports $100,000,000 in gold because of an extraordinarily large favorable balance of payments. The gold comes into the hands of bankers in the first place and either is gradually put into circulation or serves as a reserve for other forms of currency. Sooner or later its influence is likely to be felt in

credit expansion, increased business activity, and a rise in the general price level. Similarly the withdrawal of $100,000,000 from circulation in B may cause, other things remaining the same, a restriction of credit, a slackening in business activity, and a fall in the general level of prices.

While gold was flowing into A the rate of exchange must have been down to the gold importing point. Now however there comes a transformation in the exchange market. The rise in the price level in A will make A a better market for the sale of goods from B but a poorer market from which to buy. Similarly B will become a cheaper market from which to purchase and a poorer market in which to sell. Imports from B into A will therefore tend to increase, while exports from A to B will tend to decline. The increase in imports means an increase in the demand for bills of exchange, while a decrease in the supply is caused by the falling off of exports. The simultaneous increase in demand and decrease in supply will have the effect of raising the rate of exchange to some point above the gold importing point in A, thus putting a stop to the further inflow of gold.

Likewise an outflow of gold from country A to country B would be only temporary, for the export of gold would reduce the price level in the gold exporting country and raise the price level of the gold importing country. In the first country the export of goods and services would be stimulated and the imports checked until the increase in the supply and the decrease in the demand for bills of exchange would cause the rate to fall below the gold exporting point, so stopping the outflow.

We conclude that gold cannot flow in one direction, either into or out of a country, in large amounts for an indefinite period, because the flow of gold provides its own automatic check through its effect on the price level, the exports and imports, and the rate of exchange.[1] Temporary discrepancies in

[1] The automatic operation of this check can of course be nullified if the country receiving the gold "sterilizes" it; *i.e.*, prevents it from entering, directly or

the balance of payments are settled from time to time in gold, but economic forces work in the direction of reducing the flow of gold to the minimum, and, while in a given year a country may have considerable net imports or net exports of gold, consideration of a longer period will show the imports and exports of gold tending approximately to offset each other. The tendency over a period of time is for the debits and credits — the amounts owed by residents of a country to foreigners and the amounts owed by foreigners to them — to equal each other. We pay foreigners for goods and services by the export of our own goods and the rendering of services to foreigners; in the last analysis the actual transfer of money becomes trifling.

Reaction of exchange rates upon imports and exports. Aside from the regulating force of the international flow of gold, the exchange rates themselves react upon imports and exports and tend to bring about an equilibrium. This relation will become clear if, placing ourselves in the position of the merchant whose business brings him into the exchange market as a buyer or seller of bills, we consider the effect of the exchange rates upon his transactions. Let us assume that this typical merchant is exporting to England and that sterling bills are selling at a high premium in New York. With exchange in this position, the merchant may make an extra profit from his sales in England. His normal trading profit is computed in the sale price of his goods and is therefore included in the face of the sterling bills which he draws against the English buyers. In turning his bills into dollars at the prevailing high rate of exchange, a second profit is derived from the fact that each pound sterling is worth more dollars than normally. Now if this merchant is in active competition with other exporters, as is the normal case, the profitable market for American goods in England will result in lower

indirectly, into circulation and consequently from affecting the price level, and if the country losing the gold neutralizes the effect of this on its own price level by a more liberal extension of bank credit.

sale prices to the benefit of English buyers, and this situation will increase exportation, since it is axiomatic that more can be sold at a low than at a high price.

Obviously the position of the American importer of English goods is the reverse of this when sterling exchange sells at a premium in New York. Since payment of goods imported from England is made by the purchase of sterling bills in New York for remittance to the English exporter, the American importer finds his profit cut down by the premium on exchange. The premium becomes an additional cost in the business of importing English goods, and the prices of these goods will rise and their import diminish.

These illustrations should suffice to make it clear that the relation of the rates of exchange to the balance of payments is one both of effect and of cause. The rise and fall of the rates is controlled fundamentally by the balance of payments. But when an excess of imports forces foreign exchange to a premium in any market, the high rate tends to restrict imports and encourage exports and thus works toward a restoration of equilibrium in the foreign trade. Conversely when a large volume of exports makes the rates on foreign bills fall to a discount, imports are stimulated and exports restricted, which again tends to bring the trade into balance. In normal times the rates, confined as they are within the bounds of the gold points, can rise or fall so little that their influence upon the balance of payments may be imperceptible. It must be borne in mind however that this narrow range exists only because gold is permitted to flow in sufficient volume to offset the balance of payments; when gold is prevented from supplying this corrective, there is no definite limit to the fluctuations of the rates, and a premium or discount in the rate of exchange may therefore have greater influence on international transactions.

Mercantilist fallacies. The early Mercantilists believed that, if the merchandise exports of a country exceeded in value its merchandise imports, precious metals would flow in to settle the balance. Since to them the precious metals were

especially desirable forms of wealth, this situation came to be called a "favorable balance of trade." Similarly when the value of merchandise imports exceeded the value of the exports, the condition was known as an "unfavorable balance of trade," since it was expected that precious metals would have to be exported. Even today popular belief clings firmly to these notions, and it seems as logical to many of our contemporaries as it did to the Mercantilists to stimulate exports and repress imports in order that prosperity may be brought to a country through import of the precious metals.

There are three fallacies involved in this concept of a favorable balance of trade. The first lies in conceiving the exports and imports of merchandise to be the only items in the balance sheet; the second involves the concept of gold as the most desirable of all imports; and the third is the supposition that a steady inflow of gold can be induced by artificial means.

The first of these fallacies has now been disposed of in connection with our development of the nature of the international balance of payments. A few words on the second fallacy will be in order here.

Is gold a peculiarly desirable import? Many persons would immediately answer this question in the affirmative. They experience a feeling of satisfaction when the statistics show that gold is coming in, they "view with alarm" an outflow of gold, and they would favor governmental action to encourage the import of gold.

In the first place, it should be evident that gold obtained through international trade adds nothing to the total wealth of the nation, since it is obtained only by the sacrifice of other goods presumably of equal value. If a country lets these other things go in order to get gold, it presumably gains in utility through the import of gold. But it just as truly gains when it exports gold in order to obtain other things which it wants more.

As regards the use of gold as money, the important fact is that, within limits, the absolute quantity of money is pretty

much a matter of indifference. Through the adjustment of prices, a given volume of exchanges may be consummated with either a small or a large quantity of money; what occurs is that the price level is low in the one case, high in the other case. It must be recognized however that there are times when the movement of gold is a matter of more than usual concern to the treasury and bank officials of a nation. A country which is with difficulty maintaining the gold standard quite naturally looks upon the export of gold in any considerable amounts as a matter of the utmost consequence and as justification, so it thinks, for any restrictive measures which may so operate on the international balance of payments as to prevent it. Likewise a country which is contemplating a return to the gold standard in one form or another finds the acquisition of gold reserves essential and will undoubtedly consider gold so valuable an import as to warrant substantial sacrifices in goods and services.

In general however there is nothing, except for the needs of a monetary system at critical periods, to set gold apart as a peculiarly desirable import and nothing to be gained by forcing its import in excess of what would enter under the normal operation of international trade. Under conditions such as existed prior to the World War the flow of gold is an automatic thing, controlled by economic forces which result in a state of equilibrium which is for each nation the most favorable situation possible in view of all its other circumstances.

Ignorance of this principle lay at the foundation of the third of the aforementioned Mercantilist fallacies. This particular fallacy has been to a considerable extent exposed in our development of the automatic operation of the natural forces which normally control the international flow of gold. The futility of attempts to create a favorable balance of payments by means of tariffs and other artificial restrictions of imports will be shown in the chapter after the next.

EXERCISES

1. Mr. Jones, an exporter of grain from the United States to France, sells a shipment worth $10,000 to M. Guerne, agreeing to receive payment by drawing on the buyer. Draw the bill of exchange, and trace the steps by which the transaction is handled.

2. Suppose that a New York City bank will pay $4.80 per pound sterling for a demand draft drawn by an American exporter on an English importer.

(*a*) What price would the bank pay for a time draft payable 90 days after sight, if the rate of discount were 4 per cent? 2 per cent?

(*b*) How does a decline in the rate of discount affect the "spread" between the prices of these two types of bills of exchange?

3. Suppose that a New York City bank asks $3,600 for a bankers' sight draft in the amount of 10,000 rupees drawn on Calcutta. How much would the buyer have to pay for a 10,000 rupee draft payable 90 days from date? Assume that the rate of interest is 4 per cent, and the mail time from New York to Calcutta is 30 days.

4. Construct an international balance sheet from the following items, and determine the balance of trade and the balance of payments.

Exports of merchandise	$2,400,000,000
Shipping services rendered by foreigners	80,000,000
New American investments abroad	100,000,000
Interest received by American investors	65,000,000
Imports of merchandise	2,100,000,000
American tourist expenditures abroad	320,000,000
New foreign investments here	600,000,000
War debt receipts	5,000,000
Immigrant remittances	35,000,000
Foreign tourist expenditures here	80,000,000
Interest paid to foreign investors	75,000,000
Gold	.....?

5. Explain how each of the following tends to affect the price of foreign exchange in this country:

(*a*) Decreased export of American cotton

(*b*) Development of an American merchant marine

(*c*) Stoppage of war debt payments to this country

(*d*) Decline of immigrant remittances

(*e*) Increased imports of Canadian lumber

(*f*) Stoppage of interest payments due American investors

(*g*) Redemption of American securities held abroad

(*h*) Loss of confidence in the stability of a foreign currency, causing foreigners to increase their investments here

(*i*) Increased sales of American automobiles abroad

(*j*) Decline of American tourist expenditures abroad.

CHAPTER XXXIII

Contemporary Problems of Foreign Exchange

Pre-war equilibrium. In the chapter immediately preceding, we have followed the workings of foreign exchange as it prevailed in the period before the World War, when the important nations were on a gold standard. It was a period not only of stable exchange relations but of substantial equilibrium as well. Changes of course were taking place, and there were periods of excessive business activity and periods of recession, but there was a mutual adjustment of fundamental factors in national and international life which we do not have today. Price levels were in equilibrium. International trade followed lines determined by a fairly well-settled international division of labor and was not so constricted by universal high tariffs as today nor so distorted by frequent changes in tariff. International investment proceeded along conservative lines under the leadership of the experienced investment houses of London.

This international equilibrium was brought about largely by the normal operations of the gold standard, which provided, as we have seen, automatic checks to factors which might make for disequilibrium. Abnormal factors operating through the price level or the balance of payments could be taken care of by exchange operations without causing disaster either to the exchanges or to the internal economy of a country. The operation of these automatic controls tended to keep the rates of exchange normally within the gold points, but it should be noted also that central banks often intervened in the market to bring the rate of exchange down from the gold exporting point when it was believed that this rate was the result of temporary dislocations in demand and supply. The standard method was for the central bank to raise its rate of redis-

count. A rise in the rate of rediscount at the Bank of England, for example, would make it more profitable to leave funds in London than to transfer them to New York, or it might even induce funds to come there, thus altering the demand and supply of bills of exchange on that centre in such a way as to check the outflow of gold or to reduce its volume.

Post-war exchange conditions. It is unnecessary to emphasize the internal and external dislocations brought about by the World War, but it is important — even though we can do no more than mention a few of the major ones — to show why these dislocations and others, the product of the post-war period, have given the world its heritage of disordered exchanges. In the first place, it is clear that the practical problem of financing war needs made both the abandonment of the gold standard and some degree of inflation almost imperative.

Inflation came about as a result of unbalanced budgets and was resorted to in degrees which varied with the financial strength of the government concerned and its ability to raise revenue either by taxation or by bond issues. It meant of course the printing of inconvertible money and the expansion of deposit currency through government borrowing from central banks for the purpose of meeting current governmental expenditures. The United States did not find it necessary to balance its budget by the emission of paper money, but it did not escape inflation, which was occasioned by the importation of millions of dollars in gold and by increase in the issue of federal reserve notes (a comparatively new form of money at this time) and in the volume of bank credit made possible by this influx of gold.

Two facts of special significance emerge from this monetary disorder. The first is that exchange rates became erratic and subject to violent fluctuations, leading to a vast increase in the risk entailed in foreign transactions and to an unaccustomed volume of speculative operations in the exchanges. The second is that inflation proceeded to different degrees in the inflating

countries, so that the natural relation of price levels to each other was distorted.

Preoccupation with war and the production of war materials led both to the growth of huge war industries in the European countries and also to the decay of other industries, notably some which had been most successful in the export trade. Lack of shipping space, as well as lack of goods for export, transferred much of the export trade of European countries to the United States. The creation of new national boundaries by the peace settlement separated regions which had been economically unified, and this dislocation was intensified by an increase of tariff barriers which a new and fiery nationalism demanded. International trade was thrust from its accustomed channels and impeded in ways which brought disastrous results on the balance of payments of various countries.

Finally the heritage of war debts between the Allied and Associated Powers and the enormous reparations payments demanded by the Allies added another strain to international finance, a strain intensified and indeed made intolerable by the unwillingness of many countries to receive payments in goods and services.

Irredeemable money and the par of exchange. Before discussing the problems of the exchanges in the post-war period it is essential to understand the relation existing between the money of a country on an inconvertible paper standard and that of a country on a gold standard, and also the relation between the respective units of two countries both of which are on an inconvertible paper standard. When the British government in 1919 "unpegged" sterling exchange in New York — that is, when it ceased to give it an artificial price by a continuing offer to buy sterling at a fixed rate — it was said that sterling was free to find its "natural level." Today it is said that Great Britain cannot return to the gold standard until the pound has found its natural level in terms of the dollar. What is a "natural" rate and why is another rate an "unnatural" one?

The purchasing power par of exchange. The answer to this question is found in what is called the purchasing power par of exchange, which may be defined as a rate of exchange between two currencies which expresses an equivalence in the purchasing powers of the respective units in the home countries. Thus if $6.00 were to be assumed to be the purchasing power par of exchange between the dollar and the pound when both the United States and Great Britain were on a depreciated paper standard, it would mean that the pound had six times as great a command over commodities in general in Great Britain as one dollar had in the United States.

The purchasing power par of exchange is of course not a fixed par but fluctuates with changes in the purchasing powers of the respective units, as measured by changes in the index numbers of the price levels. If the price level in the United States were to rise while the British price level remained unchanged, the purchasing power par would move upward from our assumed figure of $6.00, while if the British price level were to rise and the American price level to remain stationary, the purchasing power par would decline from $6.00.

Perhaps the most important characteristic of this par of exchange is that it is neutral in its influence on the currents of commerce. It is a rate which would neither stimulate nor check imports or exports but would permit them to pursue their normal course. It would neither discourage nor encourage capital imports or exports; it would neither bring about an influx of tourists nor would it serve as a bar to their coming.

Purchasing power par and market rate of exchange. The purchasing power par of exchange is always a useful concept in any period of economic equilibrium and when we are considering two countries whose trade with each other is sufficiently large to influence the general level of prices in each country. Market rates of exchange will actually tend to revolve around this neutral or equilibrium par of exchange. A rate higher than our assumed par of $6.00 would mean that

sterling exchange was at a premium in New York. Imports into the United States would consequently be checked and exports would be encouraged. An increase in the supply of sterling exchange and a decrease in the demand in New York would cause the rate of exchange to fall towards the purchasing power par of exchange. Similarly a discount on sterling in New York would affect the currents of commerce in such a way as to bring the rate of exchange up towards the purchasing power par.

In a period of rapid economic change however the adjustment of prices to disturbing factors is very slow, both nationally and internationally, and in such a time the market rates of exchange may bear little direct relationship to the purchasing power par. Let us suppose for example that the United States should inflate its currency. The immediate effect would be a rise in the rate of foreign exchange; it would take more American dollars to buy a pound sterling. Eventually there would also result a corresponding rise in the prices of American goods. If this effect were to come instantaneously, the monetary inflation would have no influence on export trade. A pound sterling would cost more in American dollars, but American goods would be worth more in dollars, so that in terms of domestic goods foreign exchange would be no more costly than before, and there would be no tendency to check imports. From the foreigner's side, American dollars could be obtained more cheaply in terms of, let us say, British pounds, but American goods would be correspondingly higher priced in dollars, so that there would be no added inducement to buy American goods; American exports would therefore not be encouraged.

Actually it takes considerable time for the general price level to adjust to monetary inflation. In the meantime therefore the high exchange rate, caused by monetary inflation, exerts its normal influence to check imports and stimulate exports. The effect of this on the rate of exchange, other things being equal, would be to check the speed with which

it was rising. None the less, so long as inflation proceeded, we would expect the rate of exchange to be continuously at a premium over the purchasing power par of exchange. When inflation ceased and the domestic price level had adjusted itself to the new quantity of currency in circulation, we should expect the rate of exchange again to be determined around the purchasing power par.

The purchasing power par of exchange is today a matter of practical importance. To countries contemplating a return to the gold standard it is essential to discover if possible this neutral rate of exchange; to officials charged with the duty of eliminating excessive fluctuations in the market rates of exchange it is desirable to know what figure may be taken as representing a natural rate of exchange between two countries; and in periods of rapid change it is useful to know the direction in which this par of exchange is moving, even though the causal relationship between the purchasing power par and the exchange rate may be obscured by the operation of other factors.

Unhappily the practical determination of the purchasing power par of exchange is beset with apparently insuperable difficulties; at least thus far no method of computing it has been discovered which is accepted as theoretically sound. The experimental method is the only one available, and it is the one which must be used by officials in charge of exchange stabilization funds. On the basis of indexes of price levels and of all other pertinent economic data, some rate must be chosen as a tentative par of exchange. The officials must then by appropriate operations in the exchange market strive to keep the market rates of exchange from diverging too widely from this assumed par of exchange. If it is found that as a result of these operations the market rates are continuously discouraging exports, it is clear that they are below the purchasing power of exchange and the officials must revise their tentative par of exchange. By a series of shifts of this character they may hit upon the purchasing power par, but even so it should

be noted that their difficulties are not at an end, for the par is or may be a moving par.

Stabilization of exchange rates. In the immediately preceding chapter we saw that a premium or discount in the rate of exchange exercises a corrective influence on the rate by affecting the prices of exports and imports and the other items which appear in the international balance sheet. Nevertheless unstable exchange rates are a usual characteristic of currencies which are divorced from gold. One would expect this to be found in times of economic disequilibrium, when the internal price level is fluctuating, but it is found also in periods of relative economic stability. Not only may temporary or seasonal influences play upon the rate of exchange and force it to an extreme discount or premium — extreme at least in comparison with the limited range of fluctuations of a gold standard currency — but the corrective action of the premium or discount is rather slow and halting in operation.

Unstable rates of exchange are a source of much disturbance in world finance, both by increasing the risk of international financial relations and thereby tending to diminish their volume, and also by inducing speculative dealings in exchange which may themselves cause grave disturbance. Many countries on a depreciated paper standard have therefore felt that they could not afford to permit their rates of exchange to be subject to such wide fluctuations and have accordingly set up some form of conscious exchange control in order to confine the fluctuations within narrower limits.

There are many forms of exchange control to be found in the world today; we shall consider the commonest and simplest type only. This involves active dealings in foreign exchange, constantly or intermittently, by national treasuries, central banks, or special boards constituted for the purpose of eliminating undue fluctuations in rates caused by temporary variations in the demand and supply of bills of exchange or by excessive speculation. Intervention for this purpose is best illustrated by the Exchange Equalisation Account of Great

Britain. This was established in 1932, with resources of £175,000,000, later increased to £575,000,000, and with power to deal in foreign bills of exchange, to buy and sell gold, and to hold balances in foreign banks. If the London rate in New York is declining below the point thought desirable, the officials can sell dollar exchange which they have acquired. A rise of the sterling rate in New York above the desired point can be counteracted by the purchase of dollars in London. By such operations the fund can throw its influence on the demand or the supply side of the market as it desires and can usually exercise a compelling influence on the movement of the rates.

Attention should perhaps be called to two difficulties in the operation of a stabilization fund. The first concerns the choice of the rate of exchange for purposes of stabilization; the second relates to the ability of exchange officials to distinguish between those movements of rates which are caused by temporary variations in the supply and demand of bills of exchange and those which reflect a change in fundamental conditions underlying the balance of payments. Errors of judgment in these two directions may serve to augment rather than to smooth out difficulties. Nevertheless the belief that it is wiser to run the risk of occasional mistakes in judgment than to leave the exchanges entirely uncurbed appears to have been confirmed by the success of stabilization funds in securing relative stability of rates.

Dollar exchange. In concluding this chapter on recent developments in the foreign exchanges, a word must be said about dollar exchange. The changes in the American monetary and banking system have been described in previous chapters in sufficient detail to give the reader some idea of the impact of these factors on foreign exchange. The extent of the changes however and the degree to which control over the exchanges has been lodged with the executive branch of the national government are perhaps not fully appreciated.

It will be recalled that after a period of depreciating dollar

rates in foreign markets, the United States, in January, 1934, set up a system whereby gold might be obtained for export, with a devalued gold dollar defined as containing $15\frac{5}{21}$ grains of gold nine-tenths fine. This represents a decrease in weight of about 41 per cent from the old gold dollar.

A more important test of the gold standard than the weight of gold ascribed to an imaginary coin is the presence or lack of a free gold market, and a free gold market the United States does not now have. Her paper money is not redeemable in gold, either coin or bullion, and present Treasury regulations do not permit the export of gold by private individuals or institutions, though gold may still be imported for private account. Gold does not therefore move in or out automatically, as it does in a country on a gold standard.

A further development has been the establishment in the Treasury of a stabilization fund under the direction of the Secretary of the Treasury. This fund was established by the gold reserve act of 1934 and was provided with $2,000,000,000 of the $2,811,375,757 profit secured from the revaluation of gold. In the original act the life of the stabilization fund was limited to a maximum of three years, but by successive amendments this has been extended. In its objectives and operations it is quite similar to the British fund. The Secretary of the Treasury is empowered to deal in gold and foreign exchange and to conduct other operations which will stabilize the value of the dollar exchange in foreign currencies and maintain the market for government securities. It is obvious that the operations of this fund are to be directed against undesired fluctuations in the rates of countries on an inconvertible paper basis. This might conceivably result in a clash between the operations of national funds. The manipulations of the British to depreciate the sterling rates because of the encouragement to their export trade might be offset by the operations of the United States Secretary of the Treasury if he felt that the dollar rate was too high in terms of sterling and other currencies. The result might be a stalemate or a verdict awarded

to the largest fund — neither of which would tend to improve international amity. At the moment this danger from competing stabilization funds is somewhat lessened by an agreement which was reached in the fall of 1936 by the United States, Great Britain, and France and later subscribed to by the Netherlands, Switzerland, and Belgium. This agreement provided for coöperation with respect to some phases of stabilization operations. Its weakness lies in the fact that any one of the signers can withdraw from it at very short notice.

American exchange is definitely a controlled exchange. The authority of the Secretary of the Treasury to monopolize dealings in gold, the power of the President to alter the weight of the dollar, and the operations of the stabilization fund all point to this, and there is the further power vested in the President to bring all foreign exchange transactions and all international transfers of credit under governmental regulation and control in the event of any national emergency as declared by him. It is true today that the daily routine of foreign transfers is not greatly affected by executive action, but none the less the authority exists to put into effect a much more rigorous control.

Conclusions. Study of recent developments in foreign exchange seems to show that the present trend is in the direction of a greater degree of control over the exchanges rather than a return to the free exchanges of former times. Many students of monetary theory believe this is a desirable trend, and certainly those who advocate a managed currency divorced from all connection with gold are forced to a belief in a controlled foreign exchange market. The weight of the evidence would seem however to indicate that the virtues of an international gold standard outweigh its defects. Although the improbability of securing in the immediate future a general return of nations to the gold standard may be granted, it would appear that international coöperation to select appropriate gold parities and to bring about a more equitable distribution of the gold reserves of the world offers one of the most profitable

approaches to a return of world prosperity. That stable exchanges and the removal of exchange restrictions would lead to an increase in world trade and in world financial dealing seems inevitable. It might indeed be true that certain of the less drastic methods of control could be retained with profit, that some sort of an international clearing house could be devised to eliminate unnecessary gold movements, or that international coöperation might on occasion find it wise to revise gold parities if the ones originally chosen were shown to be exerting an unwholesome influence on the balance of payments. But the strangulation of world trade by exchange methods, the futile competition in exchange depreciation in order to force exports, the singularly vicious attempt to balance the exports and imports between every pair of countries, and other such practices cannot continue in a world that demands the advantages of international division of labor.

EXERCISES

1. Suppose that as a result of a reduction in weight of the American monetary unit its value in terms of foreign currencies declines by 40 per cent. What is the effect of the depreciation of the dollar on each of the following transactions?
   (*a*) An export of American automobiles valued at $100,000
   (*b*) The import of French wine valued at 100,000 francs
   (*c*) Repayment by a foreign corporation of $10,000 due to an American bondholder
   (*d*) The expenses of American tourists abroad
   (*e*) War debt payments to the United States.
2. (*a*) How would the rate of exchange with France be affected by transfers of investment funds by Frenchmen to this country?
   (*b*) How would this development affect the export-import trade with France?
   (*c*) How might the American stabilization fund be employed to offset the transfers of investment funds to this country?
   (*d*) How might the stabilization fund be used to offset withdrawals of investment funds from this country?

CHAPTER XXXIV

Industrial Nationalism and the Protective Tariff

International division of labor. That division of labor between individuals promotes individual well-being in the economic sense has been fully demonstrated, by experience as well as logic. The advantages to be gained by division of labor between different regions or nations are of the same order and equally demonstrable. This subject has already been developed in connection with our study of interregional trade,[1] where it was made clear that any nation or region obtains for its people the highest possible standard of living (other things being equal) when it devotes its industrial efforts to those lines of production in which it has a comparative advantage and depends for other products upon the industrial efforts of other parts of the world.

National self-sufficiency. In spite of these clear truths, the peoples of the world have never accepted fully as a matter of practical policy the corollary of freedom of international trade. Considerations of various sorts have led to governmental restrictions upon the trade between nations. Sanitary and moral considerations demand restrictions upon the import of certain products, such as narcotic drugs. Tariff duties are laid upon imports and exports in order to bring in public revenue. National preparedness for war is sometimes believed to require the fostering of certain industries within the nation, even though this may be accomplished only by placing obstacles upon the import of foreign-made products. In recent times certain nations have believed that the protection or control of their monetary systems required them to place restrictions upon imports. But most potent of all the influences leading to restriction of the freedom of international trade

[1] See Chapter XXXI.

has been a popular belief, widespread and persistent, that by checking imports, a nation may enhance its economic well-being in terms of greater wealth and a higher standard of living.

In consequence of all these considerations — and especially the last — there has developed an ideal of national economic organization which is the antithesis of international division of labor and which may be called "national self-sufficiency" or "industrial nationalism." This ideal requires that a nation be, so far as possible, free from reliance for important products upon the industry of other peoples. For its accomplishment there is required obviously severe restriction of imports from abroad; indeed a tendency to regard all imports as a menace to national prosperity is evident. While the theory of nationalism does not necessarily imply a forcing of exports, the fact is that those who hold to this ideal generally regard exports with a favor no less than their disapproval of imports. Imports are bad; exports are good; the former impoverish, the latter enrich the nation.

This ideal of industrial nationalism was a cardinal principle in the philosophy of the Mercantilists, and its accomplishment was consciously sought in the practical policies of the leading nations of Europe in the seventeenth and eighteenth centuries. Its theoretical foundation was demolished by Adam Smith. It later acquired a new lease on life through the teachings of the German economists of the so-called "historical school," predominantly Friedrich List, and Henry C. Cary in the United States. It is an article of faith of the protectionist movement which has flourished in many nations for many generations. At the moment, the idea of industrial nationalism has gained a hold upon popular thinking stronger perhaps than at any time since the days of the Mercantilists. It is closely interwoven with commonly held doctrines of the causes and cure of depression and is apt to be an essential part of current conceptions of "the planned economy." It must finally not be forgotten that among the most persistent apostles of na-

tional self-sufficiency are many whose own self-interest would be promoted either by the restriction of imports generally or — what is more to the point — of the particular good which they produce in competition with foreign makers.

Effects of the protective tariff. For the accomplishment of the ends of national self-sufficiency a variety of means is at hand. Embargoes, limitation of permitted imports by national quotas, restrictions upon foreign exchange, depreciation of the national monetary unit — all these are devices which have been employed, never perhaps so widely or vigorously as at the present time. But the chief instrument for the restriction of imports is the protective tariff, and the whole subject of industrial nationalism may be elucidated in large measure by analysis of the problem of protectionism.

For our excursion into this problem the principles of international trade and the principles of the shifting and incidence of import duties, as developed in earlier chapters,[1] furnish the starting point. We know that, in the absence of restrictions upon its foreign trade, the people of each nation tend to buy goods in the cheapest market and sell in the dearest market and that this leads them to devote their productive resources and energies to those lines in which their nation has comparative advantage, as evidenced by the ability to produce at comparatively low cost. At the same time they rely upon other nations for those goods in whose production they are at comparative disadvantage. Thus develop international trade and international division of labor.

International trade may be checked by means of duties on imports.[2] As we have seen, a duty high enough to be protective may diminish or entirely stop the import of any chosen commodity. Carried to its extreme limit, the protective tariff might be used to prevent virtually all imports. Whether applied moderately or to the limit, the protective tariff has the

[1] Chapters XXXI and XXV.

[2] Since they seldom enter into the controversy — never in the United States — no consideration need be given in this discussion to duties on exports.

effect of reducing imports and building up corresponding industries at home. It thus restricts international trade, promotes "diversity of industry" in the home country, and diminishes reliance upon the goods of other nations. National self-sufficiency gains at the expense of international division of labor.

As to the truth of the foregoing statements no question can be entertained by anyone who is familiar with the elementary principles of economics. The controversy over protection starts from this point. Granting the respective effects of freedom of trade and protection as we have stated them, the "protectionist" asserts that it is to the economic advantage of a nation thus to increase diversity of industry and national self-sufficiency; the "free-trader" [1] says that this means an economic loss and that the nation gains most when it enjoys all the advantages of international division of labor. Thus the issue is sharply drawn.

Protection and national self-sufficiency: The case for protection. The advantages of trade, as already developed, are so evident and so important that we are justified in placing the burden of proof upon the protectionist, who claims that there is an economic gain from putting artificial restrictions upon a nation's foreign trade. Let us proceed then to examine the principal arguments advanced in support of the policy of protection.

a. The favorable balance of trade argument. Some of the most popular arguments for protection are such naïve applications of old Mercantilist fallacies that they may be dismissed with nothing more than brief reference to the established principles of trade. Such is the argument that a nation gains in its foreign trade only from its exports, that imports are a national loss, and that the tariff is therefore beneficial

[1] The reader should perhaps be reminded that "free trade," as the term is used, does not mean no tariff; it means no protection. The free-trader has no quarrel with the revenue tariff; he is opposed only to protection. There is no country in the world entirely without a tariff and no appreciable sentiment in favor of the abolition of all tariffs.

through its effect in checking imports. We have already learned that, so far from exports being superior, it is imports which are the real goal of international trade; we have learned also that a country's balance of payments must in the long run be in substantial equilibrium and that therefore anything which reduces imports must eventually also reduce exports. It is neither advantageous nor possible by means of a tariff to make imports decline relatively to exports.

b. The gold argument. Similarly naïve is the argument that a favorable balance of trade, created by the tariff, causes gold to flow into the country and is therefore advantageous. This is simply a corollary of the preceding argument and rests upon other Mercantilist fallacies with which the reader is acquainted. We know that it is not the balance of trade (*i.e.*, merchandise exports and imports), but the balance of payments (including all debits and credits) which determines the inflow or outflow of gold; and we are also aware that it is neither possible nor advantageous for a country to pile up gold as the result of artificial restrictions upon its foreign trade.

But there are those who are not content with this summary disposal of cherished beliefs. Is it not true, they say, that the rule for amassing wealth is to sell much and buy little? Is it not thus that the individual business man accumulates money from the profits of his business? And must not the same rule hold for national gain from trade? At the risk of adding proof of the obvious, we may give some further attention to these Mercantilist notions.

The wealth of the people of the United States in 1925 was estimated at about 355 billion dollars. At that time the nation's total stock of gold coin and bullion was about 4½ billions, or about 1¼ per cent of the national wealth. In 1880 the national wealth was, according to estimates of the United States Census, about 44 billions. The stock of gold coin and bullion was then 352 millions. In this period of 45 years therefore the national wealth was increased by 311 billions, while slightly over 4 billions were added to the stock of gold. It is certain

enough that this great gain in national wealth was not due to accumulation of gold from foreign trade. The growth of a people's wealth is not principally in the form of gold; it consists in the increased value of its lands and houses, its factories and railroads and ships, its homes, and countless other forms of wealth vastly more important than its stock of gold. And the gain from its foreign trade is not in the form of gold but in the useful commodities imported.

But what of the analogy of the business man? The answer is that neither does he accumulate wealth in the form of money or by selling more than he buys. If at the end of a profitable year in which a man has made gains of $50,000 we should ask him to show us his profits in concrete form, he would not point to an increase of his money on hand or his bank deposits. The probability is that these items are no larger than at the beginning of the year. His profits were not made by selling more than he bought. As he received money from the sale of goods, he paid it out — partly for current consumption, partly for additions to his factory and machinery, for lands, for improvements to his home, for a new automobile, for government bonds and corporation securities. Thus he will show concrete evidence of his year's accumulation. So it is with the nation. A nation, like an individual, grows wealthy not by selling more than it buys but by producing more than it consumes.

c. The added industry argument. It is argued that the new industries which admittedly may result from protection are an addition to the industrial resources of the nation and that therefore the tariff has increased the national wealth. Let the reader recall that the tariff causes the establishment of a new industry by making it too costly to buy a foreign product previously imported. He will further recall that this diminution of our imports must be balanced eventually by a corresponding decline of our exports. The balance of payments must maintain its equilibrium; if we will not buy of foreigners, they cannot buy of us. The result is that, while a new industry is arising, some old industry finds its foreign

market declining and is forced to reduce its operations accordingly. The new industry is thus not an addition to the nation's industry but merely a diversion of capital and labor from one field to another.

Whether capital and labor already employed are shifted or new capital and labor are directed into the protected field, there is clearly no gain in total industry. Moreover the diversion of industry is from the more profitable into the less profitable fields, and so results in a net economic loss.

d. The wages argument. This argument appears in two forms. (1) It is asserted that the protective tariff, by increasing the total amount of industry in the country, adds to the demand for labor and so tends to increase wages. Protection is therefore of advantage, not merely to the entrepreneur class, but to all laborers, and so to at least the great majority of the people. It is evident that this argument rests for its foundation on the added industry argument and falls with it. The mere diversion of industry, particularly when it is from more advantageous to less advantageous lines, can at best do nothing to increase the demand for labor or to raise wages.

(2) The wages argument is not always put so crudely as this. By American protectionists it is more often argued that, in the absence of the tariff to keep out cheap foreign goods, American manufacturers could not pay their present high scales of wages but would be compelled to lower wages or perhaps to go out of business entirely. Thus the tariff is in the interest of the workingman. This argument overlooks the fact that many American industries seem able to get on well enough without protection. These are the industries naturally fitted to prosper in the American environment, the industries in which America has a comparative advantage. Without protection, they are nevertheless able to pay the prevailing high level of American wages; they pay as high wages as the protected industries. If any industry is really unable to pay the prevailing rate of American wages it is because that industry is not adapted to our conditions. Its presence is due

to the artificial stimulus of the tariff, and it exists at the expense of some more favorable industry which could pay high wages without tariff assistance. Removal of the tariff might cause the artificially favored industry to disappear, but its loss would be offset by the gain in the naturally favored industry, and there would be no decline in wages.

Wages and labor cost. There is much loose talk about wages in this connection. Protectionists argue as though, simply because wages are higher here than elsewhere, it must cost more to produce anything in America. They express alarm at the prospect of competition with the products of "foreign pauper labor." But the fact is that in numerous fields of agriculture and manufacture goods are produced more cheaply in America than anywhere else in the world, in spite of high wages. In these fields employers are able to pay high wages for high-grade labor, and that without making the cost of production high.

