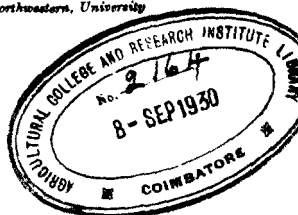


PRINCIPLES OF ECONOMICS

BY
FREDERICK S. DEIBLER, Ph.D.
Professor of Economics, Northwestern University



FIRST EDITION
FOURTH IMPRESSION

McGRAW-HILL BOOK COMPANY, INC.
NEW YORK: 370 SEVENTH AVENUE
LONDON: 6 & 8 BOUVERIE ST., E. C. 4

1929

COPYRIGHT, 1929, BY THE
MCGRAW-HILL BOOK COMPANY, INC.

PRINTED IN THE UNITED STATES OF AMERICA

PREFACE

This volume is designed as a text for an introductory course in economic principles. Its scope is confined to a consideration of the more general and basic principles of the subject. An attempt has been made to start with those facts and postulates that are fundamental in the development of the science and from these basic notions to develop an orderly movement from production, markets and exchange, value and price, money, banking and credit, to distribution and the sharing of income. Finally, the thought is brought back to the end of economic activity in the consideration of consumption and the notions of economic welfare.

In the selection and organization of the topics treated, I have followed in the main, though not in all particulars, the traditional plan of presentation. Many of these topics have been proven by long experience in the class room to be serviceable. However, instead of chapters on economic history there has been substituted material descriptive of the present structure of economic life. The chapters on applied subjects, such as labor, corporations, railroads, taxation, etc. have been omitted because it is thought that the elimination of these topics will effect an economy in the students' time. In most institutions today, advanced courses are devoted to these subjects, and if they are covered in the first course, there would seem to be an undue duplication of effort. Hence this volume is confined to a treatment of general principles, leaving the applied topics for development in the more advanced courses.

Students taking up this subject for the first time are warned against certain difficulties that they are likely to encounter. In the first place the subject is concerned with matters of every day business experience, such as wealth, value, capital, rent, interest, profits, etc. We have to employ terms that are in common usage in business, but common usage will not satisfy the demands of scientific exactness. Therefore, the economist has to assign specific limitations to the meanings of these terms, or otherwise be led into hopeless confusion and endless controversy. Not-

withstanding the greatest care in this particular, one finds that after more than a hundred years of thought and study, there still remain wide variations in the usage of terms, even among the best informed authorities in the field. Because of this difficulty students should proceed with caution, not only in their own use of terms, but in putting the usage of an author to the test.

In the second place, the phenomena which constitute the subject matter of economics are as variable as human desires. This variability makes a vigorous demand on our powers of thought. If we hope to deal scientifically with this material, we must be prepared to exercise a high degree of flexibility in our thinking and must recognize that knowledge is not absolute in character. Those whose minds crave absolute knowledge are likely to be perplexed and disappointed, when they discover the limitations to the application of the principles developed in this field.

But who can devise an absolute rule for the adjustment of a wage controversy? Or who can predict with precision the price of wheat one year hence? Such questions do not lend themselves to an absolute solution. However, it is a great gain to be able to predict the direction of change, even though we are unable to express its quantitative effects with precision. Economic principles accomplish this purpose and enable one to formulate judgments, or render decisions, that serve most practical ends. Students very early need to learn that the principles developed in this field express tendencies, but that these tendencies narrow the range of personal judgment. Even though the conclusions based upon these principles are only approximations, they are better guides for individual conduct, or for the development of social policy, than the spontaneous thoughts of untrained minds.

Lastly, most students come to the subject of economics with very definite prejudices and preconceived notions. Without conscious thought on our part most of us have accepted as fundamental certain assumptions in regard to the existing economic relations. To challenge these is to shake one's faith in the correctness of the reasoning underlying the principles themselves. On this account, the attitude of mind of the average person toward economics is different from what it is toward other fields of learning. One seldom challenges the conclusions of the physicist or the chemist on the basis of his own notions of the laws of nature.