In international competition, as in industrial competition everywhere, the controlling factor is not high or low wages, but high or low *labor cost of production.* Wages, as such, tell us nothing. It is only when wages are divided by quantity produced, to give the labor cost per unit of product, that we have the key to the problem of relative advantage in international competition. Thus it is that high relative wages when accompanied by still higher relative output give low labor cost. This is the condition in the great body of industries which are naturally adapted to American conditions. Here high wages are paid without the aid of protection. In general these are industries which either are favored by peculiar advantages of soil or climate or natural resources, such as cotton growing, wheat growing, copper mining, etc., or have the advantage of skillful use of labor-saving machinery, such as the manufacture of automobiles, sewing machines, farm machinery, etc. Industries not so favored by nature or which must lean heavily upon hand labor do not flourish here. To them the high level of American wages is a handicap, which can be

overcome only by a protective tariff which removes the competition of the cheaper foreign product. Protection is required therefore, not to maintain the high level of American wages, but to make possible the payment of American wages in industries not suited to American conditions.

e. Effect of withdrawing protection. The wages argument is without doubt urged in good faith by many American manufacturers who honestly believe that protection is necessary, not merely to their own profits, but also to the well-being of American labor. Being engaged in lines of industry which possibly could not exist without tariff protection, they believe that without protection they could not continue giving employment to their labor forces at the present scale of wages. They visualize the distress which would follow a material cut in wages or a shut-down of industrial plants. They fail to see that discontinuance of the artificially fostered industry would eventually open up other fields in which these laborers would find employment at American wages.

It is not to be denied that the shifting would impose inconvenience and loss upon the laborers and still more upon the capitalists. Industry that has become adapted to the artificial condition of protection cannot adjust itself to freedom of trade instantly or without some loss to all concerned. This is no argument for retaining permanently the artificial condition. But it does dictate caution and consideration in making the change. Protection has long been an established American institution; entrepreneurs have directed capital and labor into certain fields of industry relying upon the government's policy of protection. They and their employees have the right to ask that a change in policy be made gradually and with every possible care to minimize the unavoidable losses of adjustment.

The fact is that American industry would not be wrecked by a careful, gradual removal of protection. Some few industries would doubtless succumb. In most cases certain grades or lines would be discontinued while others advanced, the

sort of adjustment which normally goes on as a matter of course in every progressive industry. For industry as a whole the final effect would without question be beneficial. American tariff history shows protection surprisingly less potent, either for good or for evil, than both its friends and its enemies would have us believe. The course of American industry has not been vitally influenced by its imposition, nor would it now be revolutionized by its withdrawal.

f. Alleged saving in transportation costs. As an argument in favor of protection, attention is sometimes called to the supposed waste involved in sending raw materials to a foreign country only to be returned to the country of origin in the form of manufactured products. An example is American cotton, much of which goes abroad to English cotton mills, and some of which finds its way back to America in fine British cotton fabrics. Here is said to be a wasteful double cost of transportation, which might be avoided by a tariff that would keep out the British product and compel its manufacture in America.

The fallacy here is obvious. There is no loss in the operation in question; otherwise it would not occur. Transportation costs are no worse than other costs. The double transportation of cotton across the Atlantic is undertaken because that is the way to get the product for the American consumer at the least total cost. Forbidding this method by a tariff would save this transportation cost, to be sure; but only at the expense of a still greater cost of manufacture. The *reductio ad absurdum* of this argument appears if we imagine ourselves trying to apply it universally. Practically all domestic industry involves such double costs of transportation. Wheat is raised on Nebraska farms, shipped to Minneapolis, ground into flour, and returned to the Nebraska farms. Iron and copper and coal are brought from Minnesota and West Virginia to New England, and the finished steel and brass goods are shipped back to these states. The sole reason for all this double transportation is to minimize total costs of production.

Its avoidance would be possible only at the cost of abandoning territorial division of labor and putting the human race back in the Middle Ages economically.

g. Meeting the tariffs of other nations. There are those who are ready to admit the advantages of universal freedom of trade between nations but who yet feel that, in a world in which protective tariffs are common, no one nation can afford to permit free entry of foreign goods. Nor can one nation safely remove its tariff barriers unless other nations do the same. This misgiving overlooks the experience of Great Britain, which, during half a century in which she led the world in an amazing industrial development, maintained freedom of trade regardless of the tariffs imposed against her products by other nations. It overlooks also the fundamental principle. It is true that our own industries and our economic welfare are injured by the tariff barriers of other nations, which deprive us of some measure of the benefits of international trade. But the whole conclusion of our analysis has shown that our national welfare is injured by the curtailment of our foreign trade occasioned by our own tariff. Freeing our trade from the shackles both of our own tariff and the tariffs of foreign nations would be doubly advantageous. But if we cannot secure the removal of foreign tariffs, that is no reason why we should forego the gain to be derived from removal of our own tariff.

The trouble with reciprocity agreements, tariff treaties, and the like is that they are usually conceived in the spirit of Mercantilism, which regards exports as a national blessing and imports as a national curse. There follows a sort of inverted horse-trading negotiation, in which each nation seeks to give the other the utmost of its products while taking as little as possible in return. Once the notion is reversed and imports are recognized as the true goal of foreign trade, with exports desirable only as the means whereby imports may be obtained, the whole spirit of mutual tariff concessions is changed. No nation need grudge other nations whatever advantage they

may derive from relaxation of its own tariff. If concessions in foreign tariffs may be obtained in return for reduction of our own tariff, this is all to the good, but it does not follow that our own relinquishment of trade obstructions should be abandoned because other nations do not see fit to act likewise.

h. Mutual dependence of nations. Mercantilist notions sometimes run to the extreme belief that whatever injures the industry and trade of another nation must therefore be beneficial to us. It follows that one of the chief services of our tariff policy should be to injure the economy of foreign nations by excluding their goods from our markets. That can of course be done. We have seen that in that case the principal injury is to ourselves. It now remains to point out that we suffer also from whatever injury our tariff policy may inflict upon other nations. The whole world is bound together by the universal principle of division of labor. Just in proportion as the industries of other nations are efficient and productive will they be able to serve our needs by enabling us to obtain by exchange certain products from them at less cost than if we had to make them ourselves. In general each nation stands to gain economically not from the adversity but from the prosperity of other nations.

Considerations against protection. Having placed the burden of proof on the protectionist and found his arguments not convincing, the free-trader might conceivably rest his case here. There are however some important positive considerations on the free trade side of the argument which may serve to throw further light upon the question.

What is the territorial unit for protection? It is an interesting characteristic of the foregoing arguments for protection that each assumes that the tariff barrier will follow national boundary lines. Yet the arguments themselves are purely economic; they contain nothing to show that protection would not be exactly as beneficial if directed against a competing region within the same nation. We are tempted to ask why the New England textile mills should be denied pro-

tection against the rising industry of the South, and *vice versa*, or why the automobile industry in various states should not have protection against Michigan. Historically many of the first tariffs were imposed by towns or other minute political subdivisions. They proved such a burden on trade and such a drag on economic progress that they were swept away. In like manner it came to be recognized that the tariff systems of some score or more little states in central Europe were proving fatal to the economic development of that region, and the German *Zollverein* — customs union — swept them all away and opened the path to a later political union and to the marvellous industrial development of the German Empire. Originally the American states had each its own tariff system directed against the others. They had to give way in the interest of the economic welfare of the whole region. Few today would seriously propose to retrace any of these steps already taken in the direction of freedom of interregional trade. No argument exists to show that still further progress would not follow the removal of the remaining barriers to trade between nations.

Benefits of foreign competition. It is a fallacy to regard foreign competition as especially heinous. Manufacturers naturally dislike competition. They seek to avoid it whenever they can. But competition is in the interest of the consumers, for whom industry in the last analysis exists. And the competition of foreign producers is just as wholesome as that which is nearer at hand. If the home manufacturer cannot meet the foreign competition, then the home consumers are entitled to the less costly product. If the home manufacturer, spurred on by the foreign competition, does succeed in meeting it and holding his market, then again the consumers are benefited. Moreover the nation gains from the increased efficiency of an industry which might otherwise have lagged behind. There are instances of American industries which, if pushed to do their utmost, could have competed on even terms with the best of the foreign producers but which, shel-

tered by the American tariff wall, have grown lax and fallen behind the industrial procession.

Conclusion. Analysis of the arguments for and against protection thus appears to demonstrate that industrial self-sufficiency secured by means of the protective tariff, so far from enhancing the national welfare, is a cause of economic loss. No gain can be shown to offset the demonstrated advantages of international division of labor. Protection is an economic burden, a cause of loss to the people of the nation which employs it. The economic advantage is all on the side of freedom of international trade.

It is probably not necessary to remind the reader that these conclusions do not imply that any nation — least of all the United States — could instantaneously abolish its entire tariff system without some unhappy consequences. Tariff systems have grown up during a long period, and the industry of each nation has adapted itself to the tariff situation as it has been and is. In America a system of very high tariff duties upon most of the important products of manufacture and upon some raw materials has been in effect for three-quarters of a century. There have in this period been variations in the level of protection, but never anything approaching free trade. American industry has grown up in this environment, to which it has had constantly to adapt itself. To change this condition, suddenly and without notice, by going over to the free entry of all foreign-made goods would throw industry into confusion and do serious injury to the economic welfare of the nation. From this proposition — which must be well-nigh self-evident — the conclusion is, not that free trade is not the ideal or that protection must be retained, but rather that the change from a condition of protectionism to the healthier condition of free trade must be made gradually and with due notice, to the end that the costs of readjustment to industry and the inconvenience to the people be made as small as possible.

Temporary protection. Thus far we have been considering the question of protectionism from the viewpoint of the

relative merits of international division of labor and national self-sufficiency, the protective tariff serving as a permanent device for promoting the latter ideal. In contrast we must now observe the possibility that the tariff may be called upon from time to time to serve certain temporary objectives not involving any permanent artificial interference with the freedom of international trade. The classical example of this application of protection is the use of the protective tariff on behalf of young industries.

Infant industry protection. Use of the protective tariff has been frequently urged and occasionally employed as a means to facilitate and hasten transition from the agricultural to the industrial stage of national economy. A nation on the threshold of this change will be taking the first steps toward establishing numerous manufacturing industries. In these first steps manufacturing will have to contend with the competition of imported goods coming from older nations whose industries are already firmly established. They have the machinery, the technical knowledge, the skilled workers; they are directed by experienced entrepreneurs. None of these essentials has as yet been attained by the young industries of the first country, and the competition of foreign manufacturers is felt as a severe handicap.

Under these circumstances the tariff may be used to give relief from foreign competition until such time as the young industries have become firmly established. This line of argument is known as "the infant industry argument" for protection. When logically used it rests upon the following assumptions: (1) the argument applies only to industries which are suited to the national conditions, (2) the obstacles to the establishment of such industries are temporary only, (3) the tariff protection need be temporary only; after the removal of the tariff the industries will be able to stand alone. When these conditions are fulfilled and when it is further proved that the nation will actually benefit from the establishment of the industries in question more rapidly than they would

come in the natural order of events, the infant industry argument is undoubtedly sound.

Conditions in the United States at the beginning of the nineteenth century were fairly typical of those here assumed, and the first tariff acts were perhaps to some extent justified by this argument. Even more than the tariff, the international conflicts of the period from 1807 to 1815, by interfering with foreign trade, gave protection to the young American industries. But the conditions to which the infant industry argument was applicable had generally passed with the first third of the century. This argument is little heard of today. It must be noted also that it has not often been urged in good faith; having used it to obtain tariff favors, its advocates have later shifted to other arguments to prevent removal of protection. Most protectionists today advocate permanent protection.

Protection against "dumping." From time to time producers of some nation undertake to unload a surplus product upon foreign markets at prices lower than are charged in their own domestic market, sometimes even lower than cost of production. When goods are thus exported at less than cost the process is called *dumping*. Dumping may be resorted to by some particular firm whose position sufficiently approaches monopoly, or by a combination of producers acting in concert for this purpose, for example the closely organized chemical industry of Germany. Often such export trade is encouraged or actually subsidized by the government. For a long time it was the practice of the state-owned railroads of Germany to grant special low freight rates to certain great kartels on goods destined for export. During the past decade dumping has become more and more a governmental policy, especially of certain authoritarian nations, employed in connection with their efforts to control their foreign trade and foreign exchanges and protect their stocks of monetary gold. Similar results are caused when nations, as has been quite common in recent years, deliberately devalue their currencies

in order to take advantage of the temporary stimulation to export trade which tends to follow.

The producers of any nation whose markets are invaded by such dumping of foreign goods are subjected to competition of special severity. Were this a stable and presumably permanent condition, it would be the part of wisdom for the nation concerned to accept the situation, congratulate itself that its people were able to obtain some needed products thus cheaply, and let its own capital and labor transfer to more advantageous lines. But dumping, from its very nature, is fickle and unstable. Such temporary competition may work serious injury to a nation's industries, only to leave its people poorly supplied when the foreign producers return to a more normal price policy. The reader will recognize the similarity between foreign dumping and certain practices of monopolistically inclined enterprises in the domestic market, practices which have earned the title of "unfair competition."

It is only natural that a nation should seek to protect its own producers against dumping. The protective tariff is the obvious means. Wisely employed, the tariff may serve to offset the price cutting of foreign producers, with the result that the foreign goods enter the domestic market at prices not less than their normal costs of production. This policy involves frequent adjusting of tariff rates to meet changing policies of foreign producers, and it requires skillful handling. There is moreover always the danger that selfish interests will use the dumping bogie to secure excessive or permanent protection. Suitably employed however protection against temporary dumping is in the interest of the whole people of the nation concerned. Protective tariffs and other measures against dumping are at present being employed by an increasing number of nations.

Protection on account of military necessity. Protection may be urged from an entirely different standpoint. The tariff is necessary, it is said, in order that the nation may be assured of a supply of certain products vital to its safety in time of war.

If it be proved (1) that a certain industry is thus necessary and (2) that this industry will not exist unless relieved of foreign competition, then it follows unquestionably that protection is advantageous. Whether these two assumptions are actually fulfilled is a matter of fact to be determined in each case; it is doubtful whether, in a country of the size and variety of resources and industries of the United States, there is much room for this line of argument.

Without doubt the extraordinary development in recent years of protectionism among the nations of Europe has been in large measure based upon military considerations. These nations were preparing for war and seeking the utmost of national self-sufficiency in order that in the next war they should not be weakened by enemy blockades as was Germany for example in the World War. Thus we have seen tariff rates of unprecedented height, embargoes, quotas, and the other devices which marked the nationalistic hysteria of the past decade or two. As a result of these policies the outbreak of war in the autumn of 1939 saw the chief belligerents possessed of a far greater degree of national self-sufficiency than they enjoyed in 1914, though doubtless at the cost of some sacrifice of living standards.

It should be noted that the military necessity argument for protection is not based upon the traditional economic grounds. It does not rest theoretically upon the false assumption that the people will be richer on account of the tariff. It can admit that they will be poorer, through having to pay more for some important commodity than they need have. But so are they made poorer by the taxes which defray the cost of building battleships. The justification is the same in either case.

Protection by bounties. Occasionally the object of protection is attained, not by a duty on the foreign product, but by a bounty paid by the government to the domestic producers. The United States for example has on occasions sought to encourage the growing of sugar by means of bounties. As to their economic significance the bounty and the protective

tariff are alike, with one important exception. The cost of protection by tariff is imposed upon the consumers in proportion to their use of the goods in question. The cost of the bounty rests upon the whole group of taxpayers. Since protection can be justified, if at all, only on the ground of public advantage, the apportionment of the cost under the system of bounties is more equitable than that of the protective tariff. Nevertheless bounties are very rarely used, probably because it is here so evident that protection costs something, that a burden is placed on the whole people for the benefit of a favored few.

Ship subsidies. Ship subsidies are bounties paid to ship owners or operators in order to encourage the development of national shipping — the "merchant marine." The economic nature of ship subsidies is precisely the same as of protective tariffs and bounties. The justification of subsidies to shipping must be found, if at all, in the military necessity argument.

Preferential trade and tariff arrangements. At the same time that the recent wave of extreme nationalism has led most nations to curtail severely their trade with other nations, there has developed a contrary tendency to enlarge the area of comparative freedom of trade by means of special tariff agreements. The British Empire furnishes a conspicuous example. In 1932 Great Britain took definite action terminating her traditional policy of free trade, and in the same year a scheme of imperial preference was set up by agreement among Great Britain and the various self-governing dominions. Thus was the British Empire bound together in an area of tariff preference as against the rest of the world. France, without going so far toward a formal organization, has made preferential tariff arrangements with her colonies. Germany has been seeking to build up a preferential tariff group, including nations under her special influence in central Europe as well as certain South American countries. Japan is quite evidently seeking to dominate a great Eastern tariff area including Manchukuo and China.

Reciprocal trade agreements of the United States. A somewhat different line has been followed by the United States. The trade agreement act of 1934 (continued by act of 1937) authorized the President to enter into foreign trade agreements with other nations, in connection with which he was empowered to make reductions up to 50 per cent in existing tariff rates upon certain imports from a given nation in return for corresponding concessions in that nation's tariff in favor of American exports. The Secretary of State, Cordell Hull, has acted with great vigor in negotiating such reciprocal tariff arrangements. By the fall of 1939 twenty-one such agreements had been concluded, the most important, and possibly the most difficult to arrange, being the agreement with Great Britain and a new agreement with Canada, which were signed on November 17, 1938, to go into effect on January 1, 1939.

These agreements are reciprocal, but not preferential like those of most other nations, since the United States still lives up to the liberal "most-favored-nation" principle, in accordance with which any reduction granted to one country is "generalized" to all nations with which the United States has most-favored-nation treaties. This does not mean that there is no advantage in being the party to a trade agreement with this country, since the agreements are based on the "chief-supplier" principle; *i.e.*, the United States makes tariff reductions only on articles coming exclusively or chiefly from the nation which is the other party to the agreement.

It would be too much to suppose that this reciprocal agreement policy is making any real breach in the traditional wall of American protectionism. Indeed care has been taken not to permit competition that would seriously affect any American industry. This is accomplished both by the choice of articles for tariff concessions and by limiting the amount of import; for example, in the agreement with Czechoslovakia her producers were permitted to send shoes at the reduced rate, but only up to 1.25 per cent of the average annual pro-

duction of the United States. Nevertheless in a world that seems rapidly to be throwing away important economic benefits through an exaggerated nationalism, the present attitude of the United States is furnishing a powerful influence on the more liberal side. This comes with special significance from the nation which for a century has been the chief exponent of the policy of high tariff protection.

CHAPTER XXXV

Labor Problems

Problems of labor. In this chapter will be discussed various problems confronting the wage earner in the modern system of industry, selected from a much larger number of labor problems because of their importance to society as a whole as well as to the individuals affected by them. Since these problems are broadly social in cause and effect, they have given rise to attempts to solve or alleviate them by governmental action. These remedial policies will be considered in the following chapter. Because the problems of labor affect large groups of people alike, it is natural that these groups should have organized for their own benefit and protection. The organized labor movement with its programs and tactics and its violent clashes with employers springs from these characteristics of the labor market. This phase will be presented in Chapter XXXVII.

Unemployment: Nature and extent. Accurately defined, unemployment is involuntary idleness on the part of people able to perform useful labor. Although we lack complete information as to the extent of unemployment, it is known that the amount is large even in times of normal business activity and that it rises to astonishing proportions during depression periods. One careful study of the major American industries over a period of twenty-one years has given us the table on page 676. Here unemployment is measured as the percentage of the idle to the total available labor supply. As shown by the caption many occupational groups are not included in the figures.

This table provides a basis for estimating what may be called *normal* or *chronic* unemployment. The average for the entire period in these industries was about 10 per cent of all

the workers. The percentage fell below 6 per cent in only two of the best of these years, and it rose above 16 per cent in the worst. It has been estimated that inclusion of certain more stable occupational groups would reduce these percentages by about 2 points all along the line.[1] Allowing for cyclical fluctuations the volume of unemployment changed very little over this long period. We may say then that on the average some 8 per cent of all the workers are involuntarily idle all the time.

UNEMPLOYMENT IN MANUFACTURING, TRANSPORTATION, BUILDING TRADES, AND MINING, 1900–1920[2]

Year	*Percentage unemployed*	*Year*	*Percentage unemployed*	*Year*	*Percentage unemployed*
1900	10.0	1907	6.9	1914	16.4
1901	7.5	1908	16.4	1915	15.5
1902	6.8	1909	8.9	1916	6.3
1903	7.0	1910	7.2	1917	6.0
1904	10.1	1911	9.4	1918	5.5
1905	6.7	1912	7.0	1919	6.9
1906	5.9	1913	8.2	1920	7.2

Unemployment becomes acute during the lowest points of a business depression, and at such times it is recognized as the major practical problem of the day. Through their effect on the employment market, the forces causing a recession of business activity renew themselves and tend to carry the depression to even lower levels. The initial maladjustment in the business world throws large numbers of men and women out of work. The cessation of their buying power in the market prolongs and increases the maladjustment, causing further recession of business activity and in turn additional unemployment. It is likely too that some of the things done by the government and other social agencies to relieve distress and make work for the idle act as independent causes retarding normal economic recovery or even aggravating the depression.

[1] P. H. Douglas and A. Director, *The Problem of Unemployment*, pp. 32–33.
[2] P. H. Douglas and A. Director, *op. cit.*, 1931, p. 28.

The following table shows the course of unemployment since 1920, both in actual numbers and in percentages of the wage-earning population. These estimates cover all types of employment except agricultural workers. In these statistics persons employed by government on relief projects are regarded as unemployed.

ESTIMATED UNEMPLOYMENT IN THE UNITED STATES [1]

Year	*Average monthly number unemployed*	*Average percentage of all wage earners*
1920	1,400,000	5.1
1921	4,300,000	15.3
1922	3,400,000	12.1
1923	1,500,000	5.2
1924	2,300,000	7.7
1925	1,800,000	5.7
1926	1,700,000	5.2
1927	2,100,000	6.3
1929	1,100,000	2.3
1930	4,100,000	8.4
1931	8,200,000	16.6
1932	12,700,000	25.4
1933	12,900,000	25.4
1934	11,200,000	21.8
1935	10,300,000	19.9
1936	8,600,000	16.4
1937	7,400,000	13.7

It will be seen that even in 1926, which was a fairly good year industrially, upwards of a million and a half workers were unemployed. The disastrous unemployment which accompanies a prolonged depression is reflected in the figures following 1930. These figures, large as they are, do not tell the whole story, for the well-established practice of spreading work by placing the whole labor force on part time has reduced the percentage of complete unemployment while throwing a burden of partial idleness upon the workers who still hold jobs.

Types of unemployment: Basis of classification. In the broadest view of the matter the causes of unemployment,

[1] Figures to 1927 from L. Wolman, *Recent Economic Changes*, Vol. II, 1930, p. 478. Later figures based on studies of the National Industrial Conference Board. The year 1928 is not included by either of these sources.

however they differ in detail, must have one essential characteristic in common. They are all aspects of maladjustment between the demand for labor and the supply of labor in the employment market. Since the supply of labor is not subject to sudden or capricious change, the forces which upset the equilibrium of demand and supply must operate principally on the demand side of the market. There is an unexpected reduction in the need for wage labor or in the profitableness of its employment in some large occupational group, or in several such groups simultaneously, or in all industries together. Discovery of the major causes of these changes enables us to identify three principal types of unemployment, each with its own characteristics of cause and remedy: *technological*, *seasonal*, and *cyclical unemployment*.

Technological unemployment. The invention of new and the bettering of old machine equipment, reorganization of the lay-out of production within an enterprise, improved methods of management, and the like are salient features of modern industry. Any such change is considered an improvement only if it reduces costs of production. The saving of cost usually takes the form of a reduction of wages per unit of output, accomplished either by diminishing the number of workers employed or substituting a cheaper for a higher grade of labor. The immediate effect of technical progress therefore is to reduce the demand for the labor of some occupational group.

It is hardly necessary to present evidence regarding so commonplace a feature of our industrial order. Technological change is incessant in all lines of production. In the vast majority of cases the specific event has a relatively small effect on the labor market; but even minor changes, if they are real improvements, tend to unsettle the equilibrium of demand and supply at some point. There are times, however when technological progress seems suddenly to affect large areas of the industrial system and to exert far-reaching influences upon the wage-earning class as a whole. Workers in a given industry

are not affected by changes in that industry alone, but may find the sale of their own product and hence the demand for their labor decreased by a reduction of costs in an enterprise producing a competing commodity. Thus the influence of technical advance is broadly diffused through the laboring population. For example during the decade 1919–29 physical productivity per worker in manufacturing increased 45 per cent and in agriculture 25 per cent because of improvements in the processes of production. At the end of this period there were 500,000 fewer workers in manufacturing than at the beginning; 200,000 fewer in the mines; while of the 3,800,000 people who left the farms during this decade, 1,500,000 were of working age.[1] No doubt many of these displaced workers suffered unemployment for various periods of time. Eventually most of them found work in new occupations. Though we lack complete information on the subject, it is known that large numbers were absorbed in the expanding clerical trades, in merchandising, and in the services clustering about the use of the automobile — the filling stations, repair shops, roadside businesses, and the like.

We have said that the initial effect of technical progress is to reduce the demand for labor; but it is well to carry our thought a little further. Ultimately improvements in the productive process must lower the price of the good in question; the quantity sold will increase; the industry will expand, thus tending to reabsorb the displaced workers. The amount of expansion and reabsorption of labor resulting from a given price decline will depend on the elasticity of demand for the good in question. If the demand is very elastic it is possible that even more labor will ultimately be employed in the same industry after the change than before, although the increased demand may be for a different kind or grade of labor. If elasticity is comparatively slight there will probably be a permanent decline in the number of workers in that particular industry. But in this case consumers will be spending a smaller

[1] Leo Wolman, *Recent Economic Changes*, Vol. II, p. 478.

total amount of money on this particular good, and the purchasing power released by the change will be expended on other goods and services. Under either alternative therefore the final effect of the change on employment is not an absolute reduction of the number of workers employed but at most a shifting of demand from one labor group or one locality to another. If other alternatives are examined it will be seen that these ultimate consequences remain the same. The new invention or scientific discovery may create an entirely new industry which undermines an older one, as the automobile has reduced the manufacture of horse-drawn carriages, or the radio has brought hard times to the vaudeville stage. But these developments when they have worked out their effects merely reduce the demand for one kind of labor and increase that of another. Or we may assume, as is sometimes argued, that the gain from a cost-reducing improvement is not passed on to the consumer but retained by the owner of the industry. Even in this case the owner will spend his monopoly profits either through the investment market or for consumable goods, thus increasing demand for labor at some point in the productive system.

In the long run therefore improvement of production methods cannot permanently lessen the demand for labor as a whole; it merely redistributes the demand among the vocational groups which make up the wage-earning population. This final judgment on the matter is important as pointing the way to appropriate public policy. But it does not disprove the existence of an immediate problem. Each case of technological unemployment may be temporary in the long-run view of things, but it is none the less disastrous at the moment to the workers affected by it. It may even be permanently disastrous to certain groups whose crafts are rendered obsolete by new machines and whose means of livelihood on the former level of comfort have vanished. Even for society as a whole the net result in eras of broad scale and continuous change is not temporary but prolonged unemployment. The individual

cases add up to a large total of people out of work at any given time; and the unceasing march of technical progress replenishes the army of the unemployed year after year as different individuals drift in and out of its ranks.

Seasonal unemployment. Unequal distribution throughout the year either of the sale of the product of an enterprise or of the supply of raw materials it utilizes results in seasonal changes in the demand for labor. The causes of these fluctuations are so many and so various that no detailed account can be given of them here. In some cases it is habit and custom — fashion, the observance of holidays, etc. — which alter the demand of different seasons for the products of labor; in others climatic changes affect either the demand for some product as clothing and fuel, or the supply, as in agriculture, or the facility with which production can be carried on, as in the building trades. Very few industries are entirely immune from seasonal influences, either directly or through the medium of other industries initially affected.

That these rhythmic fluctuations in the tempo of business activity create unemployment goes without saying. In the coal industry 40 per cent of the workers employed in the best month are idle in the worst; employment in the construction industry has fluctuated as much as 79 per cent in the course of a year; in retail trade by 30 per cent; in clothing manufacture by 35 per cent; in automobile manufacture by 24 per cent.

It is characteristic of the irrevocably seasonal industries to attach to themselves a supply of labor equal to the demand at the highest seasonal point. Relatively high wages, bonus payments, overtime pay offered by employers to stimulate output at these peak points mislead workers into the belief that these vocations are unusually profitable. So there accumulates a *reserve of labor*, or a supply of workers in excess of the number that would be required to produce the same annual output on a stabilized plane of operations. In coal mining, for example, it was found at one time that 200,000 workers would be

continuously idle if production were spread evenly throughout the year. This wasted labor is the ultimate social loss from the seasonality of industry. From the point of view of the worker the loss is self-evident.

Cyclical unemployment. The business cycle has been studied in another place. Its effect on employment varies with its different phases: an increasing demand for labor during the recovery phase; a relatively full and steady demand during the prosperity phase; declining demand during recession; abnormally small demand, more or less prolonged, during depression. How extreme may be the resulting unemployment in the trough of a depression is shown by the figures given on a preceding page for 1921 and 1933. When things were at their worst following the crisis of 1929 a quarter of American wage earners were out of work, many millions were destitute, and the combined resources of city, state, and national governments were strained to keep the people from starving.

Depression ultimately finds its way into recovery, and cyclical unemployment gradually disappears. But recovery from a bad depression always brings with it far-reaching readjustments within the industrial structure, causing some forms of economic activity to shrink below previous levels, others to expand. The restoration of the demand for labor therefore may never reach some occupational groups, and many people may suffer unemployment prolonged far into the recovery period or be obliged to shift to new occupations or new localities to their enduring loss.

Remedies for unemployment. Under this title we shall include only those policies which, by attacking the causes of unemployment, show promise of working permanent improvement. Each type of unemployment calls for remedies pertinent to its own causes. But there are certain policies or programs of action which may be applied to the problem as a whole. All types of unemployment result in the last analysis from the failure of the labor supply to adjust itself promptly

to changes in demand. While steps may be taken to control the abruptness of these changes, it is obviously profitable also to adopt measures to reduce the lag in the response of labor to them.

Greater mobility of labor. This requires that wage earners be ready and able to desert localities or occupations whenever shifts in demand make them relatively overcrowded. A few moments' thought will disclose the obstacles which prevent this response. Wage earners are human beings, attached to their places of residence by those ties of sentiment, human relationships, and material interest which bind all social groups to their home communities. In the skilled trades the individual's knowledge of the craft is a possession not applicable to other vocations and therefore lost when he changes his calling. Moreover willingness to make these sacrifices is not sufficient to enable the worker to move from one vocation to another, for it is often impossible for him to acquire the new skills demanded of him in the new line of work. Any policy which increases the worker's mobility in the geographic sense or his versatility in the craft sense will reduce the amount of unemployment.

Public employment offices. Frequently workers simply do not know that jobs are more plentiful in another locality or another vocation, and they may have no reliable sources of information on these matters. The employment market can be better organized, so that obstacles to the redistribution of labor due to mere ignorance of the facts will be removed. Organization of the markets for securities and the staple commodities has been perfected through exchanges which bring demand and supply continuously together. The same principles can be applied advantageously to the labor market by means of the employment office and the labor exchange. A complete system of employment offices would cover the entire industrial area with an intercommunicating series of local agencies, each registering and classifying the idle workmen and the available jobs in its locality, and all coöperating to

distribute unemployed labor over the entire area with as little delay as possible.

Most modern industrial nations, except the United States of America, have long had efficient nationwide systems of this sort under public management. In the United States until recently this service had been left to the individual states or to private profit-making agencies. A few of the states worked out fairly efficient systems, but for the most part the state employment office has been a perfunctory affair, staffed by political job holders. The private agencies, since they profit from a continuance of the evil they purport to remedy, are frequently even worse than this, adopting practices which partake of outright fraud.

The passage of the Wagner-Peyser act in June, 1933, marked the beginning of a new national policy with regard to this problem. This law created the present United States Employment Service with a central office in Washington, staffed with trained people under a director-general, who has the assistance of an advisory council representing employers and workers. To stimulate state action the law empowered the federal government to make an annual appropriation of $4,000,000, of which $3,000,000 is allotted to the states according to population, on condition that each state which accepts the grant shall match the federal subsidy dollar for dollar and that the state's system of employment offices shall measure up to standards defined by the federal service. The remaining $1,000,000 is used to conduct an interstate clearing house of information, to inspect the state services, to carry on research and publication, and to render certain direct services to labor in the District of Columbia and among the war veterans.

At first the states were slow to respond to the invitation of the federal government to revamp their employment offices and affiliate with the United States Employment Service. By June, 1935, only 25 states had done so. But the passage of the social security act in 1935 hastened the process. This law

embraces a broad group of policies which we examine in another place; but we must note here that its provision of federal aid to state unemployment compensation systems, by necessitating the creation of state-wide employment offices to record the unemployed and distribute the compensation, has spread the public employment office system throughout the entire country. By 1938 all the states and territories had set up these systems and all were affiliated with the United States Employment Service; at this time there were 2,000 offices in the system, employing a staff of 15,000 people.

During the brief period of its existence the service has been to a large extent deflected from its normal long-range functions to discharge placement services required of it by various federal emergency agencies such as the CWA and the PWA. Its record nevertheless is impressive. From 1933 to 1938 some 26 million unemployed people were registered in the employment offices of the country as applicants for jobs; about 22 million placements were made, including those in the various work relief services. Toward the end of this period an increasing proportion of placements was being made in private industry.

Greater flexibility of wage rates. As in the case of other goods, the behavior of the price of labor has much to do with the adjustment of supply to changes in demand. A falling wage rate in a locality or a vocation is at once the most accurate sign of oversupply and the most effective stimulus to its correction. A rigid wage rate set above the marginal productivity of a labor group will create and tend to prolong unemployment at that point in the market. Wage rates are at best sluggish prices; their rise and fall is retarded by sentiment, by custom, by contract. Any arbitrary or mandatory policy which makes them still more rigid and unyielding must be viewed with suspicion. Such policies — legal minimum wages, code agreements pegging wage rates, and the like — are frequently advocated by well-intentioned people, usually in times of distress and always in respect of some labor group

whose low or diminishing rate of pay reflects the oversupply of their services. It may be desirable for moral or humanitarian reasons to adopt such policies on occasion, but this should always be done in full knowledge of their danger as a possible aggravation of the problem of unemployment. We may take it as a sound principle that anything that increases the plasticity of the wage rate will tend to prevent or diminish unemployment.

Stabilizing the demand for labor: Voluntary action by industry. Specific remedies for the different types of unemployment take the form of attempts to stabilize the demand for labor. These attempts may be made by employers acting voluntarily, or by the government acting independently outside the system of private industry. Voluntary programs of this sort require organization on both sides of the labor market, collaboration among competing employers, and collective action by the workers. We have space here only for a brief summary of these policies of the business world.

The interest of society as a whole in technological progress is so obvious and so paramount that no proposal to stabilize industry by putting an end to invention and improvement is worthy of consideration. But for this very reason — because the gain goes to society as a whole — it is now being recognized as neither just nor expedient that the cost should be borne solely by one social class. Policies can be adopted by agreement between employers and workers which, while permitting technical advances, will slow down their rate and provide time to forestall unemployment. The linotype machine furnishes the classic example. A rapid wholesale adoption of this invention in the printing trade would have destroyed an entire craft and caused much distress. By agreement it was introduced gradually, the former typesetters were trained to operate it, and the result was that no unemployment occurred even temporarily, while ultimately the demand for labor was greatly expanded through the decline of printing prices. The

moulding machine was introduced into the iron industry under similar controls. Very recently the textile workers and the national association of their employers have attempted by joint agreement to provide for more gradual adoption of the automatic loom.

The sharply seasonal industries are experimenting with policies of stabilization in harmony with their differing technical conditions. The following is a summary of the more important of these devices. (1) An industry whose major product is subject to fluctuating consumers' demand may take on *compensating side lines*, whose market varies inversely with that of the major product. An example is the manufacture of felt hats or cloth caps by a concern specializing in straw hats, or the production of some staple product by a concern devoted primarily to the manufacture of Christmas specialities. (2) Somewhat similar is the development of *compensating markets*. Cement manufacturers have regularized their output by exploiting the South American markets where the seasons are the reverse of those at home. Much can be done on a smaller scale along this line within the continental area of the United States. (3) Goods may be *made to stock* during the slack months. This device involves expense of storage and can be used with safety only in connection with types of product which do not deteriorate in storage and are not subject to capricious changes of fashion. (4) Sales methods may be adjusted to encourage *advance orders* with a long delivery time. By offering particularly favorable terms, a large concern specializing in the Christmas trade has succeeded in inducing many of its customers to place their orders more than a year in advance, thus allowing a stabilizing of production. Similar results have been obtained by many of the larger clothing manufacturers. (5) Companies which perform several different processes may set up a system of *vocational training*, increasing the versatility of their workers so that they may be transferred from slack to busy departments without loss of time.

Public works: Theoretical analysis. Remedies for cyclical unemployment have been suggested in connection with our study of the business cycle and need not be treated in detail here. It is in relation to this phase of the unemployment problem that governmental activity is usually proposed. The program of public works which has figured so largely in the activities of the American national government in recent years is one example of this type of policy.

The public works program contemplates governmental activities on a large scale during periods of general business depression, involving works which would not otherwise be undertaken at all, such as slum clearance, reforestation, road building, irrigation projects. As advanced by serious students of the problem, the program is carefully safeguarded against the most obvious dangers involved in extravagant governmental spending. It is to consist entirely of projects worth while for their own sake; it is to be financed as much as possible from funds reserved for the purpose during prosperous years, thus avoiding the necessity of heavy taxation or large public borrowing during a time of unsettled business; it is to be administered in a businesslike way, employing laborers on the basis of their usefulness to the enterprise, not merely because they are unemployed, and holding them to the same standards of wages, working conditions, and performance demanded by efficient private enterprise. In addition to these safeguards, a sensible public works program must be based on reliable statistics of unemployment, in order to permit of advance planning in prosperous years and to regulate the scope and the regional location of the government's activity in harmony with the trend of unemployment. Assuming that all these features of the program can be assured, there is something to be said for such a program as a corrective of cyclical unemployment.

As things go in the political world there is grave reason to doubt that any such businesslike handling of so large and complex an undertaking could be obtained through the agencies

of government. But even if wise administration were assured, it is still to be feared that a public works program would exert injurious indirect effects upon the situation. If funds reserved for public works in prosperous years are invested in income-bearing securities, grave difficulty may be encountered in liquidating these reserves when the money is needed, and the attempt to do so in a falling market may add to the difficulties of the business world. The ultimate solution of business depression must be such readjustment within the price structure and correlatively within the system of production as will bring economic forces to a new point of equilibrium. With reference specifically to unemployment we have seen that anything that increases the rigidity of wage rates will impede solution of the problem. A gigantic public works program can hardly fail to create an artificial demand for certain types of goods and labor which will tend to reduce the flexibility of the price structure. To prevent this sort of unforeseen and indirect effect of governmental intervention in economic affairs is virtually impossible, even by the most careful administration.

Public works program of the United States. The merits of an ideal public works program have never been put to the test of experience, for nowhere in the world has such a program been attempted. It is true that a vast program of public spending has been undertaken by the United States government as a remedy for unemployment and as a device to stimulate business recovery, beginning with the national industrial recovery act of 1933, which appropriated three and a third billion dollars for this purpose. This and subsequent appropriations extending through 1938 have been spent, in part as subsidies to state and local governments to assist in financing approved projects designed and administered by these governments, and in part on a number of large-scale enterprises undertaken by the federal government itself. The program is under the immediate direction of the President and the Secretary of the Interior, who must approve all projects proposed by the state and municipal authorities and see

to it, as a condition of federal subvention, that these projects receive equivalent financial support from the local governmental units.

This experiment has undoubtedly done much for the relief of distress and probably exerted some influence in the business recovery of 1934–37. But however that may be, this experience did not constitute a real test of the public works method of solving unemployment. The program was adopted as a substitute for an existing policy of unemployment relief, which was draining the resources of governments, local, state, and national. It had no such background of long-range planning as is proposed as a fundamental feature of the ideal public works program. For these reasons the labor employed was selected or managed, not by tests of efficiency, but as objects of charitable relief; and the projects undertaken have been devised on the spur of the moment in the midst of a great national emergency. Moreover the program was financed, not out of reserves accumulated over previous years, but by current taxation and borrowing at a time when the investment market was already disturbed and the taxpayers already impoverished by the prolonged business depression. In all these important respects the present activity of the government fails to qualify as a true test of long-range planning of public works.

Child and woman labor. The labor of women and children is not a phenomenon peculiar to the modern system of industry. It is characteristic of all types of societies at all times. That it should be viewed today as a social problem is due chiefly to the fact that gainful labor by these members of the population has economic consequences in a society like our own which did not result from it in former times. We shall not stress the moral aspects of the problem, or the health hazards encountered by women and children in industry. There are also various economic consequences which need only be mentioned, such as inability of the child prematurely employed to acquire the skill demanded of the higher wage group; the

likelihood that he will be lodged inescapably in a blind alley occupation; the resulting maldistribution of the labor supply as between the overcrowded and the undersupplied occupations. Similar problems arise in connection with the employment of women.

Employment statistics. No official figures were kept prior to 1870 of the amount of child labor. The United States Census of that year reported 739,164 children between the ages of ten and fifteen at work in America. The numbers in the succeeding decades were as follows:

CHILDREN AGED TEN TO FIFTEEN GAINFULLY EMPLOYED, UNITED STATES, 1880–1930 [1]

Year	*Total*	*Percentage of all gainfully employed people*
1880	1,118,356	6.4
1890	1,503,771	6.5
1900	1,750,178	6.0
1910	1,990,225	5.2
1920	1,060,858	2.6
1930	667,118	1.4

These figures are based on the official reports of employers to the census takers. No doubt many child workers were unreported; and it is estimated that the inclusion of the ages sixteen and seventeen would double the figures. Nevertheless the outstanding feature of the table is the rapid decline of child labor since 1910, both in the aggregate and relatively to the total working population.

Agriculture has always been the chief employment of child labor. In 1930 agriculture accounted for 45.4 per cent of all gainfully employed children, as compared with 21.7 per cent in manufacture, 3.5 per cent in transportation, 8.8 per cent in trade, 6.7 per cent in domestic service, and 8 per cent in clerical occupations.

[1] Adapted from *Census of Occupations*, Abstract Summary, 1932, pp. 3–5, United States Bureau of the Census.

On the other hand, the number of women engaged in industry has rapidly increased, except in the decade 1910–19. The official figures are as follows:

GAINFULLY EMPLOYED WOMEN AGED TEN YEARS AND OVER, UNITED STATES, 1880–1930 [1]

Year	*Number*	*Percentage of all women of age group*	*Percentage of all employed persons*
1880	2,647,000	14.7	15.2
1890	4,005,000	17.4	17.1
1900	5,319,000	18.8	18.3
1910	8,076,000	23.4	21.2
1920	8,550,000	21.1	20.5
1930	10,752,000	22.0	22.0

The four-fold increase in the number of women workers during this period outstripped by far the rate of increase of new workers. In 1930 a fifth of all the wage earners in this country were women. Analysis of the figures for this year will show that there is scarcely an occupational field in which women do not make up a substantial part of the labor supply.

PER CENT DISTRIBUTION BY SEX AMONG THE PRINCIPAL OCCUPATIONS, UNITED STATES, 1930

Occupation	*Male*	*Female*
All occupations	78.0	22.0
Agriculture	91.3	8.7
Manufacturing, total	86.6	13.4
Clothing factories	29.8	70.2
Textile mills	56.7	43.3
Trade	84.2	15.8
Professional service	53.1	46.9
Domestic service	35.8	64.2
Clerical	50.6	49.4

Wages of women and children. The feature of the situation which makes child and woman labor — particularly the latter — a distinctively modern problem is the effect of their

[1] Taken from *Census of Occupations*, Abstract Summary, 1932, pp. 3–5, United States Bureau of the Census.

competition on the welfare of men workers, tending as it does to undermine wage standards throughout the employment market. All investigations agree in the conclusion that the pay given to women in industry is very low. In 1914 the average weekly wage of women in factories was $7.75; in 1921, $15.63; in 1929, $17.61; in 1937, $16.95. Such wages are inadequate to provide a wholesome standard of living for the woman worker.

But the most serious feature of the problem is the discrepancy between the wages of women and those of men. All studies of this question concur in the finding that women are paid less than men for the same task and in the same industries. One investigator, comparing the levels of wages in 1899 and again in 1923, discovered that in both years women's annual earnings were 54 per cent of men's in comparable employments. The National Industrial Conference Board found this same ratio in 1929 between the wages of women and men in skilled employments, while in unskilled occupations women received in that year 72 per cent as much as men. The best available information indicates that this differential was about the same in 1937 as in 1929. Some part of this discrepancy is to be explained by the inferior productivity of the woman worker; but when all possible allowance has been made for this factor, it still remains true that women are consistently paid less than men for no other apparent reason than that they are women.

This systematic underpayment of women cannot be accounted for except on the ground of their weaker bargaining power. As compared with men women are notably weak in asserting their claims against the employer. Moreover they are even more immobile than men; it is difficult for them to play off one employer against another when this involves a change of residence. In many cases also they are not self-supporting and make no pretense of demanding a "living wage." Because they are partially subsidized by the male wage earners of their families, they are not always intent upon

exacting the highest possible price for their labor. Finally they are unorganized and to some extent unorganizable; their bargains are made as individuals, and they labor under all the handicaps which the isolated worker must bear in the modern labor market.

Solution of the problem of child and woman labor. To the extent that it is desirable to prevent or regulate child labor, the necessary steps to this end must be taken under governmental compulsion. The problem is one which will not be solved by the free play of economic forces. It is unfortunately true that many parents are eager to get income from the labor of their children at as early an age as possible. The children too are often eager to go to work as an alternative to going to school. Employers in a competitive industry adaptable to the labor of children can scarcely refuse, as individuals, to employ them, lest the competition of the less scrupulous competitors drive the others out of business. Here then is an appropriate field for public regulation; what is being done will be described in the following chapter.

It is neither possible nor desirable to prohibit wage labor by women. The forces in modern society which are impelling women to seek their livelihoods in the vocations and professions formerly occupied solely by men are far too powerful to resist. They spring from currents of change which are fundamental in our developing culture — the trend of education, of moral and social ideas, of life in the urban home. Moreover it is certainly better both for the women and for society that women should be occupied in useful labor rather than suffer the inanity and stultification of enforced idleness. As in the case of evils resulting from child labor, some of the abuses which develop in the employment of women in industry can be removed only by legislative means. However, except in the more extreme cases, such as occupations hazardous to health and morals, betterment of the conditions of the woman worker must be brought about by the action of these workers themselves. Particularly with regard to the underpayment of

women, both in the absolute sense and relatively to the wages of men, permanent improvement will come only through devices which increase their bargaining power. Experience has shown that organization for collective bargaining is an effective means to this end.

Industrial accidents. One of the principal costs involved in the modern system of industry is due to the temporary or permanent destruction of the productive power of the workers through injuries resulting from the normal hazards of their occupations. Only within recent years has serious attempt been made to estimate the magnitude of this cost. The official statistics of work accidents in the United States are even now so unreliable and incomplete that any statement of the total loss can be no more than an estimate. In some states no records at all are kept. Certain industries have accurate records of the fatalities resulting from the hazards of the occupation, though the less serious accidents which resulted in temporary disability are allowed to go unrecorded. Even on a conservative estimate the loss can be shown to be surprisingly large. To give the result of only one study out of many, the National Safety Council found that in 1929 industrial accidents caused 20,000 deaths. The number of minor accidents has been variously estimated, usually in the neighborhood of 2,000,000 a year.

In recent years strenuous efforts have been made to reduce this waste. Factory acts and compensation laws have made it obligatory or profitable for individual owners of business enterprise to safeguard the lives and health of their employees. Groups of employers in the major lines of industry have cooperated through their trade associations to perfect devices for the prevention of accident and to urge their adoption upon the more indifferent individuals among their number. In the United States this effort has given rise to a widespread "safety first" movement, which enlists the services of experts and adopts many ingenious devices to make its program effective. Similar movements have been under way for decades in other

countries. The most significant result of these activities is their inability to solve the problem of occupational hazards. Indeed accidental deaths and injuries in industry have increased during this period, both in absolute figures and in proportion to the numbers employed. They are caused by the nature of industry itself — the increasing mechanization of the productive process — to which the human mind and body seem unable to adjust themselves. As the basic cause is social, so is the ultimate loss — the temporary or permanent destruction of productive power.

Hours of labor: The shorter work day. In industry as a whole a strong trend toward the shorter work day became apparent over a century ago and has continued down to the present time. The effect of this movement in America upon normal hours of labor in succeeding decades from 1840 to 1930 is shown in the following table:

NORMAL HOURS IN MANUFACTURING AND MECHANICAL ESTABLISHMENTS, UNITED STATES, 1840–1930 [1]

Year	*Average normal hours per day*
1840	11.4
1850	11.5
1860	11.0
1870	10.5
1880	10.3
1890	10.0
1900	9.8
1910	9.4
1920	8.5
1930	8.3

The depression following 1929 produced an abnormal situation as regards hours of labor, since many of the policies adopted by individual employers or governmental bodies to combat unemployment were of the share-the-work variety, involving part-time employment for many workers. In response to these influences there has been a strong movement

[1] Taken from C. R. Daugherty, *Labor Problems in American Industry*, 1933, p. 241.

toward a five-day week with not more than eight hours per day. In 1931 less than 800 concerns, with a total labor force of some 700,000, were known to be operating on the five-day week schedule; in 1935, 250,000 firms, employing 12,000,000, had adopted this schedule. Of course too much reliance must not be placed on these developments during depressions as forecasting the future trend.

Hours and productivity. The major problem presented by this tendency toward shorter hours is the effect on wages. When workmen demand a shorter day they expect the rate of wages to be raised simultaneously so that their earnings may remain unchanged. Now it has been proved historically that earnings can be maintained or even raised while hours are reduced. During most of the period covered by the preceding table real wages rose steadily while hours of labor were declining. To what extent this can continue must depend in the final analysis upon the effect of the shorter work day on the productivity of the workers.

Within limits a reduction of hours has been found to increase the productivity of labor. It has been proved, by the experience of many different enterprises in Europe and America, and established as a rule of wide applicability, by researches carried on by the British government and by various bureaus of investigation in America, that a shortening of the more extreme work day, a fall of hours from twelve to eleven or ten for example, results in an actual increase in the total output of labor.[1] To be sure, this effect of shortening the long work day is not uniform for all grades of workers. It is most noticeable in the case of those occupations which demand continuous mental alertness combined with manual dexterity and is least so of the automatic types, such as the tending of machines whose pace is governed by the flow of steam or electrical power. But even the most unskilled labor responds to a

[1] The evidence in support of this conclusion is summarized in the *International Labor Review* for April and November, 1924. See also H. M. Vernon, *Industrial Fatigue and Efficiency*, London, 1921.

reduction of an extremely long day by some increase of output.

There are many reasons for this. When the length of the work day is excessive, the period of rest is not sufficient to counteract the effects of fatigue. In extreme cases accumulated fatigue predisposes the workman to sickness, which results in lost time, and where this consequence is avoided the workman tends to be slack in attendance through voluntary idleness. It has been found that the accident rate varies directly with the length of the work day, reaching a maximum during the late hours, and also that it rises with the days of the week, thus disclosing another waste due to fatigue. Waste of materials in the form of defective output, waste of power, and necessity of more labor of superintendence through increased slackness of the workers also result from protracted hours of labor. A saving of these wastes can be effected by shortening the hours of work to avoid the worst effects of fatigue. It is probable that the ten-hour day is more productive at the present time for all grades of labor than is the twelve-hour day and that for the most skilled groups an eight-hour day is at least as productive as ten hours of work.

But eventually the shortening of hours of work must strike a point where productivity of labor is not increased but diminished, unless this effect is forestalled by advances in the technique of production. Given continuous technological advance, the workers may choose between two alternative ways of taking their share in the general improvement. They may elect to receive their former rate of real income and take their gain in shorter hours of work or, working the same number of hours as before, to raise their standard of material comfort. During the past decades their actual choice has taken a middle ground between these extremes. As their bargaining power has increased, the workers have pressed for a shorter work day and for a simultaneous increase of wages. That they have won both concessions is to be attributed, in part to the positive increase of productivity resulting from reduction of long hours

which we have discussed, in part also to the fact that these demands were made at a time when industrial progress was working in the direction of increasing the hourly productivity of labor. For some groups of workers the shortening of the work day has maintained their productivity at its maximum; for others it has resulted in an output below the maximum of which they are capable on the present stage of the industrial arts but higher than was possible on earlier stages.

CHAPTER XXXVI

Labor Legislation

Unemployment insurance: General nature and origin. The movement to protect workers as far as possible against the hazards of unemployment is supported by sound business considerations as well as by sentiments of justice and fair play. It seems manifestly unfair that a loss attributable to general social forces should be thrown upon the one class in society least able to bear it. Moreover it is a well-recognized fact that the fear of unemployment acts as a deterrent to the efficiency of workers, inducing them to adopt job-saving or time-wasting practices. Society would gain much by the removal of this fear as an offset to the cost of administering the policy of insurance. Finally, although insurance is strictly speaking only a means of distributing the loss, the policy may be so devised as to reduce the amount of unemployment, thus preserving for society a portion of the productive power which would otherwise be totally destroyed through the idleness of workers. Certainly little progress can be made in the solution of unemployment as long as the problem itself is concealed from view. An insurance system could disclose the extent and location of unemployment; and its details could be so adjusted as to make it worth while for employers to keep their employees continuously at work. These considerations of justice and economy have caused most modern nations to adopt some form of unemployment insurance. In eight European countries as well as in Australasia insurance systems made compulsory by law cover over 35,000,000 wage earners. In other European countries voluntary systems receiving financial support from the government apply to an additional 4,000,000 workers.

Unemployment insurance began with the efforts of the

trade unions to provide benefits for their members when out of work. The procedure followed is simple. Membership dues are set at a figure large enough to accumulate a fund out of which to pay benefits to the unemployed. As a safeguard against fraudulent claims the benefits are disbursed by local officials who know the applicant personally, and maximum limits are set to the duration of relief in individual cases. These voluntary systems are to be found in practicable operation in all countries where trade unionism has attained strength. They have obvious defects as a method of handling the problem. The contributions made by the members are not based upon actuarial data; the reserves set aside against future benefits are inadequate to meet the drain of widespread unemployment, and accordingly the systems collapse when the need is greatest; at best only a minority of the working class (*i.e.*, the part which is organized in unions) is benefited. Furthermore, since employers bear none of the cost of relief, trade union insurance adds nothing to their incentive to keep their workmen employed.

The Ghent system. Public insurance began as a half-way measure, taking the form of subsidies from the government treasury to support systems of trade union insurance. This combination of mutual and public insurance has come to be known as the *Ghent system*, taking its name from the city which first adopted the policy. Since its beginning in 1901 the system has spread widely among the municipalities and larger governmental units of Europe, until at the present time it is a well-established policy in Belgium, Czechoslovakia, Denmark, Finland, France, Holland, Norway, Spain, and Switzerland. In all these countries the policy has similar features. Public contributions are made on condition that a substantial portion of the cost — usually one-third or one-half — be borne by the trade union; administration is placed in the hands of union officials; certain standards determining the amount and duration of benefits are set by the contributing governmental body.

The Ghent system is superior to simple trade union insurance because of its greater financial adequacy; otherwise it has all the weaknesses of that type of insurance mentioned above. Furthermore it raises serious problems when applied in countries where employers oppose independent trade unionism, for by basing its policy upon the unions, the state will assist them to enlarge their membership and so will appear to be taking sides in the controversy. The failure of voluntary methods to provide adequate relief for the ever growing burden of unemployment has led many governments to pass laws making insurance compulsory.

Compulsory insurance in Great Britain: Original plan. Great Britain was the first important country to adopt the policy of compulsory unemployment insurance and the one country which has given the system thorough trial on a comprehensive scale and over a long period of time. The original law, passed in 1911, applied only to a few selected trades and covered about two and a half millions of the total fifteen million wage earners in the country. This law created a central unemployment fund, made up of equal contributions from the wage earners, the employers, and the public treasury. The wage earners' contributions were made weekly and varied according to the age and sex, but not the trade, locality, or wage rate, of the individual. These contributions were held back by the employer from the workers' pay and passed on to the government, along with an equal contribution from the funds of the business; these funds were again passed on to the Unemployment Fund, augmented by a grant from the treasury of the same amount as the workers' contribution. The payment of benefits was carefully safeguarded. No insured worker was eligible for benefit unless he had made at least thirty payments into the fund during the preceding two years. A simple device was adopted to provide proof on this point. Employers kept a book for each insured worker, in which he affixed a stamp each time a weekly payment was made. This book was given to the worker when he lost his

job and constituted his proof of eligibility. The unemployed worker was required to register immediately in the nearest employment exchange and was obliged to accept any suitable employment offered him, the term "suitable" being carefully defined to protect the worker against being used as a strike breaker or forced to accept conditions of employment materially below the standards of his craft. A waiting period of two weeks was required before any benefits were paid; and the duration of the benefit period was limited to one week for every six weekly contributions previously made by the individual and to a total of fifteen weeks in any year. The system operated to reduce unemployment, as well as to relieve distress, by giving the employer a refund of one-third of his payments for every worker continuously employed in each year.

The chief administrative features of the system have remained unchanged to the present time and are still in operation. The rates both of contributions and of benefits have been altered from time to time. They have never been large; in 1937 the workers' contributions ranged from eighteen cents per week for adult males to nine cents for youths, and the weekly benefits from $4.25 for adult males and $3.75 for adult women to a bottom figure of $1.25 for youths, with small extra allowances for dependents. But these minor changes of detail have been of little consequence as compared with certain sweeping alterations in the scope of the system and in the rules of eligibility for benefits. For ten years the system operated successfully on an insurance basis; contributions into the fund had been large enough to meet all demands for benefits and leave a surplus of $1,000,000,000; benefits were not "doles" but businesslike payments to which the recipient was entitled by insurance contract.

Breakdown and reform. Beginning in 1920 and 1921 the situation underwent a radical change. In the first place, the number of workers covered by insurance was greatly increased, without adequate provision of reserves against the

increased risk. The amendment of 1916, which added some two million workers in war industries, caused little trouble, because there was no great increase of unemployment during the next few years. But in 1920 the system was abruptly broadened to include all wage earners receiving less than $1,250 a year except those in agriculture and domestic service, thus increasing the number insured to about twelve million people. This was done on the verge of a severe and prolonged business depression, so that there was no time to accumulate reserves for these new workers before great numbers of them were thrown into unemployment. Moreover as unemployment became more widespread, the standards governing benefits were relaxed through compassion or in response to political pressure. The waiting period was reduced from two weeks to six days; the duration of the benefit period was extended from fifteen to twenty-six weeks; the thirty contributions required for eligibility were not limited to the preceding two years but could be counted over any previous period. Even these changes did not suffice to meet the pressure on the system, and the labor exchanges were authorized by the government to go on paying benefits to unemployed individuals after the twenty-six week period had passed; that is, after the insured had exhausted his legal claim. Financial exigency led also to the abolition of the practice of refunds to businesses which provided stable employment, thus robbing the insurance system of its incentive to reduce the amount of unemployment. It must be said that the British government began this policy of paying excess benefits on the understanding that it was making temporary loans to the fund which would be repaid when the emergency was passed; but continued unemployment postponed the return to the original principles of the system until the government found itself in 1931 carrying an impossibly heavy financial load. By this time the fund was in debt to the government to the extent of $500,000,000.

In that year drastic reforms were made. The benefits were reduced in amount and the contributions increased; eligi-

bility rules were made more severe; and, most important, additional benefits to workers who had exhausted their legal claims were discontinued. These victims of prolonged unemployment were transferred to the charitable relief agencies of their local governments to be treated as cases of pauperism. The reforms had to be carried through in the face of violent political opposition, which has continued down to the present time. The insurance system is now reëstablished on a businesslike basis with all the "dole" features removed, but it has become an issue in partisan politics and may again succumb to popular pressure.

Lessons of the British experience. The British experience discloses both the merits and the weaknesses of compulsory unemployment insurance. It has proved that in times of fairly stable industrial conditions unemployment is an insurable risk susceptible of treatment in a businesslike way. Under the proper safeguards it does not promote malingering, as has been feared, and it does act to reduce unemployment by stabilizing the demand for labor. But it also shows how vulnerable are these safeguards, particularly in times of widespread distress, and how readily the system may give way to the demand that insurance be turned into a policy of charitable relief. It seems clear that unemployment insurance will not stand the strain of severe business depression. Even if the political dangers of such times be avoided, cyclical unemployment cannot be handled by insurance methods. The cost will outrun any reserves that can be accumulated in advance. Even if enormous reserves are accumulated during long periods of good times, the attempt to liquidate them in an unsettled money market must tend toward further depression of business conditions.

Unemployment insurance in the United States. Under the American constitution the policy of unemployment insurance lies within the province of the several states. The federal government may take steps to encourage the adoption of state-wide systems by offering subsidies out of the national

treasury; and if these subsidies are accepted it may assume supervision of the standards and methods employed in the state plans. The social security act of 1935 makes use of these federal powers.

This law embraces a broad program of social legislation, including provisions for old age pensions, care of the blind and crippled, maternal care and child welfare, as well as compensation for the unemployed. To handle the new insurance and benefit programs the Social Security Board was created, consisting of three members with headquarters and a large staff in Washington and in twelve regional offices in various parts of the country. The law appropriates $4,000,000 during the first fiscal year and $49,000,000 in each succeeding year to be allocated to the states as a contribution toward the administrative cost of such insurance systems as they may set up. As an additional inducement to the states, the act imposes a tax on the payrolls of every business employing eight or more workers, at rates rising to 3 per cent in the third and each succeeding year, and provides that the employer may offset against this tax 90 per cent of any payments he makes to compulsory insurance funds under his state law. The law allows the states complete discretion in setting up systems of insurance suited to their own requirements or rejecting the policy outright but requires that to qualify for grants from the federal fund the state system must meet certain tests with regard to its administrative and benefit features, that funds accumulated by the states must be deposited in the Federal Unemployment Trust Fund of the United States Treasury to be invested in federal securities, and that the state system must report regularly to the Social Security Board and hold its records continuously open to the board's inspection. The act was declared constitutional by the Supreme Court in 1937.

At the present time the forty-eight states, the District of Columbia, Alaska, and Hawaii have adopted approved unemployment insurance laws. The different systems exhibit

such great variety in matters of detail that it is impossible to do more than summarize their chief features. In all cases certain groups of workers are excluded by definition: agricultural labor, domestic servants, public employees, employees of religious, charitable, and educational institutions, and some others of minor importance. Moreover in all but ten states workers in small concerns — usually defined as consisting of less than eight employees — are not insured. It is estimated that some twenty million wage earners in all are covered by the insurance systems. In all but seven states, where the workers are required to make small contributions, the cost is borne by the employers in the form of payroll taxes. All the laws require a waiting period after the loss of employment before the benefits begin; all set a maximum period, frequently sixteen weeks, on the duration of benefits, requiring a minimum period of reëmployment before additional benefits will be paid; and all adjust the benefits to the previous rate of earnings with a stated maximum, usually fifteen dollars a week.

These insurance systems have been in operation for too short a time to be tested by experience. In fact benefit payments were not available in any state until February, 1938, and it was not until July, 1939, that all the systems became effective. Accordingly no real appraisal of the policy is possible at the present time.

Legal regulation of the labor of children. Reasons were advanced in the preceding chapter to show that the problem of child labor is not capable of solution either by the action of the wage earners or by the competitive forces of industry. Realizing this, all modern states have attempted by law to prohibit the worst abuses and to set standards in protection of children allowed to work. A comparison of the first child labor law of Great Britain with the present model code of that country will show how ideas of what is appropriate have changed in the course of time. This first law (1802) applied only to pauper children; it was considered a violation of indi-

vidual rights to legislate for others. The law prohibited the labor in factories of paupers below the age of nine; children over nine were allowed to work twelve hours a day. By the law of 1920 fourteen years is made the minimum age for industrial employment for all children; the length of the work day for young persons between sixteen and eighteen years is limited to eight hours; no work is allowed at night for any wage earner below the age of eighteen. Other European nations of advanced industrial civilization have approximately the same standards. These laws are sanctioned by heavy penalties and administered by an efficient system of government inspection, which makes evasion extremely difficult.

In the United States. The regulation of child labor in the United States lies in the hands of the states, and practically every state has a law on the subject. All but two states prohibit the employment of children under fourteen in factories and stores; five set the minimum age at fifteen; two at sixteen. Thirty-four states set a limit of eight hours to the working day for children under sixteen. Only one state fails to prohibit night work for children, either entirely or in selected industrial occupations. The laws do not apply to work in agriculture or domestic service, and only a few of the states have attempted to regulate the street trades of children. The practical effect of these laws varies, not only because of their different standards but also because of the wide difference in the rigor with which they are administered in different states. In general the southern states, where the factory system is of late development and where a necessitous population with low living standards creates a large supply of child workers, have the lower standards of regulation.

Lack of uniformity has been found to affect injuriously those units of a competitive industry which are located in the states having the more rigid standards of regulation, and this fact has tended to arouse opposition to child labor laws in these advanced states. Consequently various attempts have been made to bestow on the federal government power to impose

uniform standards upon the country as a whole. When the first federal laws regulating child labor, passed in 1916 and 1918, were declared unconstitutional, Congress attempted to give the federal government the necessary authority by amending the constitution. This amendment, submitted to the states in 1924, would bestow on Congress power to prohibit or regulate the labor of children under eighteen years of age; but ratification by the necessary majority of the states has not been obtained. The latest attempt to prohibit child labor by federal statute was embodied in the fair labor standards act of 1938. With numerous exceptions, industries engaged in interstate commerce are prohibited by this act from employing children under sixteen years of age, while in hazardous industries the age limit is set at eighteen years. This new policy has not as yet been tested in the courts, nor has it been in operation long enough to justify an opinion as to its probable effect on industry.

Regulation of the labor of adult women. Attempts to regulate by law the labor of adult women have encountered even more stubborn resistance from the advocates of individualism than have the laws dealing with children. In America it was not until 1908 that such regulation was declared constitutional by the Supreme Court in so far as it dealt with the hours and working conditions of women. The constitutionality of minimum wage laws for women, a policy discussed later in this chapter, is still in doubt. The action of the Supreme Court in upholding Oregon's ten-hour law in 1908 led to a rapid spread of this type of legislation. At the present time only four states set no legal limit on the work day for women. Thirteen of the existing laws fix the limit at eight hours; eighteen at nine hours; seventeen at ten hours; and the rest at eleven or twelve hours. In seventeen cases night work for women is prohibited, and most of the laws exclude women from designated hazardous occupations. All these regulative policies however specifically exempt farm work and domestic service from the legal control.

Minimum wage laws: General summary. Belief that wages too low to provide a decent standard of living are immoral or socially inexpedient has led governments to pass laws prescribing minima below which wage rates must not fall. Beginning with the legislation of New Zealand in 1894, this movement has spread through many countries until today sixteen of the principal industrial nations — among them Great Britain, France, Germany, Austria, and the British Dominions — have statutes of this character.

In the United States this form of social legislation has been in the hands of the several states except for the brief period 1933–35, when the wage codes of the national industrial recovery act were in effect. A Supreme Court decision in May, 1935, which held this action of the federal government unconstitutional, left the states again with sole jurisdiction over the matter. After ruling variously on the constitutionality of state minimum wage laws at different times, the Supreme Court declared such laws valid in March, 1937. The spread of state laws, which had previously been retarded by doubt as to their constitutionality, began again. In 1938 twenty-four states had adopted minimum wage laws, applicable (except in one state) to women and children only.

The United States fair labor standards act of 1938. The latest attempt of the United States federal government to establish minimum wages on a nationwide scale took the form of a law bearing the above title, which was approved on June 25, 1938. This law applies only to industries engaged in interstate commerce and, within this broad classification, exempts many specific occupations from its provisions. For example seamen, fishermen, agricultural workers, workers engaged in processing milk and other perishable foods, and street-car and bus-line operators are exempt. In the industries to which it is applicable the law attempts to establish maximum weekly hours as well as minimum hourly wages. It also prohibits the employment of children under sixteen years of age in these industries and sets the minimum age at eighteen

years if the industry is classified as hazardous. The provisions for maximum hours are as follows: 44 hours per week during the first year; *i.e.*, until November, 1939; 42 hours per week during the second year; 40 hours per week thereafter.

The basic minimum wages set by the fair labor standards act are 25 cents per hour during the first year of its operation, 30 cents per hour during the next six years, and 40 cents per hour thereafter. An essential administrative feature of the law however introduces a large degree of flexibility into these standards, with reference not only to the dates at which the minimum wages become effective but also to the applicability of the rates themselves to certain occupational and regional groups. The enforcement of the act is entrusted to a new officer in the Department of Labor, who is empowered to appoint advisory committees, in each branch of industry, to collaborate with him in adjusting the provisions of the law to the conditions of the business and region. These committees are charged with the task of hastening the adoption of the 40-cent minimum wage over as wide an area as possible. It is a part of their duty however to classify occupational groups within each industry in order to recommend differentials from the basic rate if conditions warrant them. When so advised the administrator may alter the terms of the act in specific cases, prescribing minimum wages between a lower limit of 30 cents per hour and an upper limit of 40 cents. In sum therefore the act contemplates a schedule of minimum wages bounded by these limits rather than a universal standard. To enforce the act the administrator is given broad powers of investigation and is supported by provisions which punish offenders by fine and imprisonment.

Appraisal of minimum wage laws. Minimum wage legislation presents an interesting problem in economic theory. As we have seen, the economic law of wages holds that the wage rate for any homogeneous group of workers will tend to settle under competitive conditions at the point of their marginal productivity. Advocates of the "living wage" and

similar standards for the payment of workers challenge the validity of this economic principle, pointing to the successful operation of minimum wage laws in many countries as proof that the wage rate can in fact be set in accordance with standards of justice or social welfare. On the other hand, others have sometimes appealed to the law of wages to prove the futility of minimum wage legislation. As a matter of fact neither of these extreme positions is valid. The principle of marginal productivity does not deny that wages can be set by statute; it merely implies that, if wages so determined are raised to an artificial level above the marginal productivity of the group, the result will be unemployment for some of the workers affected by the law. The government can require employers to pay wages at a certain rate if workers are hired at all; it cannot force the employer to hire any given number at this rate. If the legal wage is artificially high, the employer can and will equate the wage with marginal productivity by reducing the number employed.

The fact that minimum wage laws in this and other countries have raised wages without causing unemployment argues either or both of two things. First it may be that the competitive wage in operation before the passage of the law was lower than the economic wage or, in other words, that the bargaining weakness of the workers had been exploited to the end that they were paid less than they were worth. In such a case, as we have already seen, the actual wage rate can be raised by any device which strengthens bargaining power, and this device might well be put into the worker's hands through government enactment. Secondly workers of low productivity may have shown an increase in efficiency in response to better living conditions and the greater contentment which followed upon a rise in their standard of life. The initial rise of wages which produces this improvement in productive efficiency may come through legal enactment, but if the ultimate result is a rise in the marginal productivity of the group, it is evident that the higher wage rate may be-

come permanent, and that with no reduction in the numbers employed.

Whether the minimum wage will prove a real boon to labor or an ultimate detriment depends then upon the caution with which the policy is applied and the extent of improvement which is attempted through it. Examination of those laws of the American states and of England which have succeeded in raising wages without creating unemployment shows that these limitations have been carefully observed. The standard method of determining the legal minimum has been to take the wage policy of the more generous employers in a given occupational group as a standard. The statute in most cases merely makes this policy general throughout the industry, enforcing it against those employers who have been prone to use their bargaining power to exploit the workers.

Workmen's compensation laws: The problem. Though social loss from work accidents is inevitable, arrangements can be made which affect its distribution as between different individuals in society. Such arrangements do more than raise the question of justice. It has been found that different methods of apportioning the loss from work accidents have different effects upon the number of accidents and upon the magnitude of the social loss involved.

The burden must fall in the first instance upon the injured workman and his family. Where the rate of accident differs strikingly as between different occupations, this hazard will be one of the factors in the distribution of the labor supply. If the workers were entirely free to choose their lines of occupation, and if they made this choice on the basis of cool calculation of costs and income, the more hazardous forms of employment would be relatively undermanned until the wage rate became relatively higher than in occupations which were less hazardous. In this way compensation for the risk would be worked out by the forces of demand and supply, and the prices of the different kinds of goods would carry the cost on to the consumer. But we have seen that the distribution of

the labor supply is not controlled entirely by rational choice but largely by conditions of birth and tradition. Until society undertook to distribute the risk of loss by collective action a large part of the burden was borne finally by the injured workman.

Common law remedy. The workman has always been entitled to sue the employer for the recovery of his loss. In the courts the justice of his claim was tested by the principles of the common law defining the liability of employers. These common law principles were evolved at an earlier stage of industrial development to govern the simpler relations of master and servant, and their maladjustment to the conditions of the modern world placed serious obstacles in the way of recovery by the workmen. Stated briefly, the court attempted to discover who was to blame for the accident; if it could be shown that the fault was wholly the employer's, the workman was awarded damages; but if it appeared that the workman himself or someone else shared the blame with the employer, no matter in how slight a degree, the case was dismissed without compensation. This was the essence of the common law *doctrine of negligence*. Without undertaking here to discuss all aspects of the common law remedy, it may be stated as a fact that the right to sue gave the workman very little protection.