But in economics the situation is different. Whenever questions arise that affect the existing distribution of wealth, or the prevailing notions of property, the careful student will soon find that his thinking is affected by assumptions and conclusions that have come from his previous environment. Unusual care and self mastery are required to divest oneself of these prejudices, and to deal objectively with the facts and forces studied, especially if the conclusions about them seem to run counter to those which we have previously held. Few subjects, however, afford a more fruitful field for the exercise of control over our processes of thought and for the development of independence in our thinking. Herein lies much of the educational value of the subject as a part of a student's program in college.

When economics is well presented, it will not only call for as high grade mental powers to master it, as are required by the older subjects in the college curriculum, but its educational importance will be equally significant. The ideal that should guide those presenting the subject should be to explain the economic phase of human experience and to discover and promulgate the principles that have been found to be useful as a means of accomplishing this end. The only excuse for presenting another text in this field is the hope that it may aid in this endeavor, and thus strengthen economics as a part of the college curriculum.

It would be rash to assume that there was much that is new in this volume. In its preparation I am deeply indebted to a long line of outstanding names who have made economics a science, but to acknowledge them all would be a difficult task. Many ideas have been in the process of formulation during my more than twenty years of teaching experience, and to trace all of these to their source would be next to impossible. I have given acknowledgment in the body of the text to the more evident and positive influence of an author. As is true with every teacher, I owe much to the many alert minds who, by putting puzzling questions in the class room during these years, have caused me to reflect and to search for adequate answers. Finally, I have had many suggestions and valuable criticisms from my present and former colleagues, among whom I must mention Professors Ernest H. Hahne, Vanderveer Custis, Elmo P. Hohman, Fred D. Fagg Jr., and Ralph C. Epstein; in addition, I am deeply indebted to Professor Ira B. Cross of the University

of California, who read the volume in galley proof. I wish to acknowledge my obligations to all of these persons for their kindly criticisms and helpful suggestions, and, last of all, to my wife, Julia M. Deibler, who, besides reading the manuscript and assisting in its preparation, has been a constant inspiration throughout my university experience.

FREDERICK S. DEIBLER.

NORTHWESTERN UNIVERSITY,
May, 1929.

CONTENTS

	PAGE
PREFACE	V
CHAPTER I	
THE NATURE AND SCOPE OF ECONOMICS	1-21
Economics Defined—Economics and Political Economy Distinguished—The Meaning of Economy—Economics as a Science—Economic Principles Express Tendencies—Basis for Confidence in Economic Principles—Economics a Social Science—The Relation of Economics to Other Sciences; the Physical and Natural Sciences; Psychology, Sociology, Political Science and Ethics—Business Economics—Social Economy. The Method of Treatment.	
CHAPTER II	
THE CHARACTERISTICS OF MODERN ECONOMIC SOCIETY	22-44
The Meaning of an Economic Order—Four Systems of Economy—The Autonomous, the Communitistic, the Socialistic, and the Individual Exchange Order—Individual Initiative and Private Enterprise, Division of Labor and Specialization, Cooperative in Character; Spontaneous Cooperation—Private Property and the Right to Bequeath Wealth—Right of Free Contract—The Nature and Exercise of Control, may be Unconscious or Spontaneous, or Conscious and Predetermined—The Place of the State in Modern Industry—Regulation of Private Control—Competition as an Organizing Influence, Fair Competition, Competition in Relation to Economic Theory, the Basic Assumption from which Economic Principles are Developed.	
CHAPTER III	
FUNDAMENTAL CONCEPTS	45-63
Desires as Motives—Classification of Desires, for Material Goods, for Esteem, for Action—The Meaning of the Term Goods—Definition of Utility—Classification of Goods, Free, Economic, and Public Goods—Consumer's and Producer's Goods Distinguished—The Satiability of Desires—Diminishing Intensity of Desire—Balanced Desires and Economic Welfare—Wealth and Services—Characteristics of Wealth, Utility, Scarcity, and Externality—Individual and Social	

Wealth Distinguished—Wealth and Income, Money and Real Income, Gross and Net Income—Wealth and Property—Wealth and Value—Value and Price. Caution in use of Terms.	Page