Principle of compensation. The public policy reflected in compensation legislation gives up all effort to assign the blame for accident, recognizing that in the modern system of industry destruction of life and limb are inevitable normal costs not generally attributable to a particular person's fault. The financial burden involved in compensation for accident and death is accordingly placed solely on the employer. It is expected that this cost will become part of the enterpriser's expenses and enter into the price of the product, thus finding its way to the consumer, upon whom in the last analysis it is natural that the burden should fall.

One of the greatest virtues of the policy of holding the em-

ployer financially responsible for accident is that it operates to reduce the number of accidents. By making the hazards of occupation expensive to those who control industry, the compensation laws create an incentive in the group which has the chief power to prevent accidents. The "safety first" movement and similar campaigns for the reduction of accident sprang up after society had determined to hold the employer responsible for the cost of injury received in the course of employment.

Modern compensation legislation began in 1884 in Germany, requiring employers to indemnify injured workmen or, in case of fatal accident, their families, and setting up an insurance system in which employers were obliged to insure this risk. In 1897 Great Britain adopted a comprehensive act, requiring indemnification at stated rates but leaving the question of insurance optional. France and Belgium have systems similar to Great Britain.

American laws. The first laws of the American states were declared unconstitutional, and it was not until 1909 that this policy received the approval of the courts. By 1938 all the states and territories, with the single exception of Mississippi, had passed laws on the subject, as had the federal government for its own employees and the wage earners in the District of Columbia. About a dozen of the laws, including that of the federal government, cover occupational disease as well as accident.

The essential feature of these laws is their requirement of compensation by the employer to the injured workman or, in cases of fatality, to his heirs. Aside from this common feature the laws vary greatly in detail. In only seventeen cases does the compensation act require the employer to join any insurance system. Rates of compensation vary but are always a percentage of the injured worker's wages, paid in the form of life pension in cases of permanent injury, as a capital sum or an annuity to the heirs in case of death, and as a single payment in case of temporary disability. Many of the laws pro-

vide remuneration for hospital and medical costs as well. The laws requiring insurance also vary; some provide a state insurance fund to which all employers must subscribe; others provide such a fund but leave the employer free to insure in private companies; and still others make no special provision for handling the insurance problem.

Security for the aged: Pensions. It has been found that nearly two-thirds of the population of the United States who are sixty-five years of age or older are dependent upon others for their support. A similar condition prevails in all industrial countries and has induced modern governments to adopt various measures to provide economic security for aged people. These policies are of two kinds: *pension systems*, the provision of regular monthly grants to the impoverished aged from the current revenues of the public treasury; and *insurance*, a system based on actuarial principles and providing annuities for the aged regardless of their need, covering the cost by collecting contributions during the productive lifetime of the beneficiary.

At the present time old age pension systems are in operation in about a dozen countries, including France, Great Britain, and the United States. In the latter country the policy is in the hands of the several states. Definite provision for old age pensions by state law began after 1920; by 1935 twenty-eight states had such laws in operation, paying pensions totalling upwards of thirty million dollars to nearly a quarter of a million people. The plans all contained about the same features: the eligible age was sixty-five years; American citizenship and five years' residence in the state were required; the amounts paid were usually limited to thirty dollars a month; payments were made only on proof of need. This type of provision for the aged poor, developing slowly throughout the country, was given an immense impetus by the federal social security act of 1935.

The social security act provides for old age insurance as well as for assistance to the pension plans of the states. The

latter feature is covered by a section of the act empowering Congress to appropriate $49,750,000 in 1935–36 and in following years as much as may be needed to make grants to states which set up old age pension laws approved by the Social Security Board. The approval feature has been used to strengthen the administrative machinery of the state laws and to provide complete coverage of all indigent aged people within the state who qualify as citizens and residents. When approval of the state law has been granted, the federal treasury undertakes to bear half the expense of the pension system up to a maximum of fifteen dollars a month per person. At the present time all but one of the states, as well as Alaska, Hawaii, and the District of Columbia, have approved pension schemes in operation. These schemes embody the features of the earlier state plans summarized in the preceding paragraph. In 1938 a million and a half people were receiving pensions totalling upwards to $270,000,000 annually.

Old age insurance. As distinct from old age pensions, old age insurance had been set up in many other countries before it was adopted in America. All told, thirty-two nations, including France, Germany, and Great Britain, were providing such insurance in 1938, either to all aged people compulsorily, as in Germany, or to those who voluntarily entered the scheme, as in Japan. In the United States until 1937 wage earners in private employment had available in old age only the pension plans described above, although compulsory contributory insurance had been established for public employees in many cities and states and in certain services of the federal government.

United States social security act. The old age insurance provisions of the social security act, in effect in 1937, provide a nationwide plan for the wage earners, with exception of certain categories of workers excluded by definition, under the direct administration of the federal government. In addition to governmental employees of all types, the following classes are excluded: agricultural labor, domestic servants, irregu-

larly employed casual workers, the employees of religious, charitable, and educational institutions, and marine and railway workers. All other hired workers, regardless of the amount of wage or salary — the general manager as well as the manual laborer — are embraced in the system.

The details of the plan are complex, since they provide for a transition period during which the contributions to the reserves gradually increase until the scheme goes into full operation in 1949; but the essential features are these: contributions are made in equal amounts by employer and worker in the form of a percentage tax on the payroll and the income. In 1949 the employer will be taxed three per cent of the wages of every insured employee, and the employee will be taxed an equal amount on his wage income; in the intervening years the taxes rise by graded amounts toward this percentage, beginning with one per cent in 1937. In practice the employer pays both halves of the contribution, deducting the worker's share from his pay envelope or salary check. The government makes no contribution toward the scheme. Funds raised in this way are carried in the old age reserve account of the United States Treasury to be invested in three per cent federal securities. Annuities are payable at the age of sixty-five, regardless of the financial circumstances of the individual, on condition that he ceases to draw wages at that time; as long as he continues in employment, the payment of the annuity is postponed. The amount of the annuity is to be a percentage of the total wages received by the beneficiary between 1937 and his attaining the age of sixty-five, with both minimum and maximum limits fixed so that no one receives less than ten dollars nor more than eighty-five dollars per month. Provision is made for a lump sum payment of three and a half per cent of the total wages in case of death prior to the age of sixty-five and for a return to the individual's estate of this amount less any benefits paid him when he dies after the age of sixty-five. Payments of benefits will begin in January, 1942, though of course the amounts will be small at that time be-

cause of the brief period during which reserves have accumulated. To give effect to this plan the government has created a vast administrative machinery, which registers the millions of workers involved and sets up a separate account recording for each the contributions received and payments made on his behalf.

The feature of this program most open to criticism is the so-called "reserve" against future payments. Since each beneficiary must pay premiums during his productive years without drawing benefits, the whole scheme involves the creation of enormous reserves in its initial period. It is calculated that the system will have attained maturity about the year 1980. At that time and for the indefinite future annual outpayments in benefits will be balanced by annual receipts from taxes on employers and workers plus the income from the accumulated reserves, the total reserve being estimated for that year at forty-nine billion dollars. A moment's thought will make it clear that the financial integrity of the plan (in the business sense) depends on its success in keeping this tremendous reserve invested so as to return the calculated rate of interest — three per cent.

It is noteworthy that the old age insurance reserve is to be invested in government securities; in other words, it will create a permanent national debt of the amount indicated. There are serious problems of public finance associated with such a forced growth in the government's debt; but for our purpose the important point is the misleading appearance of self-financing given to the scheme. The government will have only one source upon which to draw for the interest payments on the securities held in the old age reserve account; namely, the current tax receipts of the federal treasury. The outcome would not have been different essentially if the taxpayers had been called upon frankly to make up the difference between the income from the payroll taxes and the benefit payments guaranteed the beneficiaries. This effect of the plan will not become operative during its early years, when the contribu-

tions exceed the benefit payments and the reserves are accumulating, but only when income from the reserves must be drawn upon. It is estimated that eventually some one and one-half billion dollars annually will be required from the public treasury to balance the account.

These statements are not intended as a criticism of old age insurance as such but as an explanation of the real effect of certain administrative features of the plan adopted in this country. Old age insurance is too firmly established in modern countries and too strongly supported by economic and humanitarian considerations to be open to serious challenge. The principle of requiring contributions from the beneficiary is sound. To require the employers to pay a share is probably defensible as an administrative measure, but it is somewhat misleading as to the final incidence of the cost. The employer will tend to shift the tax, since it must operate as one of the regular costs of the business. This part of the contribution will eventually come to rest in large measure on the wage earners.

CHAPTER XXXVII

Labor Organization in America

Types of American unions. We shall use the term labor union to apply only to independent organizations of workers having no organic affiliation with any employer or enterprise. In recent years recognition of the desirability of collective dealing has prompted many large employers in America to set up organizations among their own employees through which to carry out negotiations regarding wages and other conditions of work. These *company unions*, as they are called, present certain problems, which will be considered later in this chapter. They are not included in our definition of labor unions proper.

The type of organization adopted by wage earners when acting independently of their employers is adjusted to circumstances. The attempt is made to adapt the union structure to that feature of the industrial organization which exerts determining effect upon the life conditions of the workers concerned. Since economic conditions are constantly changing, the structure of the labor movement is never exactly the same at different times or in different industries at the same time. It is possible however to distinguish two types of labor organization which contrast and even compete with each other in the general movement among American wage earners.

The trade union. The first of these is the *craft* or *trade union*. This type attempts to bring together all the workers who perform the same specific function and who are therefore likely to compete with each other. For example bricklayers, carpenters, plasterers, lathers, painters, though they all belong to the same major branch of industry, the building trades, are organized into separate unions in each locality. Similar

craft divisions are to be found among the workers on the railroads; the locomotive engineers, the firemen, the trainmen, and the conductors are united into self-contained groups distinct from each other. In fact the craft union is firmly established as a type of organization in all branches of industry where skilled craftsmanship prevails. The local units of each craft have united into national organizations in response to forces tending to broaden the range of the labor market.

To bring about coöperation between the different unions, reliance is placed on the device of *federation*. These federations are both regional and functional. As an example of regional federation we find in each industrial centre the local craft unions federated into a city council for the purpose of solving problems of common interest in the locality. Similar federations within the area of each state attempt to bring about coherent action on legislative programs which fall within the power of the state legislature. At the top of the structure the American Federation of Labor draws the whole system together into a national grouping, whose chief component units are the national craft unions. A functional federation takes the form of a loose affiliation of different crafts employed in the same general productive process. Thus the building trade workers, although allied primarily to their separate craft unions, are all drawn together into councils which attempt to promote inter-craft coöperation on common problems. Similar functional federations are found in the printing trades, the metal trades, and on the railroads.

For many years this structure of independent craft unions, federated regionally and functionally for common action, has dominated the American labor movement. In 1932, out of 155 unions in existence, 123 were organizations embracing a single craft or narrow combination of closely related crafts. The American Federation of Labor has long advocated this form of organization as opposed to any other type, and its strong hold on the labor movement of this country gave it authority to speak for the organized labor movement as a

whole. In recent years this unrivaled leadership has been challenged by the rapid growth under aggressive leadership of another form of unionism.

The industrial union. The second major type of union structure is the *industrial union.* This form attempts to obliterate craft and trade divisions among the workers and to make membership in the union coterminous with the limits of each separate industry. The factory system, with its machine technique and its refined division of labor, has tended to break down craftsmanship and undermine the position of the skilled workers. This process in the United States was during many years accompanied by a tide of immigration, which flooded the labor market with alien unskilled workers, thus tending to cause lower wages and presenting a threat to the higher standards of living of the native-born artisans. When skilled and unskilled, native and alien, male and female struggle together to obtain employment, older lines of division between occupational groups tend to disappear, and the whole group of wage earners in each major line of industry tends to develop a consciousness of common interest. Organization along craft lines is out of place in this setting; adjustment to the situation calls for a type of organization which brings the workers together in a single union regardless of their diverse functions.

In 1932 there were 32 national industrial unions in the United States, some of them organizations with much power. The United Mine Workers of America is an example. In its local branches the U.M.W. brings together all types of labor working in the mines in a given locality — engineers, carpenters, draymen, unskilled laborers, as well as the actual miners of coal. These local units are merged into a national industrial union which aims to cover the entire mining industry of the country. The clothing trades are organized similarly, as are several others. Some of the industrial organizations have joined the Federation of Labor, despite its opposition to this type of unionism, in order to coöperate

with the national craft unions in the solution of problems affecting the entire laboring population of the country.

Conflict of rival types. The C.I.O. The latent hostility between these contrasting types of union organization broke out into open conflict in 1935 when the Committee on Industrial Organization — a joint body of industrial union officials — launched a nationwide campaign to spread their type of unionism. The Committee was formed primarily to attempt the organization of the large mass-production industries, such as steel, rubber, and automobiles, which were almost wholly non-union. It fell afoul of the A. F. of L. almost at once, because of its outspoken condemnation of the Federation leaders for their apathy toward unionizing these industries and because the C.I.O. favored the industrial form of unionism, which brought it into conflict with certain craft unions, affiliated with the A. F. of L., which claimed jurisdiction over some of the occupations in these industries. Despite the active opposition of the A. F. of L., the C.I.O. pressed its organization drive vigorously in 1936 and 1937, with dramatic results. Its aggressive tactics precipitated occurrences in the business world — the "sit-down strike," mass picketing, violent conflict with the police and the company guards — which aroused widespread national interest. Early in 1937 the C.I.O. won the recognition of the United States Steel Corporation, traditionally the leading anti-union concern of the country, in the form of a collective agreement covering all divisions of the corporation's enterprise. By the end of 1937, although it had lost bitter strikes with several large independent steel companies, agreements had been formed with some 450 companies in the metal industries, embracing a half-million workers. In the automobile business the new industrial unions won recognition from all the large companies except Ford. Similarly successful were the efforts to organize the rubber and electrical industries. The outstanding accomplishment of this movement was to establish unionism for the first time among unskilled and semi-skilled workers in the large-scale

enterprises of the country. Beginning with an original membership of about one million, the C.I.O. now claims three and a half million members.

The conflict between the A. F. of L. and the C.I.O. has focused attention on the relative merits of the two types of organization and promoted the belief that they cannot exist harmoniously side by side. As a matter of fact the controversy does not turn entirely on questions of fundamental importance but is prompted in part by selfish considerations. The industrial unions, arising after craft unionism had become firmly established, have begun in certain industries to draw members away from the national craft unions, thus reducing their income from dues and the prestige of their officials. The embittering effect of these selfish considerations clouds the basic issue, whether craft or industrial unionism is the better form to promote the interests of the wage earners. Whatever the outcome of the present struggle between them, it may be expected that in the long run that type of union will survive which is best adjusted to the prevailing economic conditions of the country.

Strength of unionism in America. The following table shows the growth of American unionism since 1900.

UNION MEMBERSHIP IN THE UNITED STATES

Year	*A. F. of L. unions*	*All unions*
1900	625,200	868,500
1910	1,630,300	2,184,200
1920	4,156,000	5,110,000
1930	2,961,096	4,100,000
1937	3,268,600	7,296,500

The figures for 1937 show the disruption of the union movement through the recent growth of industrial unionism under the C.I.O. Of the four million members at present outside the A. F. of L. the C.I.O. claims three and a half millions; the remainder are in unions not affiliated with either federation. At the present time the union movement is growing rapidly

under the favoring influence of "New Deal" policy. It is however too early to say whether these recent recruits to the unions will continue as a permanent membership.

Company unions. The unionism we have discussed is a class movement in opposition to the managerial and ownership factions in society. It views the interests of these two classes as antagonistic and attempts to cut the industrial system horizontally through the boundaries of the separate business enterprises. Another type of organization has developed in recent years, called the "company union." This form organizes both managers and workers into a single unit within each separate enterprise, on the assumption that their interests are mutual and that both groups will benefit from any device which promotes harmony and efficiency. The structural details of these company unions vary among different enterprises, but they all have as their essential feature a system of councils jointly representative of labor and management. These councils discuss and exercise some measure of control over business policies which most closely affect the workers, such as the wage rate, methods of wage payment, discharge of workers, and the like.

The spread of company unions was stimulated after the World War by the aggressive policies of the independent unions, which caused many large employers who believed in the principle of collective dealing to organize their own employees within the limits of their separate enterprises. A clause of the national industrial recovery act of 1933 making collective dealing obligatory further stimulated this growth. The table on page 727 gives the figures for the years 1919–34.

A bitter struggle between the independent unions and the company unions is a striking feature of industrial relations in the United States at the present time. The independent unions believe that the rival form is a creation of hostile employers whose real purpose is to destroy the union movement and weaken the bargaining power of labor. The company union is always initiated by the employer, a fact which makes labor

COMPANY UNIONS IN THE UNITED STATES, 1919–34 [1]

Year	*Number of concerns with company unions*	*Number of wage earners enrolled in company unions*
1919	145	403,765
1922	385	690,000
1924	421	1,240,705
1926	432	1,369,078
1928	399	1,547,766
1932	313	1,263,194
1934	800	3,000,000

leaders suspicious of its purposes. Its effect is to tie the workers up with their respective enterprises, separating them from their fellows in other business concerns and other localities. We shall consider the relative merits of the two forms of organization in connection with our discussion of collective bargaining.

National labor relations act. With the adoption of the national labor relations act in 1935, the federal government altered the attitude of American national law from merely permitting to positively approving and fostering union organization. This law lists a number of "unfair practices" of employers which are prohibited, including in the list many of the devices formerly used to prevent or retard union growth within the enterprise or to establish employer domination over the union when formed. Refusal to negotiate with the accredited representatives of an organization among the concern's employees is forbidden as an "unfair practice." The law creates a permanent administrative agency, the National Labor Relations Board, made up ostensibly of members not partisan to either side of the industrial dispute, together with subordinate boards covering the principal industrial regions of the country. Disregarding certain powers of the NLRB not pertinent to our immediate topic, it is to be noted that the board has taken aggressive action in favor of unionism. It

[1] C. R. Daugherty, *Labor Problems in American Industry*, 1933, p. 781. The figures for 1934 are an estimate of this author. See *Labor Problems in American Industry*, Revised Edition, 1938, p. 643.

employs the "unfair practices" sections of the law for this purpose, having been given broad authority to investigate, hold hearings, and make decisions of a judicial character whenever labor leaders in a given industry complain of "obstructive" tactics by the employer. In cases where the right of union officers to represent the employees is in dispute, or where there is disputed jurisdiction between independent unions and a company union, the board is empowered to take a vote of the wage earners, who thereby decide whether they wish to be organized and if so in what form of union. The outcome of such an election settles the question whether a given group of labor leaders is the accredited representative of the employees. If this question is settled in favor of the union, refusal to recognize the union's officers as a bargaining agency becomes an "unfair practice."

The activities of the board were at first retarded by doubts as to the constitutionality of the act, a question which was not set at rest until April, 1937. Since this date the board has held numerous elections in different concerns, for the purpose of determining the desires of the employees with regard to unionism. It has been successful in obtaining union recognition in many of these cases.

The board has assisted independent unionism in its struggle with the company unions. Among the "unfair practices" prohibited by the act is any attempt on the part of the employer to dominate labor organization in his enterprise, even if the organization is a company union. Since most of the existing company unions were created by the employers, the board has ruled that many of them are incompetent under the law and has called elections in the enterprises in question to determine whether the workers wished to be represented by the independent unions. These activities of the National Labor Relations Board have hastened the decline of company unions since 1936.

The union program. Incidental benefits. Labor organizations perform many incidental services to their members

in addition to the functions which are their essential reason for existence. Most of them act as fraternal associations promoting social activities among those who belong to them. The stronger unions provide insurance and charitable relief for the victims of unemployment, sickness, or accident. These benefit systems in the aggregate reach multitudes of people and provide an important element of security in the lives of the wage-earning population. In 1928 it was found that 78 of the unions in the United States had such systems in operation. In 1930 the unions affiliated with the American Federation of Labor paid out some 37,000,000 dollars to their members and were carrying many hundred million dollars of insurance against the hazards of employment. Important as are these activities however, they do not constitute the distinctive economic function of labor organization. This consists in the attempt of the union to raise wages, shorten hours, and improve working conditions through the use of collective bargaining. There are some unions — the revolutionary wing of the movement — which look beyond these immediate material gains to an eventual overthrow of the capitalist system of industry, but in America these unions comprise but a small minority of all organized labor, and they too practice collective bargaining as their normal day by day function.

Collective bargaining. The essential feature of collective bargaining is the requirement that the employer hire an entire labor group as a unit, the terms of employment being settled by negotiation with representatives of the labor force. From the worker's point of view the merit of this method is derived from the bargaining weakness of the individual laborer in the modern system of industry, his ignorance of the prevailing wage or of his real worth to the employer, his timidity in asserting his claims, his urgent need of a job, which deters him from withholding his labor from the market until better terms are offered. Our study of the law of wages has made it clear that the actual wage in a given case need not be equal to the productive worth of the worker, and it can hardly be

expected to settle at this point if the two parties to the wage contract are very unequal in bargaining power. Low wages are not always a sign of exploitation, nor are high wages always proof that the worker is receiving his economic deserts. The test in each case is whether the rate actually measures the marginal productivity of the labor group in question. When the wage is below this point, whether it be high or low in amount, it can be raised by a change of bargaining methods which substitutes the collective for the individual contract. This principle also marks out the limits within which permanent improvement can be obtained merely by more aggressive bargaining; the wage can be advanced beyond the point of marginal productivity only at the cost of throwing some of the laborers out of work. The history of trade unionism records many cases in which collective bargaining has permanently increased the real wages of the organized workers.

Collective bargaining however is not of advantage to the laborers alone; it results in certain indirect benefits to the employer, which enhance its value in the industrial system. The mechanism through which this type of bargaining is carried out may be turned to account in drawing workers and employers together into closer contact on a basis of mutual tolerance and good feeling. Many of the wastes in modern industry and much of the friction and bitterness engendered in the field of industrial relations flow from the absence of such contact. Organization of the workers for collective bargaining and the establishment of a permanent structure for negotiation with the employers may effect a real improvement in the situation. The workers are supplied with a mechanism for voicing their grievances and are assured of a respectful hearing. Discipline in the shop is more effective when it bears on observance of the terms of a contract in the making of which the men have had equal voice. When subordinate officials are restrained from harsh and overbearing tactics, the improved treatment of the workers increases their contentment and their loyalty. Above all labor is given a recognized share

in the control of those features of industrial policy which most vitally affect their welfare, and this also tends to give them a more cordial attitude toward the enterprise. It goes far toward supplying a substitute for that personal interest in the product and the fortunes of the concern which existed in the handicraft stage of economic evolution but was weakened, with far-reaching results, by the growth of the factory system. These benefits from organized collective bargaining are now recognized by practically all modern employers whose enterprises have grown to any size.

The struggle over union recognition. But though the soundness of the principle of collective bargaining is generally conceded, there is no such agreement regarding the method of applying this principle. One distinctively American phase of the problem of industrial relations is the quarrel over the question of "recognizing the union." Organized labor insists that the workmen in each of the different enterprises must be represented in their bargaining by the regular officials of the union. These men are not on the payroll of any employer with whom they negotiate wage agreements but are professional bargainers representing all the members of their union within a given geographical area. Recognition of the union binds the employer to adopt this method of bargaining. On this point the issue is joined between the independent unions and employers adhering to the company union principle. These employers, though they are willing to deal with their workers collectively, insist that the workers' representatives must themselves be employees of the enterprise.

Strong arguments are presented in defense of each of these opposing systems of bargaining. Employers who refuse to recognize the union insist that the "outside" negotiator is ignorant of the peculiar circumstances of the business and therefore unable to discuss terms adjusted to specific needs of that business. Moreover these professional union officials are frequently so aggressive and domineering that peaceful compromise is impossible. On the other hand, leaders of organized

labor stress the weaknesses of the company union as a device for protecting the interests of the workers: (1) the individuals chosen to represent the workers lack courage because their jobs depend on the good-will of the employer; (2) these representatives, being ordinary wage earners continuously employed at their jobs, can acquire no knowledge of general trade conditions and no expertness in negotiation; (3) final control always rests with the management, which may veto the decision of the council and abolish the whole system at will; (4) this method of organization binds the wage earners to their individual enterprise, subjecting them to its hazards and depriving them of the collective support of other workers in case they are driven to strike in self-defense.

The controversy over the method of bargaining rarely discloses the real motives of either side. To organized labor, recognition concerns not only a process of bargaining but the very life of the union itself, and it is for this that labor leaders are fighting when they oppose the company union. On the other hand, the employer's resistance to the union's method of bargaining is really inspired by fear of certain practices to which organized labor may resort when sufficiently entrenched in the industry. In particular it is the closed shop with all its possible abuses that moves employers to refuse to recognize the union.

The closed shop. Strictly defined, a *closed shop* is one in which only union men will be employed at tasks whose labor supply has been organized. It is but an extension of the union program of collective bargaining, which is defective unless the contract is binding on the entire labor force. When union and non-union men are working side by side it is difficult to enforce the standards set in the collective agreement because of the employer's power to accord non-union men different treatment and to show them preference when expanding or decreasing the work force. Aside from this general benefit, the closed shop is of advantage to the union. It enhances the worker's inducement to join the union and reduces

the difficulty of expanding the field covered by organized labor. It strengthens the disciplinary powers of the union officials; the recalcitrant member expelled from the union suffers unemployment. This in turn facilitates the collection of union dues and makes for faithful observance of union rules. It is an absolute guaranty against the victimization of spokesmen for the workers when defending the interests of their constituents. It strengthens the hand of the union when, in the last resort, it orders a strike. In fact the closed shop makes the strike so formidable that the threat of the strike often suffices to win the point for which the union has been contending.

On the face of the matter the closed shop may be an advantage also to employers by giving the union officials power to enforce faithful observance of the collective agreement and to guarantee a standard of performance. On the other hand, there is obvious danger that these officials will use their power in less desirable ways: to oppose the efforts of management to improve labor efficiency, to encourage time-wasting practices of the workers, to prevent adoption of labor-saving machinery, and the like. Like other productive groups in society, labor is tempted to seek its own ends at the expense of others when in position to do so. The closed shop is a form of monopoly. Strongly entrenched unions enjoying this privilege may restrict the admission of other qualified workers to their organization, thus increasing the scarcity of their labor and raising their wages at the expense of excluded workers, and eventually to the damage of society as a whole. They may also restrict output while on the job and in this way reap gains from the higher prices. These are not merely imaginary dangers; many closed shop unions have put these and similar monopolistic devices into practice. This is not to say of course that the powers given union leaders by the closed shop are invariably abused and that this aspect of collective bargaining must necessarily be condemned. Final judgment must depend on the social philosophy and economic behavior of the union in question.

Political activity of American unions. In one regard the American labor movement has not followed the lines of development marked out by the history of the older movements of European countries. During the past three decades, in which period the unions in the United States have made their most substantial and continuous gains in strength, they have abstained from organizing an independent labor party. Most of the larger national craft unions, as well as the A. F. of L., have incorporated in their constitutions provisions which forbid their members from attempting to resolve the organization into a political party. With the recent growth of industrial unionism there has been some departure from this practice, the C.I.O. having entered the political arena in certain important industrial areas, sometimes with its own candidates for office. Nevertheless it may still be said that on the whole American unions conceive their problem to be primarily economic rather than political; their objective is to enhance their bargaining power as against the employers and not to capture the power of the state in the effort to put through far-reaching reforms. Accordingly their members are urged to retain their independence of choice of party membership, the official action of the unions in the way of influencing this choice consisting only in their endorsement of those candidates presented by the major parties who have shown themselves "friendly to the interests of labor." The unions have exerted political pressure chiefly by this non-partisan endorsement of candidates and by maintaining strong lobbies at the state and federal capitols.

This development is to be attributed to two major causes. Several of the early attempts to organize labor in America were destroyed by ill-advised campaigns in politics, and the unions have learned to fear the divisive influence of partisan politics upon their membership. Furthermore the American governmental structure is such as to discourage independent activity on the part of any economic group. The division of sovereignty between national and state governments, with the

important power over social legislation distributed among the forty-eight different commonwealths, the separation of the governmental machine into semi-independent executive, legislative, and judicial units, the power of a Supreme Court, remote from political influences, to pass upon the constitutionality of state and federal statutes, these and other characteristics of the system multiply the obstacles in the way of effective political action on the part of any minority economic group operating independently. Moreover the predominately *laissez faire* character of governmental policy has tended, at least until the advent of the "New Deal," to minimize the importance of legislation as a means toward the accomplishment of economic ends and has encouraged workers and employers alike to rely upon their own bargaining power.

A few of the strong unions have consistently asserted the necessity of carrying on independent political action along lines similar to those followed by the labor movement of England. It is noteworthy that the unions which sponsor the formation of an independent labor party in this country are almost wholly unions of the industrial type, whose broad inclusive basis of membership compels them to adopt policies of general social import. The attitude of the C.I.O. has been noted. The tendencies which further this form of unionism therefore will probably promote the political activity of organized labor.

Other tendencies may be observed working in the same direction. In all industrial countries governments are becoming an increasingly important force in economic affairs, regulating business activities, establishing standards of economic conduct, restricting individual initiative along many lines, entering the industrial world as owners of productive enterprises and as employers of labor. These expanding economic functions of government may attain such a degree of development that organized labor can no longer accomplish its aims by economic action unsupported by its organized political power, in which case we may expect the unions to organize for political action.

CHAPTER XXXVIII

Industrial Strife

The strike. The last resort of organized labor when the bargaining system breaks down is to settle the issue by open warfare. Strictly defined the *strike* is collective quitting of work by laborers who do not intend to surrender the job permanently. Spontaneous quitting of work has always been the practice of wage earners when other methods fail to win for them acceptable terms of employment. When however an organized group of workers strikes, it does so with the intent of retaining control of the vacant jobs until an agreement is reached. This is an important element in every strike situation, though it is not always understood by the general public. The result of this attitude of organized labor is violent opposition to any workers who attempt to take the jobs while the strike is in progress. Such people are viewed as "scabs" and "job stealers," against whom strong methods are considered legitimate. Correlative with the strike is the employer's weapon, the *lockout.*

All strikes involve loss to the workers; hence the weapon is not used thoughtlessly but is generally resorted to only after deliberation and in defense of what the workers conceive to be a vital element in their standard. There are exceptions to this rule, particularly in America, where racketeering has frequently contaminated the labor movement; corrupt union officials will call strikes in their own personal interests or to victimize an enemy among employers or rival labor leaders. Although workers when well organized and well led use the strike weapon with caution, still the menace of the strike has increased in direct ratio with the spread of unionism. Strikes occur much more frequently today than when the union movement was in its infancy. They take place on a wider

scale and are more protracted and more bitterly waged, and each strike is more of a national calamity, embracing a wider section of the industrial area and paralyzing the sources of supply for goods essential to public health and safety.

Statistics of strikes in the United States. The magnitude of the strike problem will be made clear by a glance at the table below. Strikes tend to increase in number and magnitude during prosperous years and to decline during depression. Periods of numerous strikes are also correlated with organization campaigns of the unions. These relations are shown by the figures given for individual years in the table. The year 1919 was one of flourishing business activity, marked by an organization drive upon certain large-scale non-union industries; 1932, near the trough of the depression, was a year of comparative industrial peace; 1937 shows the effects of the C.I.O. campaign during the recovery period immediately preceding.

TREND OF STRIKES IN THE UNITED STATES, 1919–1937[1]

Years	*Strikes*	*Strikers*	*Average days lost per year*
1919	3,630	4,160,000	140,000,000
1921	2,890	1,110,000	56,000,000
1932	840	320,000	5,300,000
1937	3,900	1,900,000	41,000,000

The interrelation of strikes with the attempts of unions to expand their membership is of particular importance in American labor history. When the attempt of the union to organize the workers in non-union industries is fought by the employers, a strike is the inevitable result. The great steel strike of 1919, the shop-crafts strike on the railways in the early twenties, the steel and motor industry strikes of 1936 and 1937 are examples. Certain unsavory aspects of the industrial conflict in this country — destruction of property, violence to persons — are also identified with these periods.

[1] Adapted from Daugherty, *op. cit.*, p. 287.

There is evidence to indicate that the increase of the strike evil along with union growth has been due to the fact that the attempt to introduce unionism has been generally resisted in this country; in other words, it does not necessarily reflect a preference for violent methods on the part of established unions. This inference is supported by the labor history of other countries where unionism is an accepted institution throughout the industrial world.

Types of strikes. Most strikes are what is called *primary strikes; i.e.*, called by a group of workers because of a disagreement with their own employer over their own terms of employment. Sometimes however workers strike who have no grievances of their own and who ask for themselves nothing in the way of better terms. These are *sympathetic strikes*, called by one group of workers for the sole purpose of aiding another group in a different business concern. Sympathetic strikes are usually limited to two types of situations. In the first case, the employer involved in the primary strike may be carrying on his industry by the use of strike breakers, and union men everywhere are anxious to prevent him from marketing his goods. In this case the sympathetic strike is called to coerce another employer, sometimes in a distant city, into withdrawing his patronage from the concern in question. In the second case, if the primary strike is a fight to preserve the union from destruction, organized labor of the locality in quite unrelated industries or the other unionized crafts in the same industry may strike in support of a principle considered vital to all unionized workers. A sympathetic strike may reach such dimensions within a localized area as to call out all unionized workers in that locality. There have been such *general strikes* in the United States, each of them limited to a single city and effective for a short time in bringing the entire economic life of the city to a standstill.

It is not difficult for labor leaders to construct a theoretical defense of the sympathetic strike or even of the localized general strike. The interdependence of industry has developed

to such a point that a strike in one business unit may be very seriously affected by what is done in other business units in distant localities. The loss of a strike called in defense of some vital principle of unionism may seriously weaken the position of unionized workers elsewhere. These considerations are especially persuasive when employers are united in nation-wide associations in opposition to the spread of unionism. However, while subscribing to these ideas in the abstract, the officers of the American national unions do not favor the use of the strike in these secondary forms, for the very practical reason that sympathetic strikes are rarely successful and general strikes never so. The public at large reacts to a general strike as to an insufferable attack upon its security and safety and turns out *en masse* to crush the revolt. Similarly the employer suffering from a sympathetic strike has little difficulty in convincing the public that the strikers, having no grievances against his labor policy, are acting in an indefensible manner. These strikes do occur in this country, but they are almost invariably called by local union officials against the advice of their national officers.

Violence in strikes. Every strike creates a situation in which violence to persons and property is very likely to develop, and many strikes have degenerated into a state of disorder closely bordering on civil war. It is noteworthy however that violence is normally associated either with an organization campaign in non-union industries or with strikes called to prevent the extinction of a union when an employer who has formerly dealt with it has refused to continue recognition. In such cases the union is fighting for its life and the situation calls to the front the aggressive, bellicose type of leader rather than the statesman. There is an uncompromising attitude on both sides which creates a spirit more akin to war than to business negotiation. The widespread sit-down strikes of 1936 and 1937, which were tantamount to forcible seizure of the company's property, were a novel development of the war tactics of newly formed unions in such a situa-

tion. Most of the earlier instances of bloody and destructive conflict in American labor history — such as the Homestead Strike in 1892 — developed out of similar situations.

When violence occurs in strikes called by established and recognized unions in the course of bargaining over the terms of employment, the cause is usually the presence on the scene of conflict of non-union strike breakers or the police force or militia. As pointed out in another place the free lance worker who accepts employment in an industry during a strike is viewed by the strikers as a "job stealer," against whom methods of coercion may fairly be used when milder persuasive devices fail. As for the presence of the police or militia, this usually results from an appeal that property rights and the right of the non-striker to work are entitled to protection. Since these appeals always come from the employers and property holders, and never from the strikers and wage earners, the appearance of the militia on the scene of action is taken by the strikers as evidence that force is to be used to aid the employer to win his case.

Legality of the strike. In the United States there is no single law governing the activities of the unions, since this is a matter lying within the province of the states. As regards strikes, state laws are so frequently changed and the opinions of state courts are so at variance with each other that it is impossible to phrase a general rule regarding the legality of the strike in this country. The only approach to a general principle is this: all strikes are legal, provided their object is clearly and solely to improve the economic condition of the strikers; *e.g.*, by raising their wages, shortening their hours, or bettering their conditions of work. Unless the strikers can prove that their whole intent is to benefit themselves, not to work injury upon others, this rule of law does not justify their conduct. Under the American law all combinations are criminal conspiracies when their purpose is to destroy property or to interfere with the rights of others. As a general thing the local primary strike usually survives the test of mo-

tive; but strikes on a wide scale, and especially the sympathetic and the general strike, are likely to be prohibited. When strikes are held to be criminal conspiracies, the courts will attempt to prevent them or bring them to an end by the use of the injunction. This legal device has broken many strikes, but since it leaves the causes of the controversy unchanged, it offers no permanent solution of the problem but only leaves the workers with an increased feeling of bitterness. Certain strike weapons are sometimes prohibited even in the case of legal strikes — for example, picketing and the "sit-down" — as working injury on innocent people not party to the dispute or openly violating the property rights of the employer.

Public intervention in strikes. The public is always the third party to every strike, if only in the rôle of victim. As industrial disputes have increased in number and scope, the public has been compelled in its own defense to seek ways and means to bring order into this vital area of its life. In the absence of alternative methods for the adjustment of the workers' grievances, it has been found impossible to stop strikes by the simple expedient of prohibiting them. Modern efforts to solve the problem therefore consist in setting up machinery for the adjustment of the dispute before it reaches the stage of an open break.

Conciliation. The least formal of the measures used to effect a speedy settlement of a trade dispute is *conciliation.* Strictly defined, conciliation denotes a method of settlement in which the employers and the workers, or representatives of these two groups, voluntarily meet together for discussion and agreement without aid or compulsion from the outside.

Conciliation implies a high degree of organization on both sides and a habit of collective negotiation. It is a natural outgrowth of the development of trade unionism, on the one hand, and employers' associations, on the other. The method is best used for interpreting the meaning of doubtful points in an existing trade agreement or of working out a peaceful compromise between conflicting demands of an employers'

association and a trade union, to be embodied in a future agreement. Such differences as these usually relate to specific, practical matters of detail, on which compromise is both possible and beneficial to both parties. In the absence of machinery which begins almost automatically to function for the settlement of these differences, the alternative is an open break and a trial of force which causes an immediate loss to all concerned and leaves an aftermath of bitterness and resentment. The effectiveness of voluntary conciliation is not fully appreciated by the public, because in the nature of the case this process if successful forestalls a disturbance which might have attracted public attention. As a matter of fact, conciliation has a long history of orderly and successful operation in many highly organized industries.

Questions of general principle, such as the right to organize, the prerogatives of union officials, etc., cannot be settled by conciliation. Agreements to bring such matters within the range of conciliation almost invariably cause a breakdown of the whole machinery of friendly negotiation. Furthermore experience has proved that it is necessary for the conciliation board to reach a unanimous decision on all matters referred to it. The board is made up of equal representatives of workers and employers; its proceedings are informal; its decisions cannot be enforced against a recalcitrant party. Unless decisions are unanimous, the constituents of those representatives who have voiced dissent are very likely to repudiate the verdict. A strike or lockout ensues, which brings the whole process of conciliation into disrepute.

Conciliation promoted by government. Public intervention occurs when the government requires industry to organize conciliation machinery in areas where collective bargaining, the trade agreement, and joint conference have not been established voluntarily. Some countries have set up the joint council system of conciliation throughout the industrial field as a whole; but the adoption of this plan has not been possible in the United States, because of the imperfect or-

ganization of the wage earners and the refusal of many large employers to deal with independent unions. Attempts of the federal government to encourage conciliation have frequently broken down in the face of the employer's refusal to accept the principle of collective bargaining through union officials. This was the fate of the conciliation machinery set up by the transportation act of 1920 and the Watson-Parker act of 1926. The national and regional labor boards set up in 1933 and 1934 as a phase of the NRA were designed as machinery for mediation and arbitration, but they were based on the expectation that a wide structure of conciliation councils would be called into existence by the legal requirement of collective bargaining. In this case too the substructure of conciliation was disrupted in certain industries by the refusal of employers to recognize the independent unions. Governmental attempts to promote conciliation in America will continue to cause more industrial disturbance than industrial peace as long as the company union and the independent labor organization are in conflict with each other. At the present time, as noted in the preceding chapter, the boards set up by the national labor relations act of 1935 and by subsequent state laws are promoting the spread of independent unionism at the expense of the company unions. This trend may lay broader foundations for conciliation.

Mediation. This is the first step in the direction of acknowledging the right of the public to a voice in the affairs of employer and employee. It involves a friendly and informal effort, by some individual or agency not party to the dispute, to bring the disputants together for the purpose of working out a peaceful settlement. The function of the mediator is diplomatic rather than judicial. That informal, non-coercive intervention of this sort is so frequently successful may be attributed to human nature. The mediator is usually a man of prominence in the social or political world, who comes forward at the critical moment to volunteer his services. In America mayors of cities, governors of states, the Secretary

of Labor, and even the President have performed this function with success. The eminence of such a mediator flatters the disputants with a sense of their own importance and inclines them to make concessions as a personal favor to him which they would refuse to make to an opponent.

Many countries have provided regular machinery for mediation in the form of standing boards of public officials. In the United States the Erdman act of 1898 provided that the Chairman of the Interstate Commerce Commission and the Commissioner of Labor should offer their services as mediators in disputes on the railways. Power to act as arbitrators was also given to these officials. Between 1898 and 1912, 48 applications for mediation and arbitration were made to this board, the majority of the cases being settled by mediation. When the Newlands act superseded the Erdman act in 1913, provision was made for a Commission of Mediation and Arbitration to be appointed by the President for a term of seven years. Of the 71 cases in which the services of this commission were employed between 1913 and 1917, most were settled by mediation. The law of 1913 which created the Department of Labor set up a mediation service under the Secretary of Labor, which has grown in size and scope of activity until it now employs 70 officials. Between 1913 and 1931 the service acted as mediator in some 10,000 disputes, with success in over 70 per cent of the cases. Mediation by members of the national and regional labor boards and by the Secretary of Labor as well as by officials of the regular mediation service proved again the usefulness of this device during the flare-up of industrial unrest in 1933 and 1934. Most of the states have added mediation to the list of duties of their industrial commissions or departments of labor.

At present the most active mediatory agency in the United States is the National Labor Relations Board, whose powers were described in the chapter immediately preceding. The board is concerned primarily with conflicts arising out of the

recent rapid spread of unionism into formerly non-union territory. It has permission, though not the obligation, to mediate when such conflicts either have caused a strike or are threatening one, and it has used this power freely. Between April, 1935, and July, 1937, it procured amicable settlement of upwards of 6,000 disputes, some 1,200 of which had reached the point of a strike. Failing mediation, the board has power to issue mandatory orders of the "cease and desist" type, requiring the abandonment of practices which are held to be an unjustifiable cause of strife. In only 250 cases has the board felt obliged to use these powers.

Voluntary arbitration. Arbitration involves the submission of a dispute to an impartial umpire, who will survey the merits of the case and render a verdict. Voluntary arbitration usually presupposes an existing trade agreement which makes provision for the employment of an arbitrator provided both parties agree. The award is binding on neither party, since it is non-enforcible at law. However when a trade agreement contains an arbitration clause, it is extremely difficult for either party to refuse later to submit a cause in dispute without loss of public support, and continued practice of arbitration gradually builds up a moral sanction which has great force in controlling conduct. The very fact that all previous steps have also been voluntary and that the arbitrator has been chosen by both sides because of his impartial character gives the award a sanction which is often more binding than a judgment based on force. Experience has shown that few unions or employers repudiate arbitration awards, provided the arbitrator has not overstepped the bounds of his mandate.

Almost all governments have set aside public officials or boards to act as arbitrators on request of the parties to a dispute, either as an incidental feature of their duties or as their sole function. Officials or boards empowered to offer their services as mediators may also in most cases serve as arbitrators. Virtually every one of the American states has made some such provision for voluntary arbitration. The Erdman

and Newlands acts provided that the designated mediation officials were to act as arbitrators of disputes on the railways upon consent of both parties if conciliation failed to compose the controversy, and the other federal laws providing for mediation have held their machinery available for voluntary arbitration.

On the whole these *ex officio* arbitrators have had little success either in this country or elsewhere. Both sides are inclined to distrust the impartiality and the capacity of men who owe their positions to political fortune. An improvement on the system is the appointment of a panel from which arbitrators may be chosen by the two parties, rather than the designation of a specific official. This is done in some twenty American states, whose laws provide for the appointment by the governor of representatives on the Board of Arbitration from nominees of employers and workers.

Effectiveness of voluntary arbitration. Aside from questions of personnel the method of voluntary arbitration is of limited effectiveness, both as regards the range of the industrial area in which it can be made to operate and the type of dispute to which it can be applied with hope of success. It is practicable only in highly organized industries where trade unionism and collective bargaining are well established. Even here it works only if the arbitrator has to deal with definite, concrete problems, such as an increase or decrease of wages or an alteration of hours. Controversies which raise really fundamental questions, such as the recognition of the union or the employer's right of discharge, are considered so vital to the interests of one side or the other that they will not be submitted to arbitration. Moreover it is not to be expected that an organized group will favor this method of settlement if the net result is to deprive it of the advantages of a strong bargaining position. In his handling of matter of fact problems therefore the arbitrator must generally discard principles of abstract justice and right and be content to adjust his award to the relative strength of the contestants, attempting to give

the stronger party the gains it would have made through an appeal to force. At best then arbitral awards are not deductions from economic law or from premises of public interest but a method of giving a potential victor the spoils of war without the expense of battle.

Transition stages. There are several transition stages between such voluntary arrangements as have been described and a fully developed system of coercive public intervention. One intermediate stage is represented by laws which make submission of disputes to arbitration compulsory, while leaving the acceptance of the award voluntary. Another stage is reached when attempts are made to give the award binding force after arbitration has occurred through mutual consent of the two parties. In an informal way this is sometimes done by requiring the parties to an arbitration agreement to deposit sums of money, which are forfeited in event of refusal to accept the decision of the arbitrator. The laws of about a dozen American states provide for punishment, by fine or imprisonment, of a recalcitrant party who has previously agreed to arbitration by public officials. These laws have proved impossible to enforce, and their presence on the statute books has acted as a deterrent to voluntary arbitration.

Compulsory investigation. A variant of this quasi-compulsory system is the prohibition of strikes and lockouts for a stated period while the authorities of the state investigate the matter and publish their findings. The best example of this system of *compulsory investigation* is to be found in Canada, where it has been in operation since 1906. The Canadian law applies only to public utilities, transportation agencies, and mines. It provides that any change in wages and hours must be preceded by thirty days' notice. If the change is unacceptable to one of the parties, no strike or lockout may occur until after investigation of the facts by a board appointed by the Minister of Labor and public announcement of a decision by the board. Fines may be imposed for illegal strikes or lockouts during the investigation period. After the

announcement of the award however both sides are at liberty to resort to a trial of strength. Experience has shown that, though the Canadian act has not stopped strikes entirely, it has been successful in 90 per cent of the cases to which it has been applied. This record of success has been overshadowed by the fact that certain of the illegal strikes have been of large scale, causing much loss to the public. The system has brought out one significant fact: that the coercive features of the law are much less effective than its voluntary features. In fact the attempts to impose legal penalties on recalcitrant workers have proved so futile that this provision of the law has become inoperative. During the first ten-year period only eleven persons were prosecuted, although upwards of 80,000 had broken the law, and only $1,660 were collected in fines as compared with $150,000,000 as the maximum legally due.

Although twenty-five of the American states have conferred powers of compulsory investigation upon their departments of labor, only Colorado has given this procedure extensive trial. In this state a system very similar to the Canadian law was set up by the industrial commissions act of 1915. This act was passed at the end of a disastrous strike in the mines. The scope of the law has been altered at various times, but since 1923 it has been limited to the public utilities, intrastate railways, and mines. It copies from the Canadian law the requirement of thirty days' notice of change in working conditions, the prohibition of strikes pending investigation, and the public announcement of a non-mandatory award by the Industrial Commission. The period since 1915 has been on the whole one of industrial peace in Colorado; but it is impossible to say to what extent this was due to the operation of the law or to a system of conciliation set up at the same time by employers and workers.

On the whole, organized labor opposes compulsory investigation, chiefly because the enforced delay weakens their power in case a strike eventually occurs by giving the employers time to make preparations for the strike.

Compulsory arbitration: The affirmative argument. In its complete form this device prohibits strikes and lockouts entirely and enforces the arbitral award by governmental power. Many are convinced that the progress of law must lead eventually to this coercive intervention of the state in the case of all trade disputes. It is argued that the next step in the protection of the public from disorder must be the complete disarming of capital and labor for private warfare. Founded on considerations of this character, compulsory arbitration has been adopted by the Australasian republics and for a time by one American state, Kansas. The dictatorship countries uniformly prohibit strikes and require settlement of disputes by official decree. These procedures however have no pertinence to a social order like our own.

Experience in New Zealand. New Zealand was the first country to adopt a system of compulsory arbitration. The first law was passed in 1894, after a period of bitter strikes. Since that date, except for the years 1932–36, there has been an unbroken history of compulsion in New Zealand. The original system had as its basis a series of regional conciliation commissions covering the entire island, each commission under the direction of a salaried expert. Attempt was made to settle all grievances by voluntary adjustment with the aid of these commissions. Failing this, an appeal from one of the parties took the case before a Court of Arbitration, which had final jurisdiction over all disputes. Until 1914 strikes and lockouts were not illegal unless application had been made for an award by the court, but on that date all strikes were prohibited in the case of unions registered under the act. Registration by the unions was voluntary, and those which did not register were not denied the right to strike. However the ease with which an entire labor group could be registered by a minority of its members gave the system of compulsion wide sway. A law of 1913 set up a scheme of compulsory investigation, similar to that of Canada, applicable to unregistered workers. This law prohibited strikes during the investigation period.

The Court of Arbitration consisted of a president, taken from among the justices of the Supreme Court, and two other members chosen from nominees of the registered unions and the registered employers' associations respectively. The court had judicial power to compel the attendance of witnesses and the submission of books and documents. Violation of awards was punishable by fine, and each award held good for three years unless superseded by a new judgment of the court, even though the union to which it applied had cancelled its registration in the meantime. Awards issued in cases affecting a single enterprise could be made applicable to the entire industry, thus equalizing competitive conditions.

The employers in New Zealand finally united against the law. Successful in a political campaign in 1932, they amended the arbitration procedure to remove most of its compulsory features, providing that, except for unions composed chiefly of women, submission to arbitration should require consent of both sides of the controversy. In other words, New Zealand returned at that time to a system of conciliation with voluntary arbitration as a last resort. But the change was of short duration, for the election of 1936 brought a labor government into power, which reinstated the compulsory features of the law and widened its scope.

In Australia both the Federal Commonwealth and the constituent states have systems of compulsory arbitration similar in essential respects to that of New Zealand.

Compulsory arbitration in Kansas. The Kansas law of 1920 is the only example, outside of Australasia, of thoroughgoing compulsion exerted by a democratic government. This law prohibited strikes and lockouts in specified industries and created a Court of Industrial Relations to settle all disputes. Early in 1925 however the court was shorn of most of its power by a decision of the Supreme Court of the United States, which declared unconstitutional the compulsory arbitration features of the law as applied to working time and by implication overthrew the whole process of compulsory

arbitration whenever its effect was "to compel owner and employees to continue the business on terms not of their own making." This decision indicates the apparent unconstitutionality of all measures seeking to apply the principle of compulsion to privately owned enterprises in this country, except in cases of emergency in industries clearly affected with public interest, as long as the federal constitution contains a clause guaranteeing "liberty of contract."

Judgment of compulsory arbitration. Any inferences regarding the practical operation of compulsory arbitration must be drawn from the experience of Australasia, where this method of promoting industrial peace has had thorough trial. For a time the world was inclined to accept the evidence as proof of success. New Zealand was known until 1905 as a "land without strikes." But the subsequent history of the system has shown that strikes cannot be prevented by compulsion. Since 1906 New Zealand has experienced frequent and bitterly contested struggles in the field of industrial relations. Australia has had similar experience.

The early success of these laws was due chiefly to the contentment of the workers with a system which operated continuously in their favor. The Courts of Arbitration favored the principle of unionism, even to the extent of compelling employers to accept the closed shop. Early awards resulted in continuous advance of wages and shortening of hours, thus conceding to labor larger material gains than could have been obtained from the process of bargaining alone. There were no strikes, because the causes of strikes were absent. The underlying reason for this result favorable to labor was the rapid and sustained material progress of Australasia during this early period of its industrial development, which made possible a continued increase of the wage income without encroaching on the necessary rewards of the other factors of production. When the rate of progress was retarded and it became impossible to grant the demands of the workers, rebellion against the system arose immediately. An epidemic

of strikes broke out, which were the more bitter because they were levied in part against a government which was thought to be unfair to labor. When the workers used their political power to influence the process of arbitration, the result was a series of decisions which laid unsupportable costs on the business world. In New Zealand owners of industry were then driven in their turn to take political action against the system, with the result, as we have seen, that it was deprived of its mandatory features. The political victory of labor which restored compulsory arbitration in 1936 added new emphasis to the fact that the state exercising such powers is destined to become the centre of a political struggle between capital and labor.

No type of public intervention has as yet proved as effective in promoting industrial peace as have voluntary arrangements set up within industry by mutual consent of employer and worker. Conciliation, mediation, and voluntary arbitration are accomplishing much in the way of smoothing out trade disputes. Machinery for public intervention in the last resort is most useful when it is carefully designed to promote these informal devices for negotiation and settlement. Attempts to apply compulsion instantly arouse the suspicion and antagonism of labor. If they are persisted in, despite this antagonism, their ultimate effect is to force organized labor into politics, with a program of far-reaching social change which may mean the end of the present system of industry. But these voluntary arrangements, as we have seen, presuppose organization on both sides, and in the United States at the present moment there is no agreement within industry as to the mechanism through which organization can be brought about.

CHAPTER XXXIX

Socialism and Communism

Radical movements. Many people who are in revolt against the present structure of society have no faith in moderate programs of reform but insist that the fundamental institutions of our social order must be completely changed if its defects are to be removed. These people do not agree among themselves as to the type of society that should supplant the present one or the tactics to be followed in bringing about the revolution. They have organized in separate and often bitterly antagonistic movements, each stressing some distinctive feature of its program which places it in contrast with the others. These complex and many-sided movements cannot be accurately defined by a single formula or described in the small space at our disposal. It will suffice for our purposes to group them all under the general title, *socialism*, reserving for special consideration one large section of the revolutionary movement, *communism*, which, though originating in the same body of doctrine, is now operating independently of the other socialist groups. The Fascist program will also be considered in this connection, because of its close association with revolutionary socialism, although it is commonly — and erroneously — classified among the conservative movements of our time.

Definition of socialism. All socialist doctrines contain four groups of ideas: an indictment, an analysis, a program of reform, a campaign. The *indictment* is a list of charges against "any and all industrial systems based on private property and competition." The *analysis* traces the development of capitalism back to its origin, dissects its present structure, and stresses certain developmental tendencies which, so it is argued, must inevitably destroy our present order. The *program* outlines a

substitute for the capitalistic system which socialists believe will remove the evils of modern capitalism. The *campaign* is a plan of immediate tactics for those who adhere to its cause. There is greater unanimity of opinion regarding the first two branches of doctrine than the last two. In their indictment of capitalism socialists differ only in the breadth of their condemnation of the present order and the relative stress they place on its different faults. Since the socialist analysis of our institutions is derived almost entirely from Karl Marx's classic interpretation of capitalism, there is considerable agreement here.

Socialist programs differ widely in detail, but we can discern an essential similarity among them. They all propose a type of society in which there will be no private property in wealth used for productive purposes, and they agree further that the productive system of the future shall be governed, not by competition and private initiative, but by some sort of collective control and guidance. Socialism then is destructive of these two basic institutions of our order, private property and free competition in enterprise. The programs differ in their substitutes for these institutions.

In regard to their campaign socialists divide into two groups: those relying on peaceful, evolutionary methods, and those who place their reliance on violence and force. The former are *constitutionalists*, believing that gradual increase in their power on the political field will eventually bring them to their goal by thoroughly legal means. The latter are *direct actionists;* they have no faith in politics, being convinced that only through force or threat of force will the revolution come.

Any definition which attempted to embrace all these phases of the socialist movement would be too cumbersome for practical use. We shall define the movement in terms of its program, stressing the essential similarities between the different socialist schools. *Socialism is a program of reform which repudiates private ownership of the means of production and competitive control of industry; it proposes a type of society in which productive capital will*

be owned collectively and economic activity will be controlled by authority.

Socialist indictment of capitalism. The details of the socialist indictment may be summarized with reference to the institutions of competition and private property. Competition is condemned as a method of supplying the economic needs of society, because it is wasteful of resources and effort, because it is unstable and erratic, because it results in strife within and between nations. The institution of private property is condemned because of the inequality and injustice it produces and because of its seriously adverse effects upon the processes of production. The two institutions are jointly condemned because they destroy human liberty.

The socialist charges against the competitive system on the score of its wastefulness are not altogether new to us. The point is stressed that individual enterprise results in duplication of effort. Competitive advertising, competitive salesmanship, and other forms of competitive activity increase the costs of each enterprise. These costs could be reduced if the services common to all producers were provided through a common agency. In the competitive system the superior business methods and most efficient machines are controlled by individual competitors for private profit. Increased efficiency for society as a whole would result if the best processes were made the common possession of the entire industry. Thus far the argument is much the same as that which contends for the superior efficiency of combination and large-scale production. In a more general sense the socialists urge that the present system is wasteful because enterprisers in their search for profit direct capital and labor into fields of activity which are already well enough equipped to serve the needs of society. It would be more economical to use collective intelligence in forecasting our needs and conscious coöperation in adjusting the means of production to the ends in view. It is urged also that the uncontrolled pursuit of private gain exhausts irreplaceable natural resources and destroys human

productive energy, as in the case of children put to work prematurely and old people thrown into idleness while still able to perform work.

In charging the economic order with instability the socialists develop an explanation of the business cycle which makes competition its basic cause. Competition produces anarchy in the productive system and, when coupled with inequality of income, results in overproduction. Periodically the system collapses through its inability to market its product, and there ensues an incalculable loss from idle capital and labor. These cyclical catastrophes, it is asserted, could be avoided if competition were supplanted by a planned economy, controlled and guided collectively. In charging that the competitive order encourages strife it is pointed out that employers and workers strive with each other over division of the income from industry. Producer strives with producer for advantage in the market. Out of this strife spring practices which result in loss of productive forces, such as were disclosed by our study of industrial relations and governmental control of competitive practices. The anomaly of the situation is that these contesting parties are really coöperating toward common ends. Intentional coöperation is less wasteful than antagonistic coöperation. From a broader point of view, socialists believe that the competitive system is the root cause of international war, the striving of nation with nation over markets. They believe that their system will automatically produce harmony between individuals, social classes, and nations.

The charges against private property are in part implied in the foregoing attack on competition. But this institution has defects of its own. As the socialists view the matter, the great inequalities in incomes are caused by private ownership of income-bearing wealth and perpetuated by the institution of inheritance. The socialists believe that capacity is distributed with approximate equality among the different income classes. Opportunities for self-development cannot be

equalized so long as some men are born to a position of privilege while the many are born in an environment of poverty, ignorance, and hopelessness. This situation, they say, is not only unjust on ethical grounds but is inefficient and precarious; inefficient, because it smothers the latent abilities of those in the lower groups and subsidizes idleness in the upper classes; precarious, because it arouses the envy and resentment of the victims of the injustice, who are an overwhelming majority and have the power to destroy the system.

Inequality is particularly vicious because it distorts the whole productive system out of alignment with the real needs of society. Only those things are produced which will sell at a profit. Hence the kinds and quantities of goods resulting from the combined activity of mankind are determined by the spending power of individuals and not by the test of what is needed to nourish a strong, healthy population. Many things of vital importance to human welfare are not produced at all because no individual enterpriser can make a profit out of their production. The socialist system proposes a quite different basis of production. The social group will decide what things are needed to support a wholesome life, and these things will be produced first in such abundance that there is plenty for all. Only surplus labor and capital will be devoted to the production of luxuries and extravagances, while injurious goods will not be produced at all.

As a final and in their opinion most devastating charge, the socialists condemn capitalism as destructive of human liberty. The individual liberty of which so much is made by the defenders of our order is dismissed as a myth. To the socialists the present order means the enslavement of the great majority of propertyless people by the owners of wealth and the subjection of the whole body of society to the dictates of those who dominate its essential industries. The wage system is condemned as a species of slavery. It is explained as an arrangement between the few who own the instruments of labor and the resources of the earth and the many whose only hope

of survival consists in their getting access to these things. The outcome of the arrangement is that the many deliver themselves into the power of the few, their productive energies to be exploited to the profit of the masters.

The economic destiny of the social group as a whole is, according to this indictment, at the mercy of the more ruthless or more fortunate individuals. The landlord exacts his tribute for the privilege of using the gifts of nature. Those who have acquired or inherited ownership of socially necessary instruments, such as railroads, steel mills, and factories, can work their will upon the rest of mankind. These essential instruments are social in their purpose and importance; they cannot safely be allowed to fall into the hands of individuals. Finally all social classes — wage earners, employers, capitalists, landlords, and others — are seen to be slaves to the passions of self-interest and acquisitiveness which the competitive system arouses. The socialists believe that the order of society which they propose will be one of real freedom. Men will not work for the profit of other men but for the group as a whole and under its commands. The instruments of labor and the resources of nature will be owned and managed collectively under democratic government. The passions of greed and self-seeking will be held in check by arrangements which play upon better qualities in human nature.

Socialist analysis of capitalism: Karl Marx. The socialist analysis of the development and present tendencies of the capitalist system is derived in the main from the work of Karl Marx, who was the founder of almost all the modern schools of socialism. The *Communist Manifesto*, written in collaboration with Frederick Engels in 1848, and *Capital*, an analysis of the economic laws of capitalism and a forecast of its future development which appeared in 1867, have profoundly influenced all subsequent radical movements. For our purposes the essential doctrines of Karl Marx may be presented under two heads: the class struggle and the laws of capitalist development.

The class struggle. The doctrine of the class struggle is a phase of the materialistic conception of history which is the keynote of Marxian historical philosophy. This interpretation of the historical process explains all institutional development in terms of the underlying economic institutions. Government, religion, family system, art, literature, and morals are by-products of man's economic activities, harmonizing with the prevailing system of production and exchange. The changes which have appeared in these institutions in the course of societal evolution have been the effects of prior changes in the economic basis of society. This doctrine, in some degree of comprehensiveness, is now believed to be true by most students.

Marx carried the doctrine one step further. To him not only were the economic institutions basic and determining, but these institutions had their own vital centre in the class struggle. Every system of production which mankind has hitherto devised has created a hierarchy of classes, exploiters and exploited. In classical society slave, plebeian, patrician, and noble, in the Middle Ages serf, vassal, knight, and feudal lord, were ranged in an ascending order of increasing privilege. Not only is class antagonism the essential characteristic of any historical social order at any given time, but this antagonism has produced the dynamic power which has caused society to evolve and change its structure. For the interests of the classes have clashed and forced them into continuous conflict. Revolutions in economic institutions, themselves the cause of all great changes in other institutions, have resulted from the class conflict when a previously exploited class gained power to throw off the yoke of the oppressor. Thus "the history of all hitherto existing society is the history of class struggle." But each such revolution merely reëstablished the system of exploitation in a new form; for the victorious class in turn organized a system of production with itself in the seat of power and the rest of mankind subservient to it.

The last great change of this type occurred when the bour-

geoisie arose and overthrew their feudal masters, thus instituting the system of capitalism. Out of this grew the present class alignment, capitalists and laborers, and the present struggle between these two classes. The outcome of this struggle lies in the future, but it is predicted that the wage earners will eventually succeed, as did the present capitalists, in overthrowing their exploiters. Their triumph will mark the end of the class struggle, since they will reorganize the industrial system so as to remove the possibility of exploitation. This will be true because the wage earners, being a vast majority of the whole population and committed to coöperative action, will organize the economic structure on collectivist lines, in which each will work for the welfare of the whole group. Thus the socialist revolution will usher in a classless and therefore stable society.

Laws of capitalist development. As a result of his study of the capitalist society of his day Marx thought he had discovered the existence of two motive forces actively tending toward socialism and a third characteristic of capitalism facilitating the eventual transition to socialism. The active forces were: (1) the tendency toward concentration and (2) the increasing number and misery of the laboring class; the facilitating factor in capitalism was (3) the recurrent collapse of the productive system in times of crisis.

Tendency toward concentration. As to the first of these forces, it is argued that capitalist development tends in two directions toward increasing concentration: first the concentration of industry into larger and larger business units, secondly the concentration of wealth and income in the hands of a diminishing number of people. Industrial concentration is said to be due to the superior efficiency of large industries, which enables them to undersell and crush out smaller competitors. We must note to Marx's credit that this prediction was made before the trend toward large-scale production had set in. But Marx set no limit to the process, leaving his readers to infer that concentration would continue until a single

gigantic business unit monopolized each of the major fields of industry, including agriculture. In agriculture the prediction has been falsified by subsequent events. How far it has been verified in other fields we have seen in another place. This tendency was to lead to socialism for the double reason that collective control would become both necessary and easy; necessary in order that society might protect itself from the menace of these monopolies; easy to establish because of the disappearance of effective competition, and easy to administer because of the relative simplicity of operating the few centralized productive units.

Concentration of wealth was inferred, partly from the preceding tendency, partly from other evidence. Marx made the mistake of assuming that concentration in industry of necessity involved concentration of ownership. He wrote before the day of the corporation, whose joint stock principle permits growth in the scale of the business unit coincident with an increase in the number of those who hold property rights in it. Great as are existing inequalities in distribution, we have no evidence that they have grown greater since Marx's day or are now growing greater.

Increasing number and misery of the laboring class. The tendency toward concentration, coupled with other features of the capitalist system, was expected to produce an expanding propertyless class, with a steadily declining economic status. The crushing of small enterprises by the growing monopolies would precipitate former small capitalists into the laboring class. The growth of an industrial reserve army, the effects of periodic depression, and the absorption of an increasing fraction of income by the rich would cause increasing misery for this ever growing group. The rich would grow richer and fewer, the poor more numerous and poorer. This process would cause the system to break down through sheer inability of the masters "to maintain their slaves in their slavery." It would also cause the victory of the proletariat to put an end to the class struggle, since it would be a victory of "the

immense majority in the interests of the immense majority."

The rôle of crises. Marx made much of the tendency of industrial society to fall periodically into crises and depressions. In his opinion crises were the inevitable outcome of the exploitation inherent in the capitalist system. The wage system resulted in withholding from industrial laborers a substantial fraction of the values attributable to their productive efforts. This would be momentarily profitable for the employer, but for the system as a whole it would mean that goods were continuously produced in excess of the ability of the great mass of the people to take them off the market. This situation would run along for a time until the accumulated stores of unsalable goods precipitated a crash, followed by a depression, during which production would stop until the surplus of products had been liquidated. Marx believed that such paroxysms of the industrial system would not only recur periodically but would grow progressively more severe.

This theory of crises held a place in Marx's prediction of the inevitability of socialism. In the first place, each crisis and depression would play its part in increasing the number and the misery of the propertyless classes, thus promoting the trend described in the preceding paragraphs. Moreover as each collapse became more complete and recovery more delayed and difficult, the time would come when the recuperative powers of the system would prove inadequate to the task of restoring equilibrium. The crisis itself was a challenge to the validity of the capitalist system, calling attention to faults which permitted an excess of the products of man's labor to exist coincidentally with widespread misery and want among the laborers. It therefore predisposed ever increasing numbers of people to turn against the system and seek for a substitute. There would come at last a collapse so cataclysmic that revolt would occur before the protagonists of capitalism could bring about recovery. Both Marx and his immediate followers believed that the transition to socialism would come about during a business crisis.

General appraisal of Marxian doctrine. Although there are serious errors in Marx's analysis of capitalism and many of its specific predictions must be dismissed as unproved, there remains much of value to the student of social evolution. Marx's basic interpretation of history has provided a convenient and illuminating approach to an understanding of the historical process. His analysis and forecast of capitalist development, written as it was in the early stages of industrial society, displayed rare powers of insight.

But the Marxian doctrine is important chiefly for its influence in shaping the development and the present character of socialist thought. It placed socialism on a new and firmer basis. There is real validity in the central thesis; *i.e.*, that every thorough evolution in society's institutions must come about as a natural unfolding of forces at work in the preceding order. Until the rise of communism this principle had great influence on the practical tactics of all the followers of Marx. It convinced them that capitalism must run its course before socialism is possible. The factory system in industry, with its corollary of a large wage-earning class; democracy in the political field; education for the mass of mankind; experience in collective action gained through labor organizations — these phases of capitalism must be nourished to maturity. Violent premature revolutions would fail disastrously. Marxian socialists were forced by the logic of their system of thought into peaceful and legal forms of activity. Many of them believed, as did Marx himself, that the final overthrow of capitalism could not occur without violence and disorder; but their preliminary program was evolutionary rather than violent.

This outlook continues to dominate the doctrines and tactics of all the principal socialist groups except the communists. These revolutionaries, although they claim Marx as their spiritual leader and profess to derive their doctrines from him, are convinced that socialism can be established even in backward countries which have had no experience of capitalism.

They also deride the peaceful political tactics of other groups and place reliance on direct action. So complete is this break in doctrine and strategy that it would be misleading to attach the term, socialism, to both wings of the movement. Accordingly we shall deal with them separately in the discussion of program and campaign which follows.

The socialist program. We have seen that all socialist groups agree in that they propose to abolish private ownership in the means of production and to substitute for competition some form of authoritarian control over productive processes, but they differ as regards the form of collective ownership and control. The principal line of division is between (1) those who would make the present territorial state the owner of productive instruments and would vest control of economic activity in the chosen representatives of all the people and (2) those who would place each branch of industry under the government of the workers in the industry, relying upon various devices of federation to attain harmony within the economic system as a whole. The syndicalists and guild socialists are examples of the latter branch of the socialist movement. Their ideas have attained some prominence in a few countries; but on the whole they are far less influential than are the representatives of the other wing of the movement, who take the name state socialists from the pivotal feature of their reform program.

State socialism does not demand complete centralization of society's economic life in a single unit of government. Its program provides for division of ownership between different political units. The central government would own the industries whose significance is nationwide; to the smaller territorial divisions and the municipalities would belong those which supplied only the local markets. Some allowance is made for individual enterprises, provided they refrain from employing hired labor and sell their products under strict price control. In this system, when fully established, wages would be the only possible type of income, and everyone

who wished to escape starvation would be compelled to work, excepting of course the aged, the children, and the defectives, who would be cared for by the state. The state would be the dominant power in the lives of all; it would be the only employer, the owner of all land, the only saver of capital, the only seller of goods. It is intended that this great power shall be brought completely under democratic control through universal suffrage, direct election for short terms of all policy-making officials, proportional representation, single chamber legislatures, and other devices.

State socialism includes those who accept the central principle of political ownership and control but modify in some details its more extreme doctrines. In fact the majority of state socialists are not doctrinaire but opportunistic in outlook. They press for an extension of government ownership and control wherever private ownership appears to them to threaten popular welfare and competition seems ineffective or destructive, without much concern as to the final outcome of the process. They are willing to leave as large a sphere for individual initiative as is consistent with the general principle that land and the larger aggregates of capital be used for the "common good" rather than for private profit. The Fabian Society of England is best representative of this school of modified socialism.

Socialist tactics. In most democratic countries socialism as a practical political movement has taken on this moderate, temporizing character. Political parties are formed locally and nationally to agitate for an extension of government ownership and regulation of industry. The immediate objectives are usually limited: municipal ownership of the principal services; nationalization of railroads, mines, and public utilities; legislation to limit the power of employers and strengthen organized labor, and the like. The peaceful methods employed and the relatively minor changes proposed at any one time often conceal the full implications of the movement. But this gradual encroachment on the domain of capitalism is

expected eventually to produce an economic order so extensively socialized that the establishment of a thoroughgoing socialist state will be easy and natural, and it is in terms of these final results that the whole movement must be judged.

The development of socialism in the separate countries is too complex a subject to be given adequate treatment here. Everywhere the movement is divided within itself over questions of doctrine and tactics into factions, which in most countries are organized for political purposes in opposition to one another; this quite apart from the separate organization of the communists, who are bitter antagonists of the socialist movement as a whole. The divisions within the socialist movement proper usually reflect conflicts over the relative emphasis to be placed on organized labor, on the one hand, and the territorial representative state, on the other, as the sovereign body in the socialist economy, or differences of opinion as to the rapidity with which modern industrial society should be socialized. Without going into detail it can be said that the Social Democratic Party, to be found in many countries, the Labour Party of England, and the Socialist Party of the United States, usually representing the majority wing of the movement, stand for reliance on political government and a cautious experimental rate of change; while the syndicalists and the radical socialists favor revolutionary action through labor unions and a hasty continuous drive toward socialization.

Political action. The dictatorship countries, whether communist or Fascist, have crushed out a once strong socialist movement within their borders. In other industrial countries, excepting the United States, socialism is a powerful factor in the political and social life. The socialists have controlled the governments of many municipalities throughout Europe and in Great Britain. They always form a strong opposition party in the national assemblies and at times have controlled the central governments in European democratic states. Typical of conservative socialism is the Labour Party of Great Britain, formed early in this century as the political arm of British

organized labor. From small beginnings this party grew in twenty years to become the strongest single political group in the country. For a short time in 1924 and again in 1928, though still a minority party both in terms of popular vote and representation in Parliament, the party was called upon to form the ministry and conduct the affairs of the nation.