CHAPTER IV

THE NATURE OF PRODUCTION.	64-89

Production Defined as the Creation of Utility—Methods of Creating Utility, Form, Place, and Time Utility—Factors of Production, Land, Labor, and Capital—Productive and Unproductive Labor—Capital and Land Distinguished—Capital Goods and Capital Value—Division of Labor and Specialization—Limited by the Area of the Market—Forms of the Division of Labor, Occupational, Division of Processes, and Territorial—Functional Specialization—Advantages of Specialization, Develops Skill and Dexterity, Saves Time, Fosters Invention, Increases Output and Makes Possible Exact Duplication of Parts, Promotes Better Adaptation in the use of the Both Labor and Machines, makes Supervision Easier, and Tends to Reduce Fatigue—Evils of Specialization, Lessens the Importance of Trade Skill, Lessens Workers Interest in his Product, Reduces the Number of Thoroughly-trained and Experienced Laborers, Fosters Labor Turnover and Instability of Workmen, Increases Risk of Unemployment for Men with Craft Skill, and Separates Employers and Employees, thus Fostering Misunderstandings Between Them.

CHAPTER V

LOCALIZATION OF INDUSTRY AND LARGE SCALE PRODUCTION . .	90-111

Causes for Localization, Nearness to Raw Materials, to Markets, to Power, to Competent Labor, and the Momentum of an Early Start—Causes of Large-scale Production, the Area of the Market Served, Growth in Size Leads to Further Specialization, and Further use of Machine Methods, Greater Advantage in Purchasing and in Selling its Products, Utilization of By-products; Advantages that Accrue from Experimentation and the Application of Scientific Methods—Limits of Large-scale Production, Small-market Area, Limits on Managerial Ability, Friction develops between Employee and Employer—Combination and Large-scale Production—Horizontal and Vertical Combination Explained.

CHAPTER VI

THE ORGANIZATION OF PRODUCTION	112-141

The Importance of the Entrepreneur, or Director of Industry—Methods of Selection; Usually Self-appointed, may be

CONTENTS

xi

PAGES

Selected by the State or other Collective Group to Perform the Work of the Entrepreneur—Function of the Entrepreneur is to Decide what shall be Produced and how it Shall be Made—Decision of Entrepreneurs Controlling, except as Affected by Lenders of Credit, and by the Delegation of Authority to Management—Forms of Entrepreneurship, Single Proprietor, Partnership and Corporation—Advantages and Disadvantages of Each—The Importance of the Corporation in Modern Industry—The Principle of Diminishing Productivity, Diminishing Total Average, and Marginal Outputs Distinguished—The Principle Universal in its Application—Principle of Variable Proportions—Its Bearing on the Decisions of the Entrepreneur—The Principles of Least Cost and Highest Profit Combinations with their Bearing on Marginal Choices.

CHAPTER VII

RISK AND PRODUCTION 142-166

Nature of Risk—Causes of Uncertainty, Physical and Natural Hazards, Ignorance of Scientific Principles, Unpredictable Character of Human Conduct—Uncertainty Due to Time—Effect of Legislation—Methods of Dealing with Risk; Reduction or Elimination—Results of Invention and Discovery—Market Analysis and Forecasting—Consolidation of Individual Risks—Process of Compensation—Assumption of Risk—Shifting of Risk to Middle Men—to Terminal and Speculative Markets—Nature and Use of Future Contracts; Hedging Contracts; Their Purpose and Use Illustrated—Speculative Dealing and its Evils; Trading by outside Traders; by Those with Insufficient Capital, and Manipulation of price movements. Insurance as a Means of Distributing Risk—Mutual Versus Stock Companies as Risk Carriers.

CHAPTER VIII

VALUE AND UTILITY 167-207

The Place of the Market in Operation of Economic Activity—Middlemen and Market Operations—Market Defined—Two Kinds of Markets; Competitive and Monopolistic—Competition Versus Monopoly in Economic Analysis—Advantages in Explaining Market Phenomena from the Assumption of Competitive Conditions—Value Defined—Value and Evaluation Distinguished—Value and Utility—Value and Price—Demand and Supply as Value Determinants—Not Self-operative Forces—The Nature of Demand—Individual Demand Schedule—Market Demand Schedule—Diminishing Marginal, and Total Utility—Inequality of Income and

the Explanation of Value—Caution in the Use of Marginal Utility as the Cause of Value—Total Utility and Consumer's Surplus—The Nature of Supply, the Supply Schedule—Equilibrium of Demand and Supply—Only one Price in a Competitive Market—Two Meanings of an Increase of Demand and of an Increase in Supply—Elastic and Inelastic Demand.