In the United States socialism has never been strong as a political movement. The individualism of the people, fostered by the relative abundance of economic opportunity in a new and naturally rich country, the retarded development of organized labor, and the peculiar constitutional structure of the national government, which places almost insuperable obstacles in the path of minority third parties, are among the factors accounting for this situation. In local elections the socialists have at times gained control of municipal governments, notably in Berkeley, Bridgeport, Butte, Milwaukee, Reading, and Schenectady. They have sent representatives to many state legislatures and have elected members to the House of Representatives. But in presidential elections, in which they have engaged since 1904, their record popular vote, attained in 1920, was less than a million, and it has declined since then. Nevertheless the influence of socialist ideas upon governmental policy in this country has been greater than is indicated by the political record of the party. Many reforms — such as social legislation, legal immunities for trade unions, governmental regulation of business, and direct government activity in economic enterprise — originally sponsored by socialists, have come to be adopted by the major parties and put into effect in the several states and in the nation. There is in fact little difference to be found today between certain policies of the more progressive of the major parties and the immediate program of the Marxian socialists advocated by them as steps leading gradually to complete social reconstruction.

Communism: Doctrine and program. Before outlining the program of communism, it would be well to warn the

reader that this must not be assumed to coincide with the policies put into effect up to the present time by the soviet government in Russia. Internal and international conditions have caused that government to compromise with the original principles of communism. But quite apart from the influence of political expediency, the communist doctrine itself provides for an intermediate period between the seizure of the state and the final establishment of pure communism, during which the country will be carried through a stage of socialism not unlike that which the state socialists propose as their ultimate form of social organization. During this period, it is held, the ground will be prepared for a second transformation into communism.

One distinguishing feature of communist doctrine in contrast with socialism is its conception of revolution. The communists reject the Marxian dogma that socialism can be established only in countries with mature capitalistic institutions, believing that a determined and disciplined proletariat can capture power by force in any country and hold it by terror long enough to revolutionize the social system. Moreover in all countries they are scornful of the peaceful tactics of the orthodox socialists, believing that dispossession of the capitalist classes can never be brought about by these means, while to adopt them merely enervates the movement and commits it to a policy of flabby compromise. The communists base their strategy squarely on the class war, a succession of armed clashes with the forces of capitalism, culminating in a violent revolution on the Russian model to be consolidated by a reign of terror. In harmony with this outlook they repudiate democracy, both in the organization of their party and in that of the revolutionary state, and propose a dictatorship of self-appointed leaders commanding a loyal well-disciplined proletarian army.

The essential features of the ideal communist society must be inferred from the theoretical literature of the movement. They may be summarized as follows. There is to be collective ownership of all forms of industrial enterprise, administered

by the government; labor alone is to give right to income, and all labor is to be performed for the government as the only employer; the entire social income is to accrue, in the first instance, to the government, to be distributed according to communist principles. The method of distribution differs from that of other socialist schemes. Wages are to be the same for all workers regardless of function and are to be paid in labor tickets which are redeemable in goods out of the general fund. Money is to disappear, and the only form of exchange will be that of "service for counter-service"; that is, labor service of the individual for the state and counter-service by the state in the form of consumption goods supplied to the individual. Many important types of goods — such as transportation, housing, light, fuel, perhaps food — are to be free for all citizens. Equality is to be the rule, and any departure from the rule is to be justified only on the ground of the greater need of the favored individuals. "From each according to his ability" is the formula governing the social service of the individual; and "to each according to his need," the formula determining his reward. The transition stage of the revolution is counted on to give the dictators opportunity, not only to bring the country's social institutions into conformity with this ultimate purpose, but to remould the behavior of the people and give them a motivation appropriate to a communistic society. When these ends have been achieved, dictatorship is to give way to a completely democratic system of government. Indeed the state itself, as a coercive organization, is to wither away and disappear as the people become motivated toward voluntary coöperation. The term "government," as used by the communists, connotes merely the executive organization of such a freely coöperating group.

Communist organization in the Soviet Union. The Communist Party of Russia provides the prototype of the movement's organization in all countries. Its essential features are: complete centralization of authority in the hands of the party's leaders; provision for fairly liberal debate of policy

and tactics up to the time when decision is taken as to the "party line"; ruthless suppression of all differences and the requirement of undeviating loyalty to the leaders when once decision has been reached. In other words, there is avowed dictatorship within the party as within the political state, which the party controls. Membership in the party is not a thing to be undertaken lightly. To join requires a confession of faith and devotion, a willingness to submit to arduous testing during a period of probation, and a readiness after becoming a full member to bear heavy financial and labor costs on behalf of the party's program. Members are taxed heavily in proportion to their earnings and are expected to give all their spare time to party works under the orders of their officers. Periodically the membership is "purged," through a comprehensive inquisition, which eliminates those of faltering faith and lukewarm works. Membership, in other words, is like that of a strict religious order, requiring zeal, devotion, and sacrifice. So it is in the parties of other countries. For this reason the Communist Party of Russia has never been numerically large, never including more than a trifling percentage of the people. But what is lost in numbers is more than made up by the discipline and drive of the movement. This quasi-religious aspect of communism, with the irrational and uncompromising attitudes which it evokes, is a salient feature of the movement in all countries.

The structure of the party is based on a nucleus or "cell," a local unit formed in a small geographical area, an industry, or other organization by three or more members. These local units are grouped into convenient regional organizations; these in turn form larger territorial groupings, which are brought together into the national party government at the top. Disregarding complexities of detail, it may be said that the structure is erected by electing delegates from the lower units to the higher, culminating in the election by the largest territorial subdivisions of delegates to the All-Russian Congress, reputedly the supreme authority of the party.

But the actual lines of authority are concealed by this ostensibly democratic procedure. The Congress is an unwieldy body of large membership, meeting for short periods at long intervals. It elects a Central Committee to meet more frequently on matters of policy, a Central Control Commission to serve as a supreme disciplinary authority over the party membership, and a Central Revision Committee to audit the party finances. The Central Control Commission, also a large body, delegates its authority to an elected executive board known as the Presidium. The Presidium, the Central Committee, and the Revision Commission join to elect the general secretary of the party and the Political Bureau, a small body in continuous session composed of the key men in the various central organizations. Through this complex hierarchy of interlocking organs supreme authority is gathered into the hands of the secretary and the Political Bureau. The secretary controls the paid party workers in all the local units and, using them as a personal following, becomes the leader of the entire party. It was by means of this position that Stalin first rose to dominance, and he subsequently achieved his present position of absolute personal dictatorship by using this machine to crush any of his former chiefs and colleagues among the older Bolsheviks who challenged his authority or differed with his policies. The Political Bureau is now composed of Stalin's closest associates. It is in reality the sovereign body in party government under the leadership of the Secretary, since it exercises all the powers of the Central Committee, exerts party discipline through the Control Commission, prepares and steers the program of the Congress, and controls the entire press and other publicity organs of the country.

Socialism in the Soviet Union: Recent history. In dealing with the Russian performance as an example of communist economics we must disregard the experiences of the years immediately following the revolution, for this period was too greatly disturbed by civil war and foreign invasion and by the profound economic collapse of the country to permit the com-

munist leaders to concentrate on their program of social reconstruction. For the past decade however, and particularly since Stalin solidified his power, the dictators have held complete command of the situation, both within their own party and throughout the country at large, and the economic policies they have adopted during this time may fairly be taken as representative of communism in its early stages. A turning point of policy was reached when Stalin and his colleagues decided that their experiment could not wait upon successful class revolution in other countries but must proceed to establish socialism in the Soviet Union in isolation from the capitalist structure of other modern nations. To some extent the Russian program has been colored by the dictators' fear of attack from without. This fear in part accounts for the stress upon national self-sufficiency, the stimulation of patriotism in disharmony with the theoretical internationalism of the communist creed, and the large investment of capital and labor in military preparedness, which has retarded an elevation of the standard of living of the Russian people.

Planned economy. The essence of the soviet program inheres in a comprehensive economic plan, which embraces every phase of the national life and attempts to carry the entire social order toward pre-determined objectives. Certain phases of the plan are non-economic in character, dealing with such problems as the elimination of illiteracy, the promotion of education, hygiene, and public health, cultural activities, and the like. These are among the most interesting aspects of the plan, both in conception and achievement, but we must disregard them here for lack of space.

The planning mechanism is the State Planning Commission, created by the central government, together with its subordinate bodies for regional divisions and separate economic enterprises. The commission is composed of experts in the different departments of the national economy, dominated by party appointees who hold the program true to the "party line." Its function is to prepare detailed schedules for every

line of industrial and agrarian activity, covering the kinds and quantities of goods to be produced, together with the projected costs and prices, the rates of capital increase in various industries, the increase in the number and technical efficiency of laborers, and the like — the whole plan having relation to a term of years immediately following its adoption. During the operation of the plan, the Economic Council of the central government exercises supervision over it, checking current rates of production against the scheduled forecasts, and altering the schedules from time to time in accordance with changing political and economic conditions; the whole weight of governmental authority is behind the council's decisions. Planned economy in Soviet Russia does not mean rigid, hard and fast adherence to the details of a program, but a flexible policy which, though outlined in advance and moving steadily toward pre-determined general objectives, allows for considerable readjustment along the way. Planned economy really dates from the adoption of the first "Five Year Plan" in 1928. A second "Five Year Plan" followed in 1933, and on January 1, 1938, the third "Five Year Plan" was launched. The Soviet Union has thus had a decade of experience with economic planning.

The principal features of these "plans" may be summarized as follows. First a determination to socialize the entire industry of the country under government ownership and operation, extinguishing private enterprise in all branches of industrial production and distribution. Second to establish socialism in an agrarian economy dominated at the time by millions of independent small-scale farms. Third to expand on a gigantic scale the industrial capital equipment of the country, to change its technique from small-scale to mass production and from handicraft to power-driven machinery, and greatly to increase its rate of output and lower its costs. Similarly to increase the output of foodstuffs and raw products of the sort required to make the country industrially self-sufficient. These were the major objectives. As will be seen,

they involved a fundamental revolution in economic organization as well as an increase in productivity. The intention to raise the standard of living in the country by increasing the quantity and variety of consumable goods may be mentioned as a fourth objective. But the drive toward military preparedness, with its emphasis on the machine-making and war equipment industries, caused this phase of the plans to be set aside more or less completely.

Economic organization. Almost every department of the nation's economic life has been successfully transformed into a system of state socialism. The government owns and operates the railroads and other public utilities, the mines and forests, and all the factories; it monopolizes trade, both domestic and foreign; the banking and credit mechanism is in its hands. Agriculture is to some extent an exception. Under the soviet constitution legal title to all lands vests in the state, and in accordance with this principle there are vast farms owned and managed by the government. But the major form of agrarian organization is a transition stage between private enterprise and thoroughgoing socialism — a system of producers' cooperation, in which numerous individual peasants merge their land holdings into a single productive unit, which they manage collectively under governmental supervision. This system of collectivization has almost completely supplanted the former system of independent peasant holdings; in 1938 there were 243,000 collective farms, embracing 93 per cent of all the peasants and 99 per cent of the cultivated area. There is no private employment of wage labor even on the few remaining individual farms; the socialized industries provide the only source of wages. Small household industries are allowed to exist, provided they conform with the law prohibiting the hiring of labor and sell their products at prices fixed by the government. These vestiges of private industry are recognized as anomalous, their existence being justified by the government's inability thus far to provide an adequate supply of consumption goods through its own industries. The present

rate at which economic transformation is progressing indicates that reconstruction on socialist lines is virtually complete.

Growth of soviet industry. It is impossible to trace in detail the rapid development of industry under the soviet régime. At the outset Russia had only a rudimentary industrial equipment, particularly in the heavy metal and machine-making trades. Since 1928 she has vastly increased her capital equipment in these lines. Between 1927 and 1938 total industrial production increased in the ratio of 9 to 73, an eight-fold increase.[1] Many new industries have been established, notably in lines in which Russia was formerly dependent on foreign supplies; for example, tractors and other agricultural machinery, chemicals, and airplanes. From small beginnings the Soviet Union has advanced her production of iron and steel to a level now comparable with that of Germany. Electrification has proceeded at a rapid pace. The production of industrial raw materials has increased nearly four-fold during the ten-year period. In all these lines, basic to a highly industrialized society, the development of Russia under the régime of planned economy has been little short of phenomenal.

The record is not one of complete success. As stated above, the program for the consumption industries has lagged. Goods desperately needed for the everyday wants of the people remain very scarce and of poor quality. And beyond this there have been numerous shortcomings and failures in the industrial program, dramatized from time to time by the summary execution of the managers held responsible. Working with a labor supply wholly unacquainted with machine technique and with managers and technical advisers of little experience, the forced growth of industry has produced some awkward results. The new machine equipment has deteriorated rapidly through misuse and lack of repair; production costs measured

[1] The figures quoted in this section are taken from a special report of the Foreign Policy Association, *Foreign Policy Report*, June 1, 1938.

in labor units have remained high; some of the output of the machine-making industries has been defective to the point of uselessness, and much of it has been of poor quality. The soviet authorities admit these shortcomings and attribute them to an overambitious program. The third "Five Year Plan," begun in 1938, calls for a slower rate of advance in most lines, and in some even for a decline of production below the figures actually attained the year before. This new policy of "deceleration" is made necessary, according to the soviet authorities, by the "incompetence, disorganization, and waste" which have developed under the severe pressure of the planned economy.

Labor and wages. The communist program is conceived in the interest of the wage earner, who is to be set free from the slavery which is his lot in capitalist society, given an independence which he has never formerly enjoyed, and accorded a larger share in the social income. In these as in other respects the experiment in Soviet Russia has failed thus far to conform to communist principles. Labor policy has had an interesting development. At an early stage substantially equal wages were paid to all workers regardless of function, the higher types of expert and managerial labor being placed on about the same level of reward as the manual workers. This scheme broke down completely and had to be abandoned, due to its failure to stimulate the industry and efficiency of the abler workers. As the drive for rapid increase of productivity became more severe, the last vestiges of freedom and equality among wage earners were swept away. At the present time the wage system differs very little from that which obtains in capitalistic industry, except that the average real wage is very much lower and the extremes of inequality not so great. Piece-rate wages, payment in proportion to product, and higher rewards for services which are scarce are all characteristics of the soviet wage system. In like manner other socialist theories of labor management have had to be discarded. At first the powers of management were severely

limited, and the workers were given virtual control over the operations of their respective industries. This system of industrial relations resulted in such hopeless confusion and inefficiency that the soviet authorities were obliged to reëstablish managerial autocracy in industry. In fact the control and discipline of labor in Russia is now much more complete and dictatorial than in even the most conservative of capitalistic nations. It is true that labor is organized comprehensively into trade unions, but these organizations are entirely subservient to the state and, as advocates of labor's welfare on the job, are much less powerful than in capitalist countries. Strikes are outlawed as treason against the state. Even unintentional inefficiency in union management is severely punished, as in 1937, when the entire higher officialdom of the Central Trade Council was dismissed from office and four of the most prominent labor leaders were arrested. The unions no longer pretend to exercise any bargaining power in matters of wages, hours of labor, rates of output, or discipline. They are still allowed to function in adjusting minor local grievances, but on the whole they are used by the government to strengthen its drive for greater productivity and as agencies for administering the social insurance benefits.

The material comfort of the Russian wage earners under the soviet régime is not disclosed by the schedules of money wages. These wage rates are extremely low as judged by any standards known to modern industrial nations; even so they exaggerate the material welfare of the people conceived in terms of real income. So great has been the shortage of the necessities of life that until recently these things were virtually unobtainable at the government shops, with the exception of food, which was distributed by a rationing system. The food ration was abolished in 1935, and since that time the people have had less difficulty in getting enough to eat. But the prices of most consumers' goods remain so high that the wage income for the average family is sufficient only to cover the bare necessities of life.

In appraising the material welfare of the workers in the soviet régime, account must be taken of certain other benefits accorded them by the government. They are given security against unemployment, provided of course that they are willing to accept the terms offered by the only employer, the state. A comprehensive system of social legislation indemnifies them in case of accident or sickness. Maternity and child care, free education with stipends for students in the higher branches, and a certain amount of free medical service are also provided. For 1938 the government has budgeted a total expenditure of over a billion dollars for these free benefits to the workers.

Soviet finance. In the final analysis the costs of the planned economy take the form of sacrifices by the people of freedom and material comfort. This does not refer to the large-scale destruction of life, the exile or imprisonment of great numbers of people, the forced labor of political convicts, all incidents of a ruthless dictatorship. What is meant is that in its routine economic operation the program has fed on the toll taken from the standard of life of the common citizen. During this period the Soviet Union has been virtually unable to increase her resources by borrowing abroad. Yet she has imported vast quantities of capital equipment and raw materials and has drawn other large quantities from her internal resources. This has all come from the consumable incomes of the people, thus increasing the prevailing poverty. The government's power to determine the quantities of goods and services allowed to enter into direct consumption and the quantities to be employed in capital creation has forced the people to save, either against their will or without their knowledge.

But the day by day operation of the program has had to be financed, and the development of financing methods has produced some interesting results. The soviet authorities soon had to abandon the communist doctrine that money is unnecessary in a socialist economy where the state monopolizes production and fixes wages and prices. The money system

which has been set up in Russia is ostensibly on a gold basis, with a standard unit — the ruble — which represents a fixed weight of gold. But this standard monetary unit is entirely hypothetical and does not exist in reality, being used merely as a unit of account for measuring values in government records and recording transactions in foreign trade. The money in circulation is irredeemable paper, whose value is determined solely by its purchasing power within the country. This value is not governed by the principles discussed in our earlier study of monetary theory, for the reason that prices of goods have been fixed by law and at times a rationing system has determined the distribution of goods among the consumers. In such a system there is no necessary connection between the quantity of money and the general price level. A given money income may fall into two parts: a portion which can be exchanged for a quantity of goods determined by authority; a remaining portion which has little value since there is scarcely anything to be bought with it.

The development of the program has produced a situation in governmental financing which is illuminating as a prototype of a socialist economy. By 1937 the government had gathered into its own hands 95 per cent of the national income as compared with 27 per cent in 1928. This of course refers to "income paid out"; a considerable amount of real income still accrues directly to the people, particularly the peasants, in the form of products from small land holdings and the like. This large share of the national money income accrues in the first instance to the state by virtue of its monopoly of all enterprises producing for the market. Public finance under these conditions is chiefly an accounting of the receipts and expenditures of productive enterprise. The government's budget records some income from taxes levied directly on the citizens, but in the main its receipts consist of portions of the income of the government monopolies reserved for purposes of the state. These are recorded in the budget as taxes, a "turnover tax" on sales, a "profit tax" on railways, factories, farms, etc.; but

in essence all these exactions on the "profit" of business enterprise are subtractions from the incomes of the people. The "profit" is created arbitrarily by the power of the government to fix both the money costs and the sale prices of industry. What soviet public finance means in simple language is that the dictators fix the money wages which determine the people's income sufficiently below the prices of goods to leave a large surplus for the use of the state.

CHAPTER XL

Fascism. Judgment of Radical Reform

Fascism and communism. It is a common opinion that Fascism as a social and political philosophy stands at the opposite pole from communism, representing conservative capitalist principles. The opinion has probably developed from the fact that in the countries where Fascism has triumphed — such as Italy and Germany — the movement arose as a belligerent opponent of radicalism, while in other countries, where Fascism is a minority movement, it is supported by ultra-conservative groups. In reality however Fascism as a political philosophy bears a close likeness to communism. They are both revolutionary movements opposed to democracy; both adopt the same technique of revolution — direct action by a firmly organized party under dictatorial leadership, followed by terroristic methods. The Fascist conception of government by dictatorship and even its proposed organization of the political system appear to be identical with the ideas and practices of communism in these matters. Furthermore both subscribe to the principle of a planned and controlled economy, and both might well use their powers of governmental dictatorship in pursuit of the same objectives.

There are however many sharply defined contrasts between Fascism and communism, which become clear when one examines the two systems in actual operation. First there is a difference in the concept of the state, despite the fact that in both régimes the state is equally authoritarian. Fascism worships the state as a spiritual entity, superior to the sum of the people who compose it and eternal in its validity, while communism would eventually do away with the state entirely. As a corollary, Fascism is in theory militantly nationalistic, while communism is in theory international, recognizing no political

frontiers. The fact that Russia has turned to nationalism is explained by the soviet leaders as a temporary reaction to the present state of international affairs and is admitted to be contrary to their creed. There is no such disharmony between the rabid nationalism of Fascist countries and the basic doctrines of their leaders. Second there is a distinction between the underlying theories of government in the two movements. A Fascist state is to be dominated permanently by a small party, to which only the élite can belong, while communism, whose governing party is similarly organized in the early stages of the state, avows the intention of broadening its base until it includes everyone. Third the two movements differ in their conception of social classification. Communism not only believes in the class struggle as a historical fact but fosters it as a means of social evolution. Fascism, on the other hand, denies the historical necessity of the class struggle, explaining it as the artificial creation of demagoguery, and insisting that both proletariat and bourgeoisie receive their status from membership in the state and are bound by the same obligations of loyalty and service. Finally one finds a number of minor differences which it is difficult to reduce to formula. For example communism recognizes the essential equality of human beings, regardless of sex or race, granting to all the same legal rights and admitting them all to the same vocations, in the party, in the government, and in industry. Fascism, on the other hand, either excludes women from political and social activity or subordinates them to men, and it professes belief in extreme theories of racial superiority. There is a difference too between the attitudes of the two movements toward religion. Every loyal communist must be a professing atheist. Although the soviet government permits freedom of worship, it forbids church organization beyond the individual parish, thus weakening all organized religion, and it looks benevolently on the efforts of the party to stamp out religious belief by propaganda. Fascism requires of its adherents no personal opposition to religion, and it interferes with the

churches only when they, as organizations, are hostile to its purposes.

These differences of detail should not however lead us to overlook the essential similarity of the two movements. The ease with which the German Reich and Soviet Russia have recently entered into political alliance tends to support the conclusion that there is no fundamental conflict of theory or practice between Fascism and communism. Even if at a later date these nations should find themselves at war with each other, such occurrence would indicate a conflict between the imperialistic ambitions of two dictators rather than a struggle between two antagonistic social philosophies.

Fascism in Italy: Party organization. By official decrees of its Grand Council, adopted in 1926 and amended in 1929, the Fascist Party is defined not as a political organization but as a civil militia serving the state. At its base is the local unit, the facio, under its own party secretary. The fasci are federated regionally, each such grouping being headed by the provincial secretary, and the provincial organizations are drawn together at the top in the National Directorate and the Grand Council, acting under the supreme command of Mussolini. The party officers are arranged in a hierarchy of subordination under the leadership of the Secretary-General, who is also secretary of the Grand Council. This post Mussolini does not hold; his title within the party is Il Duce, commander-in-chief.

Membership in the party carries with it professions of faith and obligations of service similar in nature to those exacted of its members by the Communist Party. Candidates for membership are given a searching examination, obliged to pass through a probationary period, and when admitted required to take an oath to obey "without discussion" the orders of Il Duce. Periodically the behavior of the members is examined by a disciplinary court, headed by the Secretary-General, and shortcomings in faith or works are punished to the point of expulsion from the party. Recently it has been decreed that

recruiting of party membership shall be limited to the graduates of the Fascist youth organizations, which school the children in the creed of the movement and inure them to its discipline.

The Fascist Party and the official organs of government are so interwoven that in reality the party is the governing mechanism of the country. The Grand Council, governing body of the party, has been recognized by law as the coordinating agency of all departments of the state. Mussolini's dual position as president of the Grand Council and head of the government — officially Prime Minister appointed by the King — has also been recognized by law.

The dominance of the party over the government can be made clear by tracing the process by which official policy is determined. Legislation originates with the head of the government; it goes for consideration to the Fascist Grand Council, is then examined by the ministry, and only after prior approval by these bodies is it presented to the Chamber of Deputies for adoption. The Chamber is the lower house of the legislature. The upper house — or Senate — which has little legislative function, is made up of members appointed for life by the King, a majority being Fascists. It is permissible too for the executive to issue decrees without consulting the Deputies, the decrees having the force of law for two years unless made permanent by subsequent approval of the Deputies.

The electoral process is under similarly rigid party control. The process originates in a list of one thousand candidates for the Chamber of Deputies presented to the Grand Council by the local party organs. From this list the council selects four hundred party members in good standing, whom it submits to the country for approval, the people voting on the list as a whole and not on the merits of the individual candidates.[1]

[1] A decree of the Grand Council on October 8, 1938, indicated that the electoral character of the Chamber of Deputies may be brought to an end at some future date. The new Chamber, as described in this decree, will consist of all

Thus party and government are fused together through the person of Il Duce, legal head of both, through the constitutional powers of the Grand Council, supreme policy-forming body in both, and by means of party control of the membership and activities of the national legislature.

Fascist social organization. The "corporative state." Fascism has never developed a body of economic doctrine comparable in extent to that of communism. Its theories are stated principally in terms of the purposes or objectives which guide the dictatorship in its utilization of supreme power. The central objective has been to unify the social and economic life of the country so that all the nation's forces can be mobilized in the service of the state. There is no limit set to the extent of governmental interference with the economic activity of the citizens. Anything necessary to the political purpose of the moment may be done in the way of promoting, prohibiting, or regulating business enterprise, fixing prices, wages, and business costs, altering the details of industrial organization, or bringing the government directly into business activity as owner and manager of enterprises. In general however Fascism opposes outright governmental ownership of the means of production. It supports the principle of private property and relies on private initiative to function within the limits set by the state.

The structural organization of society prescribed by Fascist theory is known as the "corporative" system. Although developed with apparent precision in early Fascist writings, this system was not established legally until 1934, and it is even now incomplete. It has as its basis syndicates or unions of workers and employers respectively in each trade or calling, the two interests being obliged by law to organize separately. The state gives official recognition to only one such workers'

members of the Grand Council of Fascism and of the National Council of Corporations, who will hold membership *ex officiis*. The name will be changed to "Chamber of Fasces and Corporations." A description of the National Council of Corporations is given on page 787.

and one employers' syndicate for each territorial unit — commune, province, or region — and the law provides that a syndicate, whose decisions will be binding on all workers and employers in the industry, can be formed by ten per cent of the individuals concerned. The local or communal syndicates of the two factions in each industry are grouped into larger regional organizations, depending on the geographical spread of the industry, and these in turn are formed into national federations. It is understood that the principle of separate but parallel organizations for workers and employers is observed throughout the entire structure These parallel unions are then drawn together at the top into nine national confederations, eight of which comprise representatives of the unions of employers and workers in the chief subdivisions of industrial, agricultural, and commercial activity of the country, while the ninth represents the professional and artistic activities.

Here is an apparently symmetrical structure of capital and labor throughout the country. The principal function of the syndicates is to form collective agreements covering terms of employment, in both the local and the regional divisions of the specific industry, and nationally through the federations. Such agreements are binding on all employers and workers in the industry, whether they have joined the recognized syndicate or not. Conciliation machinery operates to compose disagreements over the terms or the interpretation of a collective agreement; failing this, the case is submitted to a special Labor Court for final settlement, strikes and lockouts being prohibited. Since the confederations represent all factions in the industries, they are supposed to take a statesman-like attitude toward industry as a whole, mapping out schedules of production, improvements in technique, and the like. In this they serve as a link between industry and government. They are dominated by political authority in that both the president and the council of each confederation are appointed by the government. In theory the syndicates themselves are

voluntary organizations, but in reality all the syndical officials are designated by the Fascist Party.

On top of this industrial structure was erected in 1934 an all-inclusive system of "corporations," whose relation in function and authority to the syndical organizations is by no means clear. A corporate organization of Italian national life was envisaged in the original Fascist theory, and a National Council of Corporations had existed since 1930, but the constituent bodies of this Council were not created until four years later. In that year Mussolini announced the formation of twenty-two corporations, embracing every conceivable department of productive activity, each composed of representatives of employers, workers, and technicians in its respective field. Mussolini is president of each of these corporations, and members of the Fascist Party make up the rest of their officers. The total membership of all the corporations, some 825 in all, form a Central Corporative Committee, which according to Mussolini's prediction will eventually displace the Chamber of Deputies and become the legislative organ of the state. Each corporation has a vertical structure, embracing all stages of production in its specific field from the supplying of raw materials to the completion of the finished product. At present the functions of the corporation are defined as follows: to act as advisory bodies to the government, to settle disputes between capital and labor, and to regulate wages and other costs of production. But their future as forecast by Mussolini involves nothing less than absorption of all the powers of the state in domestic affairs.

This new-born and shifting system of social organization has not yet crystallized to a degree which would permit of precise description of the permanent structure of Fascist society. Thus far in matters of social organization, as in policy, the Fascist leaders have adopted opportunistic tactics, governed by the exigencies of the moment, setting up such mechanism in industry as will best serve their purposes. Still, assuming that the present corporate organization is to be

considered permanent, and taking seriously what the Fascist leaders say as to their future purposes, the general nature of a Fascist society may be summarized somewhat as follows. The economic and political life of the people is to be organized in a single comprehensive set of institutions. *Laissez faire* is to disappear from both branches of the nation's life. Private initiative however is to be retained in economic affairs, production being left in the hands of organized groups of employers and workers, which will elevate private initiative from the plane of competing individual enterprises to the corporate plane embracing all branches of the industry. The state will intervene at will to regulate prices, wages, and profits, direct the flow of capital among different industries, and determine the relative quantities of different goods — all in the interest of maximum national power.

Fascist achievements. It will suffice for our purposes to measure the achievements of Fascism against its own avowed objectives, summarizing these for the purpose as follows: First a destruction of liberalism and democracy in the political life of the nation and the substitution of a totalitarian state controlled by the dictatorship of a single party, itself governed by a dictatorial leadership. Second the promotion of national grandeur and prestige both in the domestic affairs of the country and in relation to other states. Third the accomplishment of these aims without essential change in the basic institutions of private property and individual enterprise in economic affairs.

The success of Fascism with regard to the first of these purposes has been made clear in the foregoing description of party and governmental organization and requires no further emphasis.

The second objective has taken concrete form, principally in terms of the military strength of Italy and the employment of military power and diplomacy to bring the nation to a position of importance in world affairs. But military preparedness has in turn necessitated comprehensive reshaping

of the economic life of the country. At the outset Italy was largely dependent on foreign nations both for the instruments of warfare and for the raw materials — oil, metals, rubber, cotton — without which modern war cannot be carried on. With regard to raw materials this will always remain true unless a colonial empire containing these resources can be acquired, for the nature of the country has it so. In pursuit of their militaristic purposes the Fascists have come nearer than in any other phase of their program to the adoption of a rigid plan laid out in terms of years. They have undertaken to make the country as nearly self-sufficient as possible in war materials and within as short a space of time as may be. This has called for a huge investment of capital and labor in the war industries, such as shipbuilding, transportation, airplanes, the heavy metal trades, and other large outlays on raw materials, both in the accumulation of reserve stocks and in devising processes for the production of substitutes, such as new methods of refining oil from low-grade shale, the production of textiles from milk, synthetic products of the chemical industries, and the like. There has been a great expansion of the hydro-electric industry, evidenced by the fact that most of the large-scale factories now use electric power and some 4,000 miles of railway have been electrified. The railways have been transformed in equipment and service until they now rank among the most efficient in the world. A large merchant marine has been developed; manufacturing enterprises previously existing have been expanded and modernized and new ones created.

Without offering any judgment upon the militaristic objectives of Fascism, the movement must be credited with remarkable achievements. Taking command of a country feeble in fighting strength, of slight influence in the international arena, and apparently possessed of negligible surpluses of wealth and energy, the Fascists have brought Italy into the front rank among modern warlike nations. It is not necessary to support this statement with statistics of armament; its

truth is demonstrated by recent diplomatic and military history. It should be borne in mind of course that the ultimate success of the whole military program has yet to be brought to the test of a war with a powerful state.

With regard to the avowed purpose of Fascism to preserve private property and private enterprise in economic affairs the record is inconclusive, since the whole system is still in transition. Outright governmental ownership and operation extends over only a small part of Italian industry, but rigid political control has robbed private property of most of its customary privileges and closed to private initiative most of its opportunities. Although the banking system is ostensibly a private enterprise, credit is a virtual government monopoly, since both long- and short-term credit is available only to such borrowers, in such amounts, and at such rates as the dictators direct. Industries are apparently organized in private ownership, but they produce what they are told in kind and quantity, and their costs are fixed by decree, as are many of their sale prices. When an enterprise established under private direction by political mandate proves unprofitable the government takes over the deficit. If a private business returns a profit, the government takes most of it in taxation. When wealth accumulates in private hands, expropriation occurs, as in the case of the recent capital levy of five per cent on landed and business property. Foreign trade and finance are completely monopolized by the government. Wage rates are fixed by political authority, and many items essential to the cost of living of the common people are sold at prices similarly fixed. In all but name this is a system of state capitalism.

Economic welfare of the people under Fascism. It is difficult to appraise the record of Fascism in terms of the material well-being of the people subject to it. Moreover such an appraisal is irrelevant to the avowed purposes of the program, for these do not aim at improvement in the economic lot of the citizen. The individual is supposed to gain enough

in pride and self-esteem from the growing grandeur of the state to compensate for any sacrifices demanded of him. It must be remembered too, when measuring the economic returns of Fascism to the Italian people, that in addition to the heavy costs of her preparedness program Italy had to bear the cost of an expensive campaign in Ethiopia and is now drawing heavily on her own resources to modernize that country. There is no magic in dictatorship to prevent militarism and war from taking their inevitable toll from the real incomes of a people.

In some directions however Fascism has reacted favorably on the day to day lives of the Italian people. National pride, as well as military strength, has required that certain prevailing conditions of degradation and poverty be eliminated. Thus there has been a concerted attack on illiteracy, on unsanitary living conditions, on epidemic disease, and the like, all of which were symptoms of a low standard of living. In these matters the Fascist dictatorship has made notable advances throughout the country.

In the more strictly economic aspects of their life the people have as yet gained little or nothing from Fascist rule. The level of material comfort among the Italian people was very low before the revolution. Nearly half the people lived by farming, either on small plots of ground owned by the family or by share-cropping the lands of others. This great section of the people never obtained more than a very simple subsistence from their labors, and both in relative numbers and in economic status the rural class of Italy remains today about where it was in 1922. The wage-earning class, on the other hand, has passed through various changes of fortune under the Fascist régime. For some of them — the higher wage groups prior to the revolution — there has been a sharp and permanent decline in real income. In fact during the first decade the average real income of the whole wage-earning class declined, but some restoration of these losses has been made in the years just past. The government has recently

adopted policies more favorable to labor. Since 1930 hours of work have been reduced by about 20 per cent; in many industries the forty-hour week is now the standard. Money wages have been raised, and real wages, though still below their peak, are above their low point, despite the shortening of the work day in the meantime. One factor beneficial to all ranks of labor has been the reduction of unemployment. At the present time the average number of unemployed is about 500,000, less than 25 per cent of what prevailed in the lean years of the Fascist régime, and about equally superior to the average condition of the country prior to the revolution. This improvement in employment is due of course to the armament program and not to any distinctive merit of Fascist social theory, but its effect upon the welfare of the Italian people is none the less real. The price-fixing activities of the government, particularly with regard to foodstuffs, have recently tended to improve the real incomes of the common people. The propertied classes, including the owners of industry, are in a less enviable situation; great private fortunes and large incomes have virtually disappeared under the régime of rigid governmental control of profit-making industry.

Fascism in Germany: The Nazi Party. The National Socialist Workers Party is both a political and a military mechanism. Its military arm, the brown-shirted Storm Troops (S.A.), was organized at an early date to give leadership to the rank and file of its membership and to supplement the propaganda of the leaders with acts of terrorism and violence. The Storm Troops are today a standing army of some 2,000,000 men, drilled and equipped; they constitute an armed force outside the official standing army of the Reich, which is an adjunct of the government. It is evident that the party is itself a fighting as well as a governing machine. The élite of the Storm Troops form the Party Guard (S.S.), numbering some 250,000 armed and disciplined men, a sort of praetorian guard of the Leader.

As a political mechanism, the party is organized on the

model of an all-inclusive national government. At its head stands the Leader and his cabinet of nineteen, each of whom is chief officer of a department or section of national policy. Some of these departmental heads have to do with party affairs, such as the Chief of Staff of the Party Police and the Leader of the Nazi Youth Organizations, who are members of the party cabinet. But included among them are departments which cover all the major subdivisions of the nation's government: Foreign Affairs, Colonies, the Treasury, the Judiciary, Commerce and Industry, Agriculture, the Press, and Propaganda. Thus although the organs of official government remain in existence as administrative units, the party structure places over each of them a policy-forming body of its own. Each of the departments or sections has under its administration a complete functional organization extending throughout the country. The country is also subdivided regionally into sections, each under the control of a party chief.

In addition to its military and governmental structure the party includes certain affiliated organizations: youth leagues for boys and girls, leagues of professional men, welfare organizations, the influential Labor Front. Admission to membership in the party is now limited to graduates of the youth organizations, which serve to indoctrinate the children of the country in the National Socialist creed.

Like the state which it dominates, the Nazi Party is organized on the principle of leadership. Hitler as Supreme Leader appoints the members of the cabinet to serve not as his colleagues but as his subordinates. The structure of the party, the duties of its organs and officers, the policies it avows, are all dependent on his will. Each individual member of the party and its affiliated organizations swears unquestioning allegiance to him. A hierarchy of officials, each under complete subordination to an immediate superior, provides a channel through which this unlimited one-man authority extends from its centre into every element of the national life.

The party and the state. In the course of the Nazi revolution the German Reich lost its federal character and became a completely centralized state. Local autonomies of every sort have been wiped out, and authority has been vested solely in a single national government. The powers of this government are unlimited, not only regionally but functionally, every constitutional guaranty of personal liberty or the immunities of property having been removed. There is no separation of legislative and executive power, for both are exercised by the same body of men. The legislative process itself may be set aside and the country governed by executive decree, each decree having the force of law for a four-year period.

This authoritarian state is dominated by the party through devices which may be summarized as follows: The party is itself the only legal political organization of the people; all other political groupings in existence at the time of the revolution were disbanded, and the formation of new ones was prohibited. The citizens therefore can vote only for Nazi candidates; in fact an election is really only a plebiscite voicing approval of decisions already made. The state and the party are united at the top in Hitler, who is both Supreme Leader of the party and President-Chancellor of the Reich, holding tenure for life with power to appoint his subordinates. The bureaucracy of government from top to bottom is appointive, with the Chancellor's approval required in each case. The party cabinet not only formulates policy before it is acted upon by the government, but some of its bureau chiefs are at the same time heads of the corresponding departments of government. A Nazi organization has been set up to scrutinize and control the actions of all governmental officials, so that in execution as in purpose laws of the state may be brought into harmony with National Socialist principles.

Nazi social and economic program. As its name implies, the Nazi movement began as an avowed enemy of the existing form of capitalism, professing socialist principles. Its

official platform or program contained many specific proposals to which either the socialists or the communists might well have subscribed, such as the abolition of all income not acquired from labor, the nationalization of all trusts and monopolies, the distribution of the profits of all large industries, the communalization of the retail marketing system, the expropriation without compensation of large landed estates, and a broad program of public care for the aged, the sick, the poor, the unemployed. In reality however these proposals were adopted to win adherents, as were other parts of the platform, many of whose policies were irrelevant or even contradictory to the social purposes stated above. It seems a fair interpretation of the facts to state that in its formative years National Socialism was primarily a drive for political power, promising anything to an embittered and impoverished people which would be likely to win its support. Still its leaders did have in mind certain definite purposes to which political power would be put when once in their hands, and these may be called the broad social objectives of the movement.

The whole set of objectives coheres around a common point: the power and prestige of the German nation. With this end in view the internal program embraced the following policies. First the obliteration of democracy, with all its divisive influences in domestic politics, and the substitution of an authoritarian state under dictatorial leadership. Second the creation of a totalitarian social system; that is, an organization of society overriding all personal and property rights, which would bring the wealth and labor energy of the German people entirely into the service of the state. Third the promotion of loyalty and contentment in the great mass of the German people. This was to be accomplished in part by playing upon the emotions of racial and national pride which the humiliating position of Germany at the time had made extremely sensitive; in part by increasing the economic security, the dignity in the industrial world, and if possible the

material comfort of the common people; and also, where necessary, by ruthless suppression and extermination of dissenting groups. It is this feature of their program which has required the Nazi leaders to seek the basis of their support primarily in the labor and lower middle classes. Finally the program included the utilization of these powers of political dictatorship and social control to focus the energies of the nation on military preparedness.

Government industry. Thus far in its development National Socialism has avoided any large measure of outright governmental ownership and operation of productive enterprise. The transportation industry, 80 per cent of which is public property, is an exception to the general rule. But this industry was already publicly owned long before the Nazi régime. It has proved necessary also to monopolize the sources of long- and short-term credit in order to direct the flow of capital funds and the financing of trade; hence the Nazi régime has taken over the banks and the exchange markets. Through its control of exchange operations and by means of a quota system applied to imports and exports, the state has come to dominate foreign commerce so completely that this too is sometimes called a governmental monopoly. But in this case the actual business transactions are carried on for the greater part by private concerns subject to detailed restrictions, rather than by governmental agencies, so that technically it is not a state enterprise.

The "estates." The great bulk of Germany's production enterprise remains therefore in private hands, but in harmony with their totalitarian principles the Nazi leaders have regimented each department of economic activity into a nation-wide organization, called an "estate." These in their broadest groupings are two in number: the Estate of Industry and Commerce and the Estate of Agriculture. The organization of each estate is mapped out on paper down to the smallest detail. The Estate of Industry and Commerce, for example, is subdivided into six separate functional entities: manufac-

ture, commerce, banking, insurance, handicraft, electric power. It is also subdivided regionally into eighteen district boards. Each functional subdivision is itself broken down into smaller units, representing separate categories of productive enterprise. The functional and regional subdivisions have their own governing bodies, and at the head of the estate stands the Reich Economic Chamber, a policy-forming body for all the different branches of industry and trade. Agriculture is similarly organized into functional and regional subdivisions, with a complexity of governing boards and councils drawn together at the top into a national chamber. Membership in one of these estates or in one of its appropriate subdivisions is compulsory for all business enterprises, large or small, throughout the Reich.

It would not be profitable to describe in detail the elaborate organization of Germany's business activity under the Nazi régime. The whole system is in a state of flux, with new governing agencies appearing and disappearing, and areas of jurisdiction shifting from one agency to another as Nazi policy develops. It appears true however that the general principles and purposes which this form of compulsory, non-competitive, nation-wide organization embodies are firmly established in German economic life. The details of organization may change, but the farmer, the merchant, the manufacturer, the banker — in fact any person seeking an income from the ownership or management of an enterprise — has accepted the principle of subordination in a structure dominated by the state. These organizations are supposed to provide for self-government in industry. In reality they are all founded on the leadership principle, the government having power to appoint their officers and to prepare policies determining the kinds and quantities of goods to be produced, the methods of production or business operation to be employed, the costs to be met, and the sale prices to be accepted. In preparing these policies the government may draw in from the elaborate structure of the estates representatives of the business world to

serve in a consultative and advisory capacity. But the final decisions are political and authoritarian, and the machinery of the estates is used to enforce them upon the individual business man.

Labor organization. Labor is not included in the organizations described above, these being limited to the property and managerial interests. All previously existing trade unions have been dissolved, and in their place has been set up a nation-wide grouping of both employers and workers — the so-called Labor Front. Membership is not legally compulsory but, in view of the dangers encountered by the dissenter, is virtually so in fact. The basic unit includes the employer and the wage earners in a single concern; these are gathered into fairly small local groups, some 15,000 in number, which are drawn together into some 800 district groups, and these in turn into about 30 regional divisions. The whole organization is political in purpose. Its units from top to bottom are affiliated with the National Socialist Party, its officers must all be party members, and it is governed from the top by the cabinet minister of the party in charge of the Department of Political Organization. The government is of course dictatorial rather than democratic.

The Labor Front has no bargaining functions but exists primarily to unite employers and workers in broadly cultural activities and in numerous coöperative undertakings for the betterment of living conditions among the common people. It also administers industrial insurance, savings, and building companies.

The outlook of the Nazis repudiates the Marxian doctrines of social stratification and class struggle. In accordance with this view strikes and lockouts are prohibited as offenses against the state, and workers and employers are obliged as a matter of patriotic honor to come to an agreement as to the terms upon which they are to collaborate in service to the nation. For this purpose there has been set up a peculiarly Nazi organization of the labor market on leadership lines. The em-

ployer of each concern is designated the "leader" of that enterprise, with powers to announce the terms of employment, having first consulted a council made up of representatives of the wage earners, called his "followers." The council members, though elective, must be selected from a panel of candidates drawn up by the leader with the approval of the local unit of the Nazi Party, one such unit having been organized in every business concern of the country employing more than twenty persons. This whole procedure is in turn supervised by a system of regional Labor Trustees, whose personnel is appointed by the government, and who report to and receive instructions from the Ministry of Labor. The trustees have virtually unlimited power to abrogate a labor contract and fix new schedules of wages and hours. In reality therefore the labor costs of industry and correlatively the money incomes and work conditions of the wage earners are determined by political authority.

Economic results of the Nazi program. The period of the Nazi rule has been too brief to permit of any final appraisal in economic terms. During this time every phase of policy has been dominated by militaristic purposes, which have entailed immense investments of capital and labor in lines not productive of consumable income and the equally large sacrifices in the interests of national self-sufficiency which militarism postulates. In this major purpose the Nazi régime has been successful at least in so far as the immediate fighting strength of the nation is concerned. The rapid expansion of every branch of Germany's war machine in size, in material equipment, and in organization attests the efficacy of the Nazi principles of social unity and leadership to develop productive power and keep it focused. The drive toward economic self-sufficiency has not attained its goal and will probably never do so, because of insuperable natural difficulties. It is in this branch of their program only that the Nazi régime has set up an economic plan on the Russian model — a "Four Year Plan" for the attainment of self-sufficiency in food and raw

materials launched by Hitler in 1936. Thus far no important results have been obtained from this plan. Isolated instances of the discovery of synthetic substitutes for natural products have been dramatized in the press, but the food situation in Germany is if anything worse than it was before, and the country remains as dependent as formerly on foreign supply of raw materials.

Having in mind these costly aspects of the Nazi program, it is understandable that the people have not benefited in terms of economic welfare. Real incomes have declined. There is a shortage of goods required by the common people, which has forced the government to ration supplies. This is particularly evident in the case of foodstuffs, which have fallen so far short that the standard of nutrition for the whole country has been estimated to have declined by nearly twenty per cent. The Nazi régime has adjusted its policy to this situation in two ways: by preventing the dearth from enriching the merchants and middlemen at the expense of the consumer, and by bringing privation to bear with greater relative severity upon the wealthy classes than upon the common mass of the people. The marketing system has been subjected to such meticulous regulation that it remains a private enterprise only in form. Wholesale trade in food has been abolished entirely, and the profit margin of the retailer has been drastically reduced by setting minimum prices on his purchases from producers — designed to promote production as a phase of the self-sufficiency program — and maximum prices on his sales to consumers. The price-fixing policy also aims at equalizing living conditions among the various social classes by discriminating in favor of ordinary as compared with "quality" goods. In sum the declining prosperity of the country has produced a greater equality of real incomes, while reducing the average to a lower level.

General appraisal of radical movements. In this and the preceding chapter we have considered the abstract principles underlying the chief movements of radical change: socialism,

communism, and Fascism; we have sketched the concrete policies of the only experiment in socialism with which the world is familiar and have done the same with regard to the two Fascist régimes. It is in point now to attempt a general appraisal of the merits of these programs as compared with capitalism. The complexity of similarities and contrasts which characterizes these radical movements makes it difficult to draw any general conclusions which have final validity. In the discussion which follows we shall emphasize their common attributes and especially those which stand in most definite contradiction to the essential features of a capitalist order.

Authority versus liberty. One of the chief charges in the indictment of capitalism is that this system of society assures no real freedom for the average man. Granting this accusation all the weight it deserves, it certainly remains true that the evil has been magnified in the systems which have been substituted for the capitalist order. We need not dwell upon the incidents of massacre, imprisonment, and brutality which have horrified the democratic world; these are characteristic of dictatorship regardless of its social philosophy. But it is of the essence of capitalism, with all its inequalities in the distribution of wealth, power, and privilege, to rely upon the voluntary actions of individuals to provide for social needs. And it is equally essential to any controlled economy to substitute compulsion for voluntary action. The modern world now has experience of both methods in operation simultaneously and can easily judge of their results in terms of human freedom.

The constitutional socialists will hold this criticism inapplicable to their system, since they support democracy in industry and in political government. But it is questionable whether in practice they could avoid a wide application of compulsion. Even in a democratic régime a system of public control, instituted over the major economic activities of the citizens, must ramify until it embraces practically all the

interests of a normal human life. To most men a society where the type of work in which a man may engage, the income which he receives, and the kinds of wants he is permitted to satisfy are all determined by authority bears little resemblance to freedom.

Planned economy and efficiency. Since comprehensive planning has been put to the test only in Russia, inferences regarding it must relate primarily to that country. It would appear that the soviet experiment in economic planning has been, on the whole, a success. It is true that things have not gone always according to prediction. There have been partial failures and unforeseen difficulties in many branches of the program, but this is to be expected of so gigantic and complex an undertaking. The planning commission has succeeded in guiding the nation's economic activity in general toward the goals set up in advance, and this achievement has been interpreted as evidence of a superior virtue in communism or socialism. No such inference can be drawn from the facts. The success of the experiment in economic planning has been due to the dictatorship whose power stands behind it; and any equally strong dictatorship, whether of a socialist or a capitalist variety, would be able to accomplish similar results. Recalling again that communism in its true form presupposes a régime of democracy, the Russian experiment gives us no assurance as to the practicability of this aspect of the communist program.

Furthermore there is reason to believe that the attempt at economic planning has been greatly facilitated by certain special conditions in the country. Throughout this period the country has been laboring under conditions of extreme scarcity in all departments of its economic life. There has been a dearth of virtually every type of good and service, both for the satisfaction of human needs and for the development of basic productive equipment. In many respects such a situation greatly simplifies the task of a planning commission. There is no problem of oversupply in any line of production;

the authorities can safely assume a consumers' requirement larger than any product that it is possible for them to provide. There is no danger of an excess of labor in any branch of activity, since all are alike under-supplied with labor. No great thought need be given to the question whether the people would prefer more of this good and less of that, or would desire different styles or qualities of goods, since in the prevailing scarcity they will be happy to have any increase at all in the provision for their needs. Obviously this condition is not typical of advanced industrial societies, where the problem is one of relative overproduction in many lines of activity and relative oversupply of certain types of labor and capital, and where the preferential demands of the people must be considered. In these more typical situations the task of a planning commission must be immensely more complex and difficult, and until the planned economy has been tested under these more typical conditions it is unsafe to pass favorable judgment upon it.

What has been said applies in some measure also to economic planning in a Fascist régime. It is unthinkable that anyone seriously desires the nations of the world to remain perpetually on a war footing, their entire life dominated by the voracious demands of militarism. What success economic planning has had in Germany and Italy is pertinent almost exclusively to war. But the efficiency of a planned economy as compared with one of individual enterprise can be decided finally only in terms of the welfare of the people in peace time. Neither communist nor Fascist experiments give us evidence on this point.

Planned economy and capital formation. The incentive of an interest return is believed to be necessary for the accumulation of capital in a system of free enterprise. The authoritarian experiments have shown that capital can be provided under authoritarian management without appealing in this way to individual self-interest. The soviet government in particular has certainly acquired a vast capital equipment. There

has been some resort to borrowing through interest-bearing government bonds; but the funds raised in this way have been only a small part of the total capital accumulation, and the purchase of the bonds by the people has been more largely a response to social pressure than to the attraction of the promised interest payment. For the most part the capital has been provided by the people through savings which they either did not know they were making or which they could not avoid. Funds raised by taxation and goods acquired by governmental requisition are examples of compulsory savings. A more general form of saving, often unknown to the people who have borne the burden, has resulted from the government's power to determine the quantities of goods and services which were allowed to enter into direct consumption and the quantities to be employed in the production of capital instruments, a power which it possessed by reason of its control of the entire productive system. To the average man in Russia the effect of this governmental policy has appeared in the form of a scarcity of consumption goods; he need not know that this scarcity merely reflected the government's decision to accumulate capital as rapidly as possible or, in other words, that he was bearing the burden of saving. It is in these ways that a socialist régime, if given sufficient authority over the people, can meet its needs for capital equipment while dispensing with the incentive of personal award to the saver.

But in connection with this matter it should be noted that a socialist system has no power to alter the process of capital formation in its essentials. The accumulation of capital involves a sacrifice of consumable income by some individual or group in society. Socialism does not make saving burdenless but merely conceals the burden. This concealment is possible because the individuals who make the sacrifice do not take part in the decisions of the governing authorities. In other words, it is the dictatorial character of the soviet government which gives it power to accumulate capital without the knowledge or against the will of those who provide income for this

purpose. In a democratic system the people at large would be called upon to make the decision whether to live more comfortably in the present or provide for the increased comfort of later years and succeeding generations; and it is a safe conjecture that they would require some real incentive to induce them to forego present enjoyment. The soviet leaders foretell the time when their governmental régime will be thoroughly democratic. Until the matter is put to a test at that time we have no evidence that democratic socialism — and this is the avowed objective of all socialist systems — will be able to acquire and maintain its indispensable capital equipment while abolishing pecuniary rewards to savers.

Judgment of socialism. In conclusion it will be profitable to return to the socialist indictment of capitalism as summarized in the immediately preceding chapter. Every thinking man admits that each of the charges in this indictment contains elements of truth. The present order is no more perfect than is any other work of man. In fact it contains many and serious faults, faults of inefficiency, of injustice, above all of retarded and imperfect adjustment to the changing conditions of human life. The socialist movement is to be credited with the great service of having focused attention on these faults and accelerated their correction. Its indictment of capitalism however is not convincing, for the following reasons.

(1) Socialism exaggerates the inefficiency of the present order. The indictment of capitalism is overdrawn. After all has been said that can be said against it, this system has succeeded in maintaining the largest population the world has ever known upon the highest level of average comfort. It has crowded into the short space of a century and a half greater advance in all the arts of production than was achieved during the whole preceding space of society's history. It has equipped mankind with an unprecedented supply of capital instruments and is increasing the quantity and improving the quality of this equipment without pause. The uncontrolled individualistic system of production and marketing, despite grievous

lapses in times of economic unsettlement, has worked for many years to supply the individual consumer with want-satisfying goods drawn from the resources of the entire world. It has accomplished this by providing incentives which call into play two powerful human impulses: acquisitiveness and the desire to excel in competition. The institution of private property supplies the former of these incentives; individual enterprise, the latter. Socialism aims to suppress the motive of acquisition entirely and to deprive success in competition of all material rewards. In their place it would rely upon the motives of brotherhood, social service, and altruism. It is not necessary to deny that human nature possesses these latter characteristics or that they are of higher order than those nourished by a capitalist environment. But there is grave reason to doubt that they will prove as effective in stimulating human energy along lines of economic activity as the motives which have been responsible for the unparalleled material success of capitalism.

(2) Socialism underestimates the vitality of capitalism. The survival of social institutions depends upon their power of adjustment to new conditions as they arise and of providing within themselves mechanism for solving the problems which are created by their own growth. Capitalism has this power. It is a flexible system. The institutions of property and competition can be qualified when need arises without destroying their fundamental utility. Collective bargaining can be substituted for the individual contract when the latter works injury to the weaker members of society. Social insurance can be developed to safeguard the workman against many of the hazards of the wage earner's life. Labor legislation can assure him minimum standards of safety, sanitation, even of income. Collective provision may be made for the services which private initiative will not supply, such as education, agencies of public health, recreation facilities. Producers can be restrained from undermining and destroying each other by unfair competitive practices. The consumer can be protected

against adulterated goods and scrimped quantities by laws enforcing standards of purity and honesty of weight and measure. The overweening power of great business can be restrained by public control. All these things can be done without discarding the basic institutions of society. Although we may not be content with the rate of progress attained along these lines in the past, that there has been progress no one can deny. The true conservatives of our day are those who lend their support to these reforms; those who oppose timely readjustment of traditional ways and methods must be ranked among the enemies of the present order.

SUGGESTIONS FOR FURTHER READING

The books in the following bibliography are grouped under general headings, whose arrangement agrees with the order in which the respective subjects are treated in the text.

I. GENERAL ECONOMICS TEXTS

SLICHTER, S. H. *Modern Economic Society.* New York, 1931

ELY, R. T., and others. *An Outline of Economics.* Sixth edition. New York, 1937

SPAHR, W. E., and others. *Economic Principles and Problems.* Third edition. New York, 1936

ORTON, W. A. *Prelude to Economics.* Boston, 1932

GARVER, F. B., and HANSON, A. H. *Principles of Economics.* Boston, 1937

MEYERS, A. L. *Elements of Modern Economics.* New York, 1937

—— *Modern Economic Problems.* New York, 1939

MARSHALL, A. *Principles of Economics.* Seventh edition. London, 1916

EDIE, L. D. *Economics: Principles and Problems.* Second edition. New York, 1932

TAUSSIG, F. W. *Principles of Economics,* 2 vols. New York, 1939

FAIRCHILD, F. R., FURNISS, E. S., and BUCK, N. S. *Elementary Economics,* 2 vols. Fourth edition. New York, 1939

II. ECONOMIC JOURNALS

The American Economic Review. Cambridge, Mass.

The Quarterly Journal of Economics. Cambridge, Mass.

The Journal of Political Economy. Chicago, Ill.

The Economic Journal. London

The Annals of the American Academy of Political and Social Science. Philadelphia, Pa.

The Political Science Quarterly. New York

The Journal of the American Statistical Association. Concord, N. H.

The Economist. London

Econometrica. Menasha, Wis.

The Commercial and Financial Chronicle. New York

The Review of Economic Statistics. Cambridge, Mass.

The Federal Reserve Bulletin. Washington, D. C.

III. ECONOMIC HISTORY

DAY, C. *History of Commerce.* New York, 1922

—— *Economic Development in Modern Europe.* New York, 1933

SINCLAIR, H. M. *A Preface to Economic History.* New York, 1934

OGG, F. A., and SHARP, W. R. *The Economic Development of Modern Europe.* New York, 1926

HEATON, H. *Economic History of Europe.* New York, 1936

CLAPHAM, J. H. *The Economic Development of France and Germany, 1815–1914.* London, 1921

—— *An Economic History of Modern Britain,* 3 vols. New York, 1927, 1932, 1937

SHANNON, F. A. *Economic History of the People of the United States.* New York, 1934

FAULKNER, H. U. *American Economic History.* New York, 1931

THOMPSON, C. M., and JONES, F. M. *Economic Development of the United States.* New York, 1939

CALLENDER, G. S. *Selections from the Economic History of the United States.* Boston, 1909

BOGART, E. D. *Economic History of the United States.* Revised edition. New York, 1938

MYERS, G. *History of the Great American Fortunes.* New York, 1936

COLE, A. H. *Fluctuations in American Business, 1790–1850.* New York, 1935

GRAYSON, T. J. *Leaders and Periods of American Finance.* London, 1932

IV. CORPORATIONS AND INVESTMENT

FIELD, K. *Corporation Finance.* New York, 1938

HOOGLAND, H. E. *Corporation Finance.* New York, 1938

STEVENS, W. M. *Financial Organization and Administration.* New York, 1934

HANEY, LEWIS H. *Business Organization and Combination.* New York, 1934

DEWING, A. S. *Financial Policy of Corporations.* Revised edition. New York, 1934

—— *Corporation Finance.* New York, 1931

—— *Corporate Promotions and Reorganizations.* New York, 1914

LINCOLN, E. E. *Applied Business Finance.* Chicago, 1924

CHAMBERLAIN, L., and HAY, W. W. *Investment and Speculation.* New York, 1931

V. STATISTICS AND ACCOUNTING

CRUM, W. L., PATTON, A. C., and TEBBUTT, A. R. *Introduction to Economic Statistics.* New York, 1938

SMITH, J. G. *Elementary Statistics.* New York, 1934

CROXTON, F. E., and COWDON, D. J. *Practical Business Statistics.* New York, 1934

MILLS, F. C. *Statistical Methods Applied to Economics and Business.* New York, 1938

HUSBAND, G. R., and THOMAS, A. E. *Principles of Accounting.* New York, 1935

PORTER, C. H., and FISKE, W. P. *Accounting.* New York, 1935

KESTER, R. B. *Accounting Theory and Practice.* New York, 1930

VI. PRICE

MILLS, F. C. *The Behavior of Prices.* New York, 1927

SCHULTZ, H. *Statistical Laws of Demand and Supply.* Chicago, 1928

NOURSE, E. G., and DRURY, H. B. *Industrial Price Policies and Economic Progress.* Washington, 1938

HENDERSON, H. D. *Supply and Demand.* New York, 1922

FISHER, I. *Elementary Principles of Economics.* New York, 1912

MCISAAC, A. M., and SMITH, J. G. *Introduction to Economic Analysis.* Boston, 1937

MCCONNELL, D. W., and others. *Economic Behavior.* Revised edition. Boston, 1939

WIESER, F. VON. *Natural Value.* (English translation.) London, 1893

CHAMBERLIN, E. *The Theory of Monopolistic Competition.* Cambridge, Eng., 1933

ROBINSON, MRS. J. *The Economics of Imperfect Competition.* New York, 1933

DAVENPORT, H. J. *The Economics of Enterprise.* New York, 1913

MEEKER, J. E. *The Work of the Stock Exchange.* New York, 1922

VII. THE DISTRIBUTION OF WEALTH AND INCOME

CLARK, F. B. *The Economics of Distribution.* New York, 1933

MARSHALL, A. *Principles of Economics.* Eighth edition. London, 1920

GEORGE, H. *Progress and Poverty.* New York, 1880

TAYLOR, H. C. *Outlines of Agricultural Economics.* New York, 1931

DORAN, H. B., and HINMAN, A. G. *Urban Land Economics.* New York, 1928

FISHER, I. *The Theory of Interest.* New York, 1930

MOULTON, H. G. *The Formation of Capital.* Washington, 1935

DOUGLAS, P. H. *The Theory of Wages.* New York, 1934

HICKS, J. R. *The Theory of Wages.* London, 1932

DAUGHERTY, C. R. *Labor Problems in American Industry.* Revised edition. Boston, 1938

KNIGHT, F. H. *Risk, Uncertainty, and Profit.* New York, 1921

FOSTER, W. F., and CATCHINGS, W. *Profits.* Newton, Mass., 1925

EPSTEIN, R. C. *Industrial Profits in the United States.* New York, 1934

KING, W. I., and EPSTEIN, R. C. *The National Income and Its Purchasing Power.* New York, 1930

BOWLEY, A. L. *The Division of the Product of Industry.* Oxford, 1919
National Industrial Conference Board. *National Income in the United States 1797–1938.* New York, 1939
LEVEN, M., and WRIGHT, K. R. *Income Structure in the United States.* Washington, 1938
KUZNETZ, S. *Studies in Income and Wealth.* New York, 1937
MILLS, F. C. *Economic Tendencies in the United States.* New York, 1933

VIII. MONEY AND BANKING

WESTERFIELD, R. B. *Money, Credit and Banking.* New York, 1938
BRADFORD, F. A. *Money and Banking.* Second edition. New York, 1935
WHITE, H. *Money and Banking.* New edition. Boston, 1935
CURTIS, M., and TOWNSHEND, H. *Modern Money.* New York, 1938
KEMMERER, E. W. *On Money.* Philadelphia and Chicago, 1934
National Industrial Conference Board. *The New Monetary System of the United States.* New York, 1934
DONALDSON, J. *The Dollar, A Study of the "New" National and International Monetary System.* New York, 1937
GAYER, A. D. *Monetary Policy and Economic Stabilization: A Study of the Gold Standard.* New York, 1935
WESTERFIELD, R. B. *Our Silver Debacle.* New York, 1936
MADDEN, J. T., and NADLER, M. *The International Money Market.* New York, 1935
HEILPERIN, M. A. *International Monetary Economics.* New York, 1939
HELDERMAN, L. C. *National and State Banks — A Study of Their Origins.* Boston and New York, 1931
REED, H. *Federal Reserve Policy, 1921–30.* New York, 1930
HARRIS, S. E. *Twenty Years of Federal Reserve Policy.* Cambridge, Mass., 1934
SOUTHWORTH, S. D. *Branch Banking in the United States.* New York, 1928
The Twentieth Century Fund. *Debts and Recovery.* New York, 1938
WILLIS, H. P., and BECKHART, B. H. (Editors). *Foreign Banking Systems.* New York, 1929
DEWEY, D. R. *Financial History of the United States.* Twelfth edition. New York, 1934 (see Table of Contents)
NOYES, A. D. *Forty Years of American Finance.* New York, 1909
—— *The War Period of American Finance.* New York, 1926

IX. MONEY AND PRICES

FISHER, I. *The Purchasing Power of Money.* New York, 1911
MARGET, A. W. *The Theory of Prices.* New York, 1938
FISHER, I. *The Making of Index Numbers.* Boston, 1922
EDIE, L. D. *Money, Bank Credit, and Prices.* New York, 1928

WARREN, G. F., and PEARSON, F. H. *Gold and Prices.* New York, 1935
SPAHR, W. E. *The Monetary Theories of Warren and Pearson.* New York, 1934
FISHER, I. *Stable Money.* New York, 1935
WILLIS, H. P. *The Theory and Practice of Central Banking, with Special Reference to American Experience.* New York, 1936

X. PUBLIC FINANCE. GENERAL TEXTS

LUTZ, H. L. *Public Finance.* Third edition, New York, 1936
JENSEN, J. P. *Government Finance.* New York, 1937
BUEHLER, A. G. *Public Finance.* New York, 1936
SHULTZ, W. J. *American Public Finance and Taxation.* Second edition. New York, 1938
BASTABLE, C. F. *Public Finance.* Third edition. London, 1903
SELIGMAN, E. R. A. *Studies in Public Finance.* New York, 1925
PIGOU, A. C. *A Study in Public Finance.* New York, 1928
BULLOCK, C. J. (Editor). *Selected Readings in Public Finance.* Third edition. Boston, 1924
MILLS, M. C., and STARR, C. W. *Readings in Public Finance and Taxation.* New York, 1932
FAGAN, E. D., and MACY, C. W. *Public Finance, Selected Readings.* New York, 1934
DEWEY, D. R. *Financial History of the United States.* Twelfth edition. New York, 1934 (see Table of Contents)

XI. TAXATION

GREEN, W. R. *The Theory and Practice of Modern Taxation.* Chicago, 1933
STAMP, J. *The Fundamental Principles of Taxation.* New York, 1929
BROWN, H. G. *The Economics of Taxation.* New York, 1924
Twentieth Century Fund. *Facing the Tax Problem.* New York, 1937
—— *Studies in Current Tax Problems.* New York, 1937
National Industrial Conference Board, Inc. *Taxation and National Income.* New York, 1922
—— *Cost of Government in the United States, 1929–1930.* New York, 1932
—— *Cost of Government, 1935–1937.* New York, 1938
United States Bureau of the Census. *Financial Statistics of State and Local Governments: 1932.* Washington, 1935
HOWE, F. C. *Taxation and Taxes in the United States under the Internal Revenue System, 1791–1895.* New York, 1896
SPAULDING, H. B. *The Income Tax in Great Britain and the United States.* London, 1927
LELAND, S. E. *The Classified Property Tax in the United States.* Boston, 1928
JENSEN, J. P. *Property Taxation in the United States.* Chicago, 1931

National Industrial Conference Board, Inc. *State and Local Taxation of Property.* New York, 1930

SHULTZ, W. J. *The Taxation of Inheritance.* Boston, 1926

BUEHLER, A. G. *General Sales Taxation.* New York, 1932

—— *The Undistributed Profits Tax.* New York, 1937

National Industrial Conference Board, Inc. *State and Local Taxation of Business Corporations.* New York, 1930

FAIRCHILD, F. R., and associates. *Forest Taxation in the United States* (Report of the United States Forest Taxation Inquiry). Washington, 1935

FILLEBROWN, C. B. *The A B C of Taxation.* New York, 1909

National Tax Association. *Proceedings of the National Conference* — Annual from 1907. *Bulletin,* Monthly from February, 1916

XII. BUDGET, GOVERNMENT BORROWING, WAR FINANCE, ETC.

GUEST, H. W. *Public Expenditures.* New York, 1928

STOURM, R. *The Budget* (English translation). New York, 1917

BUCK, A. E. *The Budget in Governments Today.* New York, 1934

WILLOUGHBY, W. F. *The National Budget System.* Baltimore, 1927

BUCK, A. E., and others. *Municipal Finance.* New York, 1926

Twentieth Century Fund. *The National Debt and Government Credit.* New York, 1937

NOYES, A. D. *Financial Chapters of the War.* New York, 1916

GRADY, H. F. *British War Finance, 1914–1919.* New York, 1927

JÈZE, G. P. A., and TRUCKY, H. *The War Finance of France.* New Haven, 1927

MOULTON, H. G., and PASVOLSKY, L. *War Debts and World Prosperity.* Washington, 1932

XIII. RAILROAD TRANSPORTATION

DAGGETT, S. R. *Principles of Inland Transportation.* New York, 1934

JONES, E. *Principles of Railway Transportation.* New York, 1924

ACWORTH, SIR W. M. *The Elements of Railway Economics.* Revised edition. Oxford, 1924

VANDERBLUE, H. B., and BURGESS, K. F. *Railroads: Rates — Service — Management.* New York, 1923

SPLAWN, W. M. W. *Government Ownership and Operation of Railroads.* New York, 1928

JOHNSON, E. R. *Government Regulation of Transportation.* New York, 1938

CUNNINGHAM, W. J. *The Present Railroad Crisis.* Philadelphia, 1939

XIV. MONOPOLY AND GOVERNMENT INDUSTRY

JONES, E. *The Trust Problem in the United States.* New York, 1929

LAIDLER, H. W. *Concentration of Control in American Industry.* New York, 1931

JENKS, J. W., and CLARK, W. E. *The Trust Problem.* New York, 1929
BERLE, A. A. JR., and MEANS, G. C. *The Modern Corporation and Private Property.* New York, 1933
FILENE, L., and STODDARD, W. L. *Unfair Trade Practices; How to Remove Them.* New York, 1931
MICHAELS, R. K. *Cartels, Combines, and Trusts in Post-War Germany.* New York, 1928
PLUMMER, A. *International Combines in Modern Industry.* London, 1934
FLYNN, J. T. *Security Speculation.* New York, 1934
COREY, L. *The Decline of American Capitalism.* New York, 1934
BONBRIGHT, J. C., and MEANS, G. C. *The Holding Company.* New York, 1932
PERSONS, W. M. *Government Experimentation in Business.* New York, 1934
DIMOCK, M. E. *Government-Operated Enterprises in the Panama Canal Zone.* Chicago, 1934
—— *British Public Utilities and National Development.* London, 1933
Power Authority of the State of New York. *Report on Cost of Distribution of Electricity.* Albany, 1934
THOMPSON, C. D. *Public Ownership.* New York, 1925
O'BRIEN, T. H. *British Experiments in Public Ownership and Control.* New York, 1938
CRECRAFT, E. W. *Government and Business.* New York, 1925
REQUA, M. L. *Relations of the Government to Industry.* New York, 1925

XV. BUSINESS CYCLES AND THEIR CONTROL

SCHUMPETER, J. A. *Business Cycles.* New York, 1939
PIGOU, A. C. *Industrial Fluctuations.* London, 1927
MILLS, F. C. *Economic Tendencies in the United States.* New York, 1932
WARREN, G. F., and PEARSON, F. A. *Gold and Prices.* New York, 1935
PERSONS, W. M. *Forecasting Business Cycles.* New York, 1931
TIMOSHENKO, V. P. *The Rôle of Agricultural Fluctuations in the Business Cycle* (Michigan Business Studies, Vol. II, No. 9). Ann Arbor, 1930
MOORE, H. L. *Economic Cycles: Their Law and Cause.* New York, 1914
—— *Generating Economic Cycles.* New York, 1923
KUZNETS, S. S. *Cyclical Fluctuations.* New York, 1926
—— *Secular Movements in Production and Prices.* Boston, 1930
MITCHELL, W. C. *Business Cycles: The Problem and Its Setting.* New York, 1930
HASTINGS, H. B. *Costs and Profits.* Boston, 1923
ADAMS, A. B. *Economics of Business Cycles.* New York, 1925
—— *Analyses of Business Cycles.* New York, 1936
HANSEN, A. H. *Business-Cycle Theory.* Boston, 1927
—— *Economic Stabilization in an Unbalanced World.* New York, 1932
HABERLER, G. VON. *Prosperity and Depression.* League of Nations, 1937

XIX. COMMERCIAL POLICY AND PROTECTIVE TARIFF

International Economic Relations. The Report of the Commission of Inquiry into National Policy in International Economic Relations. Minneapolis, 1934

TAUSSIG, F. W. *Tariff History of the United States.* Eighth edition. New York, 1931

TASCA, H. J. *The Reciprocal Trade Policy of the United States.* Philadelphia, 1938

BEVERIDGE, SIR W. H. *Tariffs, the Case Examined.* London, 1932

JONES, J. M., JR. *Tariff Retaliation.* Philadelphia, 1934

BEARD, C. A., and SMITH, H. E. *The Open Door at Home.* New York, 1934

HIRST, F. W. *Safeguarding and Protection in Great Britain and the United States.* New York, 1927

BIDWELL, P. W. *The Invisible Tariff.* New York, 1939

PATTEN, S. N. *The Economic Basis of Protection.* Philadelphia, 1890

DIXWELL, G. B. *The Premises of Free Trade Examined.* Cambridge, Eng., 1881

RICHARDSON, J. H. *British Economic Foreign Policy.* New York, 1936

STALEY, E. *War and the Private Investor.* New York, 1935

—— *Raw Materials in Peace and War.* New York, 1937

XX. LABOR PROBLEMS: GENERAL STUDIES

DAUGHERTY, C. R. *Labor Problems in American Industry.* Boston, 1933

DOUGLAS, P. H., and DIRECTOR, A. *The Problem of Unemployment.* New York, 1931

ROBINSON, J. *Essays in the Theory of Employment.* New York, 1937

KEYNES, J. M. *General Theory of Employment.* New York, 1936

BEVERIDGE, W. *Unemployment: A Problem of Industry.* New York, 1930

PIGOU, A. C. *Industrial Fluctuations.* New York, 1927

FELDMAN, H. *Regularization of Employment.* New York, 1925

—— *Problems in Labor Relations.* New York, 1937

DOWNEY, E. H. *Workman's Compensation.* New York, 1924

WOODBURY, R. M. *Workers' Health and Safety.* New York, 1927

XXI. PUBLIC LABOR POLICY

SOULE, G. *A Planned Society.* New York, 1932

COMMONS, J. R., and ANDREWS, J. B. *Principles of Labor Legislation.* New York, 1927

WOODBURY, R. M. *Social Insurance, An Economic Analysis.* New York, 1917

DOUGLAS, P. H. *Social Security in the United States.* New York, 1935

Social Security Board. *Social Security in the United States.* Washington, 1937

XXII. THE LABOR MOVEMENT

HARRIS, H. *American Labor.* New Haven, 1939
CUMMINS, E. E. *The Labor Problem in the United States.* New York, 1932
DAUGHERTY, C. R. *Labor Problems in American Industry.* Boston, 1938
PERLMAN, S. *Theory of the Labor Movement.* New York, 1928
WEBB, B., and S. *Industrial Democracy,* Part III. New York, 1920
WOLMAN, L. *Growth of American Trade Unions, 1890–1923.* New York, 1925
—— *Ebb and Flow of Trade Unionism.* New York, 1936
HARDMAN, G. B. S. *American Labor Dynamics.* New York, 1928

XXIII. THE INDUSTRIAL CONFLICT

MASON, A. T. *Organized Labor and the Law.* Durham, 1925
International Labor Office. *Conciliation and Arbitration.* Geneva, 1927
SELEKMEN, B. M. *Postponing Strikes.* New York, 1927
WITTE, E. E. *The Government in Labor Disputes.* New York, 1932
FEIS, H. *The Settlement of Wage Disputes.* New York, 1921
BARNETTE, G. E., and MCCABE, D. A. *Mediation, Investigation, and Arbitration.* New York, 1916
YELLEN, S. *American Labor Struggles.* New York, 1936
MILNE-BAILEY, W. *Trade Unionism and the State.* London, 1934
BROOKS, R. R. R. *When Labor Organizes.* New Haven, 1937

XXIV. PROPOSED CHANGES OF THE ECONOMIC SYSTEM

MARX, K. (Edited by Algernon Lee). *Essentials of Marx.* New York, 1926
LENIN, N. *Imperialism. The State and Revolution.* New York, 1926
LASKI, H. J. *Communism.* New York, 1927
SKELTON, O. B. *Socialism, a Critical Analysis.* Boston, 1911
PIGOU, A. C. *Socialism versus Capitalism.* London, 1937
COLE, G. D. H. *The Simple Case for Socialism.* London, 1935
HOOVER, C. B. *Economic Life of Soviet Russia.* New York, 1931
RUSSELL, B. *Proposed Roads to Freedom.* New York, 1919
WEBB, B., and S. *A Constitution for the Socialist Commonwealth of Great Britain.* London, 1920
—— *Soviet Communism.* London, 1937

Index

Acceptance, 371–72
Accident, 695–96; compensation for, 695, 713–16
Advantage, law of comparative, 614–16; defined, 614
Advertising, 155–58; false, 566
Aged, pensions for, 716–17; pensions for, by railroads, 605
Agriculture, increasing costs (diminishing returns) in, 248–50, 254, 255; unprofitable field for large-scale methods, 137–38, 570; government aid to, 432–34; relative level of prices in, 180–82; economic rent in, 247–48; exempted in labor legislation, 704, 707, 708; in Russia, 774
Agricultural credit agencies, 432–34; Farm Credit Administration, 433
American Federation of Labor, 722, 723, 724, 725, 729, 734
Arbitration, 745–52; compulsory, 749–52
Assessment, special, 465–66; ordinary local, 483–85; of tangible and intangible property, 486–87
Assets, defined, 44

Balance of payments, international, 633–42, 634n.
Balance sheet, explained, 44–50; bank's, 372–75; international, of the United States, 635–36
Bank credit, 367–95; elasticity of, 392–93; importance of, 393–95; government support of, 410, 420–21; expansion of, before 1929, 415–16; liquidation of, after 1929, 416–17; for agriculture, 432–34; for real estate, 430–31; for speculation, limited, 428; qualitative or quantitative control of, 428–31
Bank holiday of March, 1933, 354, 420
Banking, principles of, 365–95; function of, 486–87; history of American, 397–404, 410; branch, 412–13, 425–26; chain, 412–13; centralized, 428–29; weaknesses of American, 413–15, 431–32; legislation concerning, since 1929, 418–36; functions of, assumed by American government, 420–21; political control of, 433, 436
Banking act, of March 9, 1933, 354, 420–21; of June, 1933, 355, 358, 422; of 1935, 422
Bank notes, 384; national, 399, 400–401, 406–407; federal reserve, 406–407; defined, 384; relation of, to deposits, 384–85; regulation of, 385–86; elasticity of, 406–407
Bank of the United States, first, 398; second, 398–99
Banks, closing of (March, 1933), 354, 420; number of, 411; private, 411; savings, 411; land, 433; central, 398–99; American local (state), 399–400, 411; mutual savings, 411; failures of, 417, 418; and RFC, 418–19; reopening of (1933), 420–21; home loan, 434–36; and foreign exchange, 625–26; coöperative, 433
See also Commercial banks; Federal reserve banks; National banks
Barter, 338–39; defined, 18
Bill of exchange, 622, 623, 624, 625; defined, 622; demand and supply of, determine rate of exchange, 631
Bill of lading, 622, 624
Bimetallism, 349; Thomas amendment permitting, 358
Bonds, explained, 42–43; and stocks compared, 43–44; and changes of price level, 43–44; payment of interest on, 42, 49–50; government, 522–23

Borrowing, 272–74; for consumption, 272; for production, 273–74; by government, 273, 518–21; limited by security, 367–68; to combat depression, 553–56
Bounties, protection by, 671–72
Budget, 523–30; double, 525–26; emergency, 525–26; explained, 523–24; general, 526–30; unbalanced, 526–29; deficit, 529
Business cycles, 531–56; seasonal, 533–34; major and minor, 535–37; long-wave, 534–35; and secular trend, 537–38; consequences of, 548–50; control of, 550–56; and unemployment, 675–77, 682; socialist explanation of, 762
Business management. *See* Entrepreneur
By-products, 59–60

Capital, a factor of production, 25–26, 27–28; defined, 25, 47n., 222n.; consumers', 26; producers', 26; accumulation of, 28–30; the means to income, 222–23; valuation of, 224–44; formation of, and time preference, 267–68; productivity of, 274–75; accumulation of, in Russia, 775–76, 778–80, 803–805
Capital stock, explained, 40; common, 41; preferred, 41–42; value of, 48–49; dividends on, 41; dividends of, 50–51; watered, 51–52; and bonds compared, 43–44
Capitalism, indictment of, 755–58; in Germany, 794–96; in Italy, 790; socialist view of, 758–62; vitality of, 806–807; case for, 805–807
Capitalist class, essential service of, 282–84
Chain stores, 38
Check, 376–80; defined, 377
Class struggle, 759–60
Clayton act, 563, 566, 567
Clearing house, 380–82
Closed shop, 732–33, 751
Coinage, 347–48; free, 347; gratuitous, 348; under bimetallism, 349; total, of silver dollars, 360, 362, 364
Collateral security, 419, 622
Collective bargaining, 604, 729–31; of child and woman labor, 693, 694–95
Combination, defined, 62; motives for, 62; formation of, 63–64; horizontal, 64; possible social advantages of, 65–66; specialization within, 66; vertical, 67; weaknesses of, 69–71; not necessarily monopoly, 558
See also Trusts
Commercial banks, 367; operations of, 367; balance sheet of, 372–75; collection between, 379–83; volume of business in, 382; number and size of, 411–12; legislation concerning, since 1929, 418–36; divorced from investment banking, 426–27
Committee on Industrial Organization, 724–25, 734
Common law doctrine, in partnerships, 36, 36n.; of negligence, 714
Communism, 767–80; campaign of, 771–72; doctrines of, 767–69; and Fascism, 781–83; organization in Russia, 774–75; program of, 772–74
Communist unions, 777
Company unions, 726–27, 728
Compensation laws, workmen's, 695, 713–16
Competing cost goods, 237–38
Competing goods, 236–37
Competition, 80–81; the basis of modern industry, 86–91; unregulated, may degenerate into monopoly, 558; defined, 145; pure, 146–48, 159–87; supply under, 146, 159–62; cutthroat, 558, 569; protects consumer, 557; imperfect, 148–58, 189–99, 202–20; imperfect, defined, 151; quality, 154–55; unfair, 564–65; railroad, 588–94; benefits of foreign, 500–503, 666–67; socialist criticism of, 755, 756–58, 764
Competitive prices, and public interest, 86; under increasing costs, 177–82; laws of, 177; under constant costs, 182–84; under decreasing costs, 184–87; in theory and practice, 239–44
Complementary goods, 237–38
Conciliation, 741–43
Consumer, influence of, on price, 96–98; protected by competition, 557; exploited by monopoly, 557

Coöperation, 72; simple, 72; by division of labor, 72–73
Corporation, importance of, 37–38; nature of, 38–40; forms of securities of, 40–44; property account of, 44–50; income account of, 52–55; large, 61–71 (*See also* Combination); value of, 242; saving by, 270–71; taxation of, 485–86, 489, 490–91
See also Holding company
Corporative state, 785–88
See also Fascism
Cost, analysis of, 135; average, 138–40, 159–61; constant, 167–68; controls supply, 132; decreasing, 135–36, 170–71, 184–87; defined, 12; of production and discounted income, 233–39; effects of past, 129–30; and expanding output, 135–37; fixed, 134–35; future, 133–34; increasing, 165–67 (*See also* Diminishing returns); limits prices, 128–29, 132–34; long-run and short-run, 162; marginal, 138–40, 193–96; and output, 140–41; of production, 127–42; relationship to price, 128–29, 132–34; schedules and curves, 141; and the sellers, 127–42; unit, and large-scale enterprise, 134–37; variable, 134–35, 161; of government, 507–13
Credit, explained, 365–67; depends on savings, 366; commercial, 366–67; function of, 393–94; speculative use of, 428; government, 520–21; control of, 429–30; and business cycle, 552–53
See also Bank credit
Credit money, 344; irredeemable, 351–53
Currency, defined, 341; based on money, 341–42; need of elastic, 392; velocity of circulation of, and value of money, 448–49; quantity of, and value of money, 446–48

Debt, war, 416–17; and fluctuating value of money, 453, 454; government, 519–21; of United States, 521–23, 529, 530
Defense, a function of government, 461, 462; expenditures for, 511, 518
Deficit, of federal government, 529; spending, 554–56
Demand, individual, 112–18; relation of marginal utility to, 112–18; total, 118–19; defined, 119; schedule of, 119–21; not desire, 121; law of, 121–22; single buyer and, 117–18, 122, 147–48; elasticity of, 123–25, 151, 182–84, 196–97; importance of, 155–57; indicators of, 197–99; stimulation of, 156–57; and supply, effect of, on price and quantity exchanged, 173–77, 180–87; and supply, laws of, 177; effect of decreased, 179; and monopoly, 189–91; substitution and, 205–208; discount partly determines, 231–33
Deposits, 374–76; a form of currency, 341–42; discount for, 372–74; defined, 374–75; demand and time, 375; relation of, to bank notes, 384–85; elasticity of, 392–93; reserves against, 407–408; insurance of, 422–25; interest on, 427–28; quantity of, and value of money, 446–48; velocity of circulation of, 448–49
Depreciation, of capital, 281; of dollar, 354–55, 650–51
Depression, 353, 391, 546–47; of 1929, 351–52, 353, 354, 416, 417, 418; finance, 525, 526–29
Devaluation, of the dollar, 354–58, 456–57, 651
Dill-Crosser railway labor act, 604
Diminishing returns, in agriculture, 248–50, 254, 255; law of, 166–67; from urban lands, 258; limit specialization, 615–16
See also Marginal cost
Discount, process of, 229–30; and laws of price, 231–33; bank, 368–69; rate of, 372–74; by federal reserve banks, 404, 428; control of rate of, 552
Discrimination in rail rates, 588–92; between goods, 588–89; local, 589–92; personal, 592
Distribution, unequal, problems of, 317–36; merit in present system of, 319; functional, 246–316, 317; personal, of income, 325–27; personal, of wealth, 322–24; equalizing, of wealth by taxation, 506; of national income, 507–11

Dividends, 41; payment of, 49–50; stock, 50–51; taxation of, 490
Division of labor, 72–79; by trades, 73; within trade, 73; efficiency of, 73–75; drawbacks of, 75–76; depends on extent of market, 76–77; regional, 77–78, 609; is coöperative, 78; international, 632, 665
Dole, 703, 705
Dollar, depreciation of, 354–55; devaluation of, 355–56, 651; stabilization of exchange value of, 355–56, 650–52; commodity or stabilized, 455–56
Domestic service, exempted in labor legislation, 704, 707, 708
Draft, 371; demand or sight, 624, 625

Economic rent, 246–65; defined, 247; in agriculture, 247–48; market measurement of, 250–52; explicit and implicit, 253–54, 307–308; contract, equivalent to producer's surplus, 252–53; of urban sites, 257–60; value of land based on, 260; relation of price and, 260–62; as private income, 263–65; taxation of, 265, 493–94
Economics, defined, 7; as a science, 8
Employment, fluctuations in, 675–77, 681–82; relation of payrolls to, 535, 536; part-time, 677
See also Unemployment
Employment offices, public, 683–85
Entrepreneur, 32–33, 34; function of, 32; incentive of, 310–12; and risk, 32, 310, 492; a "factor of production," 33
Equities, defined, 45
Erdman act, 744, 745–46
Exchange, defined, 18; dollar, 650–52; mint par of, 627, 629; indirect, 631–33; sterling, 631; purchasing power par of, 646–49
See also Foreign exchange; Bill of exchange; Trade; Exchange rate
Exchange Equalisation Account, 456, 649–50
Exchange rate, 623, 626, 627n.; stabilization of, 456–58, 649–50; defined, 626; gold points of, 628–31; reaction of, on imports and exports, 628, 638–39; relation of, to purchasing power par, 646–49
Exchange Stabilization Fund, 456–57, 651
Exports, 618–20; of United States, 618–20; duties on, prohibited by American constitution, 500; reaction of exchange rates on, 628, 638–39; financing, 621–24

Fascism, 781–807; and communism, 781–83; in Italy, 783–92; in Germany, 792–800
Fair labor standards act, 709, 710–11
Farm. *See* Agriculture
Farm Credit Administration, 433
Federal Deposit Insurance Corporation, 421, 422–25, 427
Federal National Mortgage Association, 436
Federal reserve banks, 403–409; notes of, 406–407; discounting of, 404; rediscounting by, 404; open-market operations of, 408; deposits in, 407–408, 427; reserve of, 407–408, 409; legislation concerning, since 1929, 418–36
Federal Reserve System, 403–409; founded, 402; organization of, 403; membership in, 403–404; powers of Board of, 427–30; as clearing house, 409; and stock market, 426–27; and open-market operations, 429–30
Federal Trade Commission, 563–65, 566, 568
Federal trade commission act, 563–66; amendment of 1938, 565–66
Fees, 464–65
Foreign exchange, 421, 456–57, 621–33, 643–53; under act of March, 1933, 421; governmental control of, 456–57, 649–50, 651–52
See also Exchange; Exchange rate
Freedom, economic, 81–83, 90; personal, 81; in Russia, 801–802
Free trade, 80, 503, 654, 657n., 667

Ghent system, 701–702
Glass-Steagall act, 419–20
Gold, why used as money, 345–46; as standard money, 346–50; ratios of, to silver, 349; embargoes on exports

of, 353; stock of, in United States, 362, 364; surrender of, in 1933, 354–55; clause concerning payment in, 355; policy of purchasing, 355–56; bullion used to settle international balances, 361, 629; nationalization of, 357; Treasury's profit in, through devaluation, 358; holdings of, of federal reserve banks, 362; international flow of, 628–34; hoarding of, 354–55, 361; "sterilizing," 637n.; not a peculiarly desirable import, 640–41
Gold bullion standard, 349–50
Gold certificates, 362, 364
Gold exchange standard, 350
Gold points, 628–31, 633
Gold reserve act (Jan. 30, 1934), 356–57, 651
Gold standard, 346–48; adoption of, 346–47; Great Britain and, 347, 645; war-time desertion of, 353; return to, during recovery, 353; desertion of, following 1929, 353–54; abandoned by United States, 354–55, 651; and exchange rates, 626–34; promotes equilibrium, 636–38, 641, 643
Government, functions of, 81, 460–62; borrowing by, 273, 276, 554–56; controls money, 339–40, 340–44; American, banking fluctuations assumed by, 399–410, 418–36; in industry, 462, 573–81; financing, 460–79; cost of, 460, 507–13, 529–30; revenue system of American, 513–18; budget of, 523–29; "New Deal" budget of federal, 524–29; in industry, pros and cons of, 79, 335, 461, 561–67, 571–73
Government regulation, in English and French history, 79–80; in Soviet Russia, 79; of industry, 79, 335, 461, 561–67, 571–73; of prices, 95–96; of public utilities, 95, 335, 573–79; of monopoly, 335, 559–81; of railroads, 595–607; of interest, 284, 552–53; of note issue, 393, 399–400; of banking, 399–410, 418–36; of motor carriers, 602–603; of foreign exchange, 456–58, 649–50, 650–52; of child labor, 694–95, 707–709; of woman labor, 694–95, 709; of wages, 710–13
"Granger" laws, 596, 597
"Gresham's law," 351

Holding company, 426, 558, 566, 602
Home loan banks, 434
Home Owners' Loan Corporation, 435
Hours, of labor, 696–99

Immigration to United States, and wages, 723
Imports, the aim of commerce, 614; of United States, 618–20; reaction of exchange rates on, 638–39; financing, 624–25
Income, defined, 12; net, 12–13; distinguished from wealth, 13; measurement of, 13, 13n.; present value of future, 227–29 (*See also* Time preference); capital the means to, 230–31; functional types of, 246–316; law of proportionality and, 303–305; national, 317–19; property's share of, 246–65; laborer's share of, 285–305; of American families, 319–24, 329–31; government, 462–66
See also Distribution
Income account, 52–55
Income tax, 487–94; and United States Constitution, 488; present law on, 488–90; exemption from, 488–89; distribution of, 489; rate of, 489, 491–92; elasticity of, 492–93; yield of, 493; shifting of, 493, 495–96; undistributed, 490; on corporations, 490–91
Index numbers, 439–46; defined, 443
Individual proprietorship, 34–35
Industrial conflict, 736–52; devices to prevent or settle, 741–52
Industrial unions. *See* Labor unions, industrial
Inflation, 458, 555, 644
Inheritance, 326–27; taxation of, 505–506
Injunction, 741
Insurance, compulsory, in Great Britain, 702–705; unemployment, 700–707; accident, 695, 713–16; old age, 717
Interest, 226–29, 267–84, 307–308; defined, 226; link between present

and future values, 227–29; formula, 227; regulation of, 552–53; determination of market rate of, 276–77; rates, lack of uniformity in, 279–80; rates, true and nominal, 280–81; changes in rate of, 281–82; and discount, 368–69; unearned, 368–69
Interlocking directorates, 566
Intermediate goods, 6, 26
Interstate commerce, regulation of, 597, 602–603; wages in industries in, 710
Interstate Commerce Commission, 597–98, 602, 606
Intrastate commerce, regulation of, 596–97, 603
Invention, 31, 680
Investigation, compulsory, of labor disputes, 747–48
Irredeemable money, 342–43; systems of, 351–53; and the par of exchange, 645

Joint cost goods, 238–39
Joint costs, defined, 587; in railroads, 586–87

Labor, a factor of production, 25; bargaining power of, 725–33; social stratification of, 285–86; demand for, 286–87, 679, 680, 686–87; supply of, 678; reserve of, 681; immobility of, 683; efficiency of, 697–99, 700; defensive policies of, against unemployment, 682–83; its share of income, 285–305; and railroads, 603–605; stabilization of demand for, 686–89; child, 690–91, 692–95, 707–709; woman, 690–91, 692–95, 709; hours of, 696–99; in Russia, 776–78
See also Labor unions; Unemployment; Wages
Labor exchanges, 683–84
Labor unions, restrictive tendencies of, 729–33; and wage rates, 730–31; and collective bargaining, 729–31 (*See also* Labor, bargaining power of); benefit systems of, 728–29; membership of, 725–26; company unions not included under, 721 (*See also* Company unions); federation of, 722; craft or trade, 721–23; conflict of rival types, 724–25; industrial, 723–24; conflict of, with company unions, 728; non-economic activities of, 734–35; recognition of, 731–32; and strikes, 736–52; and closed shop, 732; revolutionary, 729; legal position of American, 727–29
Laissez faire, explained, 80–81
See also Freedom, economic
Land, a factor of production, 25; a class of capital, 25–26; economic rent, the income of, 247–48; varying productivity of, 254–56; alternative uses of, 257–60; value of, based on economic rent, 260; private property in, and economic rent, 264–65; taxation of, 480–85, 493–95
Legal tender, 340–41; defined, 340; limited, 340; American, since 1933, 355
Liabilities, defined, 45; types of, 45
Liability, unlimited, 35–36; limited, 39; double, 39
Loans, 268–69, 272, 276–82; rate of interest on, 268–69, 276; nature of, 268; market for, 268–69, 277–79; supply of, 269–70; demand for, 272, 276; security for, 370; call, 390; payment of, 383–84
Lockout, 736, 749
Losses, 47; and insolvency, 47; and profits, 308–10

Marginal cost, 138–42, 159–61; curve of, 141, 160–61; related to supply curve, 168; defined, 138; and marginal revenue, 193–96
See also Diminishing returns
Marginal productivity, from capital, 274–75; from labor, 287–89, 299–300
Marginal revenue, 192–93; defined, 192; and marginal cost, 193–96
Market, defined, 143; clearing the, 175–76
Markets, prices in different, 235–36, 616, 617–18
Marx, Karl, 758–64
Mediation, 743–45
Mercantilism, 79–80, 639–40, 641, 655, 657, 658, 664

Miller-Tydings act, 573
Money, defined, 19, 338; primitive, 337; characteristics of, 337–38; the medium of exchange, 338–39; furnishes standard of value, 339; government control of, 339–40; classification of, 343; irredeemable, 342–43; representative, 342; standard, 334–43, 356; token, 344; credit, 344; redeemability of, 351–53; ideal qualities of, 345–46; value of, 227–29, 347, 348, 437–38; bimetallic, 349; history of American, 354–62; "lawful," 399; currency based on, 342; American stock of, 360, 364; American, in circulation, 364; effect of quantity on value of, 446–48; managed, 456–57; value of, defined, 437–38; value of, not constant, 438; velocity of circulation of, 448–49; forces determining value of, 449–53; quantity theory of, 449–53; effects of fluctuating value of, 453–54; stable, 455–56; in business cycle, 454–58; in communist society, 769; in Russia, 778–79
Monopolistic industry, 557–81; government regulation of, 560–68; power of, 568–69; problem of, 559–60
See also Monopoly
Monopoly, 69, 145–46, 150, 154, 217–19, 557–81; and competition, 89, 145–46, 557–59; and price, 217–19; and imperfect competition, 217–19; and luxuries, 219; and public interest, 219–20; and taxation, 495–96; railroad, 594–95; inevitability of, 568–69; forces favoring, 569–70; limits to, 570–71
Motor carriers act, 602–603
Mutual Mortgage Insurance Fund, 435–36

National banks, 399–401; notes of, 400–401; organized, 401–402; and branch banking, 412–13, 425–26; and members of Federal Reserve, 403–404; statistics of, 432; loans of, on real estate, 430–31
National Credit Corporation, 418–19
National housing act, 431, 435–36
National Industrial Conference Board, 509, 510, 693
National industrial recovery act, 689, 710, 726
National labor relations act, 727–29, 743
National Labor Relations Board, 744–45
National Mediation Board, 604
National Railway Adjustment Board, 604
National Recovery Administration, codes of, 710; settlement of disputes under, 743
Nationalism, industrial, 655, 656, 657, 672, 674; in Russia, 772; in Italy, 788–89; in Germany, 794–96
"New Deal," financial aspects of, 420–36, 507, 524–28, 529, 530
Notes, 384–85; government, 386; federal reserve, 386, 406; promissory, 367–80; proceeds of, 369–70; secured, 370; endorsed, 371
See also Bank notes; Discount

Partnership, 35–37; contrasted with corporation, 36, 36n., 38–39; not taxable on income, 489; unlimited liability of, 35–36; and monopoly, 558
Payrolls, fluctuation in, 535, 536; relation of, to employment, 535; tax on, 706, 707
Pensions, for aged, 716–17; by railroads, 605
Picketing, 741
Planned economy, in Russia, 772–74; judgment of, 802–805
Postal service, 460–61, 462, 575
Poverty, problem of, 327; causes of, and remedies, 331–36
Price, 18; defined, 19; distinguished from value, 19–20; controls production, 84–85; determined by bargaining, 86; unregulated, and public interest, 88; lag in adjustment of, 88–89, 176–77; controls consumption, 92; fluctuations of, 93–95, 437–59; consumers' influence on, 85, 96–98; marginal utility and, 110; relation of demand and, 174–75, 178–79, 180; fixed by sole buyer, 128; gov-

erned by cost, 132–34; relation of supply and, 164–65, 180; fixed by seller, 204–205; market, 143–44; equilibrium, 177; monopoly, 218–19, 557; under imperfect competition, 202–20; relation of economic rent and, 260–62; discrimination, 565, 566
Price level, general, 439–58; measured by index numbers, 439–46; forces determining, 449; relation of, to quantity of money, 450–58; fluctuations of, and income, 453–54; and business cycle, 456–58; and exchange, 636–38, 637n., 641
Price relatives, 439–46
Prices, behavior of, 93–95; fixing of, 95–96, 208; cutting of, 208–12; interrelated, 222–39; not determined by profits, 314–15; flexible, 551; equilibrium of, 617
See also Competitive prices; Price
Private property, 82; in land, and economic rent, 264–65; protection of, 334; social order founded on, 334; socialist criticism of, 755, 756–57; in Italy, 790
Production, 22–23; defined, 23–24; factors of, 24; direct, 26–27; indirect, or roundabout, 27; large-scale, 57–61, 558 (*See also* Combination); large-scale, field of, 58; large-scale, economies of, 58–61, 65–66, 569; large-scale, limit to efficiency of, 137–38, 570; control of, 79, 84–86; motive of, 83–84; evaluation of present system of, 89–91, 805–806; lag in adjustment of, 88–89; measurement of, 215–17; socialist criticism of, 757, 759, 762
See also Business cycles
Profit and loss statement, 52–55, 306–307
Profits, methods of concealing, 51–52; gross and net, 52–53; maximum, 190–94; defined, 308; and risk, 310–14; criticism of, system, 315–16; promoters', 310–12; accrue to owners, 312; speculative, 312–13
Property, defined, 14; division of, 14–15; in free persons, 16; distinguished from wealth, 16–17; evidence of, 17; valuation of, 242–44; profits an income of, 312; its share of income, 246–65; inheritance of, 326–27; taxation of, *see* Property tax
See also Private property
Property account, explained, 44–50
Property tax, 480–85; shifting and incidence of, 493–96; assessment of, 482–84, 486–87; rate of, 484–85; revenue from, 481; and intangible property, 485, 486–87; on corporations, 484–86
Prosperity, 543–45; limiting factors in, 545–46
Protection, 656–74; defined, 501; cost of, to consumer, 500–503, 658; fiscal weakness of, 504; pros and cons of, 657–69; effect of withdrawing, 667; military reasons for, 670–71; by bounties, 671–72
Public utilities, 573–75, 577–78; government regulation of, 573–78; monopoly in, 574–75; municipally owned, 575–76; railroads are, 582–83
Public works, to control business cycle, 553–56; remedy for unemployment, 688–89; program of the United States, 689–90

Railroad retirement act of 1937, 605
Railroads, decreasing costs in, 583–84, 585–86; fixed and variable costs of, 584–85; government regulation of, 595–607; monopoly in, 593–95; government aid to, 599; are public utilities, 582–83; competition among, 588–94; and labor, 603–605; costs of, 583–87; rates of, 587–92; state regulation of, 595–97; financial crisis of, 599–601; government operation of, 600; competition with, 600–601
Real estate, loans on, 430–31; taxation of, 480–85
Recession, 546–47
Reciprocal trade agreements, 673–74
Reconstruction Finance Corporation, 418–19; assists banks, 419; loans of, to railroads, 599, 602
Rediscount, by federal reserve banks, 404, 428; control of rate of, 552

Relief, by trade unions, 700–701; Ghent system of, 701–702
Rent. *See* Economic rent
Representative money, 342, 343, 346, 349–50, 351
Reserve, 387–92; function of, 387; ratio of, 388–92; problem of, 401–402; centralized, 402, 408–409; of federal reserve banks, 407–408, 409, 430; mobile, 409
Revenue system, of United States, 507–30; of states, 509–11, 512–13, 516–18
Revival, 541–43; breeds prosperity, 543–45
Risk, 32; of security holders, 43–44; profit and, 82–83, 306–16; reducing, 310
Robinson-Patman act, 566–67

Sales tax, 472, 473, 496–97, 498
Saving, 28–30; incentives to, 269–70; corporate, 270–71; in Russia, 778, 804
Savings and Loan Insurance Corporation, 435
Secular trend, 537–38
Seigniorage, defined, 359n.
Sherman anti-trust act of 1890, 561–63
Silver, why used as money, 345–46; as standard money, 346–50; ratios of, to gold, 349, 360; Thomas amendment and, 358; London Conference's agreement concerning, 359; domestic purchases of, 359–60; act of June, 1934, concerning purchase of, 359–60; nationalization of, 359–60; federal government's stock of, 360, 364; political significance of, 359
Silver certificates, United States, 359, 360
Single tax, 479
Socialism, 753–67; defined, 753–55; programs of, 753–54, 764–65; guild, 764; state, 764–65; capital formation under, 768–69; judgment of, 805–807
See also Communism
Social security act, 684–85, 706–707, 716, 717–20
Stabilization of exchange rates, 649–50
Standard of living, 328–29; minimum, in America, 329
Stock. *See* Capital stock
Strikes, 736–52; sympathetic and general, 738–39; "sit-down," 724, 739; devices to prevent or settle, 741–52; legal status of, in America, 740–41
Subsidies, ship, 672
Supply, individual, 162–63; controlled by costs and price, 162; defined, 162–163; not stock, 163–64; law of, 169; curve of, related to marginal cost curve, 160; long-run and short-run, 169–70; elasticity of, 169, 182–84; under competition, 159–71; under monopoly, 189–91; measurement of potential, 215–27; and demand, effect of, on price and quantity exchanged, 173–77, 180–87; and demand, laws of, 177; effect of increased, 177, 180, 182–84; effect of decreased, 177, 180, 182–84; discount partly determines, 231–33; of loans, 270–79
Supreme Court, and anti-trust legislation, 562–63; and railroad legislation, 596, 597, 605; and labor legislation, 706, 709, 710, 728, 750–51
Surplus, in property account, 46–47
Syndicalists, 764, 766

"Tandem" goods, 239
Tariff, 478–79, 500–504; protective, 478–79, 501, 504, 656–72; revenue from American, 500, 504; economic effects of, 500–503; prohibitive, 501–502; for revenue, 503; British, 503, 664; and wages, 660–61
Taxation, of economic rent, 265, 493–94; distribution of, 466–75; proportional, 469; progressive, 470–71, 473–74, 490–91; regressive, 471; of corporations, 485–86, 490–91; shifting and incidence of, 477–78, 493–94, 498–500; escape from, 475–76; indirect, 477–78; evasion of, 476, 492; direct, 477–78; purposes of, 478–79; double, 486, 490; history of federal, 487–88
See also Income tax; Tariff; Taxes
Taxes, land, 487–85, 493–95; death, 504–505; defined, 464; ability to pay, 468–69, 471–73, 474; "single,"

479; sales, 472, 473, 496–97, 498; consumption, 471–74, 496–97; poll, 474, 496; license, 464–65; shifting and incidence of, 493–94, 498–500; business, 474, 485–86, 490, 495–96; excise, 497–500; capitalization of, 494–95; state, 488, 516–17; receipts from, of federal government, 513–16; local, 517–18
See also Income tax; Property tax; Tariff; Taxation
Technological change, and unemployment, 678–81
Time preference, defined, 267; and capital formation, 267–68; consumption borrowing determined by, 272–73; individual, 277
See also Income, present value of future
Trade, and cost of transportation, 617; value of money varies with, 619; free, 80, 503, 654, 657n., 667; interregional, 609–20; gain from specialization and, 611–14; foreign, of United States, 618–20; "favorable" and "unfavorable" balances of, 633–42; restriction of foreign, 616–18, 656–72; free, permits revenue tariff, 503–504
See also Exports; Imports
Trade agreement act of 1934, 673
Trade unions. *See* Labor unions
Transfer, defined, 18
Transportation, 582–608; regional specialization dependent on, 593; cost of, and trade, 593; facilitated by government, 582; government regulation of, 595–607; coördination of, 601–602
Transportation act of 1920, 589–99, 743
Trusts, 560–61; laws against, 561–68

Unearned increment, 264–65
See also Economic rent
Unemployment, 675–92; seasonal, 681–82; cyclical, 682; extent of, 675–77; causes of, 677–82; technological, 678–81; remedies for, 682–89; insurance against, 700–707
See also Employment
"Unfair sales practices laws," 573
Unions. *See* Company unions; Labor unions
United States Steel Corporation, 67–69, 724
Utility, defined, 9; creation of, 22–23; forms of, 23; subjective side of, 98–100; marginal, 100–102; marginal, curve of, 102–107; marginal, relation of demand to, 109–11

Value, the common unit of measure, 18, 20; defined, 19, 438n.; distinguished from price, 19–20; par, 48; book, 48; market, 49; present, of future income, 222–23 (*See also* Time preference); the resultant of judgment, 241; rate-base, 242–43; tax, 243–44, 493–95; of land, based on economic rent, 260; money furnishes standard of, 437–39; of money, 437–39

Wages, 285–305; defined, 285; differences in, between groups, 285–86, 296–97, 301–303; demand for, 286–87; forces determining, 286–89; based on discounted marginal productivity, 287–89; supply of, 289; law of, 290, 712; sphere of bargaining in, 292–96; lag in adjustment of, 545, 685–86; of women and children, 692–94, 709; local variations in, 297–99; control of, to lessen cyclical variations, 551–52; and tariff, 660–61; flexible, diminish unemployment, 685–86; legal minimum, 710–13; hours and, 710–11; immigration and, 723; in Russia, 776–78
Wages of management, 300–303, 307–308
Wagner-Peyser act, 684
Wants, 3; divided into necessities and luxuries, 3–6; without limit, 6; utility the satisfaction of, 9; not demand, 121
Wealth, defined, 9, 11; income of, 12–13; costs of, 12; distinguished from income, 13; distinguished from property, 16–17; measurement of, 13; production of, 23–24; national, of United States, 322–24, 658–59; interrelation of, and income distribution, 325–26
See also Distribution