CHAPTER IX

COST OF PRODUCTION AND SUPPLY. 207-240

Scarcity and Value—Four Causes of Scarcity; "Niggardliness" of Nature; Expansion of Desires; Cost of Production; and Monopoly—Nature of Costs, Psychic Costs and Expenses of Production—Kinds of Costs; Pain costs, Fatigue costs; Opportunity costs—Diminishing Productivity and costs—Items Entering into Expenses of Production—Labor Costs, Material Costs, Overhead Costs, Selling Costs—Interest as a Cost—Rent as a Cost—Managerial Salary as a Cost—Fixed Supply and Value, Importance of this Influence in Determining Value—Reproducible Supply and Value; under Constant Costs, and Increasing Costs—Temporary and Permanent Differences in Costs—The Representative Firm in Relation to the Explanation of Value—Bulk Line Costs—Increasing Costs Found Mainly in the Extractive Industries—Decreasing Costs—Internal and External Economies—Normal Value—Joint Costs and Joint Demand—Speculation and Value.

CHAPTER X

MONOPOLY AND VALUE. 241-258

Summary of Value Theory in a Competitive Market—Obstacles to Perfect Competition—Monopoly, its Nature and Influence—Kinds of Monopoly, Social and Natural—Definition and Illustration of Monopoly Value, under Constant, Increasing, and Decreasing Costs—Class Price—Monopoly Price as a High Price—The "Corner" as a Monopoly.

CHAPTER XI

MONEY AND THE MECHANISM OF EXCHANGE 259-288

Division of Labor Necessitates the Use of Money—Disadvantages of Barter—Definition of Money—Functions of Money—Characteristics of a Good Money Material; Acceptability, Uniformity, Divisibility, Cognizability, and Convenience—The Monetary System—the Choice of a Standard—The Coinage System—Kinds of Money; Metallic, Paper, Irredeemable Paper, Representative Paper, Bank Notes, and Subsidiary Coins—Legal Tender Quality of Money—The

CONTENTS

xiii

Page

Value of Money Expressed Only in Terms of Commodity Prices—Value of Standard Money Varies in Inverse Ratio with its Quantity and Rapidity of its Circulation—The Equation of Exchange—Bimetallism, its Effect on the Value of Money—"The Limping" Standard—Irredeemable Paper Money and its Effect on the Value of Standard Money.

CHAPTER XII

BANK CREDIT AND THE CIRCULATING MEDIUM 289-304

Kinds of Credit—Banks and Bank Credit—Kinds of Banking Institutions; Investment, Savings, and Commercial Banks—Financial Functions of Insurance and Trust Companies—Two Forms of Bank Credit, Bank Notes, and Deposit Credit—Discount Explained—The Bank Statement and Bank Credit—Clearing Houses and the Check-off System—Bank Credit and the Value of Money—Relation of the Volume of Gold to Bank Credit.

CHAPTER XIII

THE BANKING SYSTEM OF THE UNITED STATES 305-324

First and Second United States Banks—Period of Private Banking—Conditions under State and Private Banks—The National Banking System; Established a Unified Note—Strength and Weakness of the System, the Note Holder was Amply Protected, Note Issue Inelastic; Surplus Reserves Flowed to Speculative Markets; Lack of Cooperation Among Banks Weakened their Strength at Critical Periods—Proposed Remedies—The Federal Reserve System—Centralizes Reserves of the Country; Provides for Cooperation through Rediscounting; the Note Issues now Elastic—The Reserve Banks Depositories of Public Funds—May Engage in Open Market Operations—This Feature an Influence in Stabilizing the Value of Money.

CHAPTER XIV

PRICE MOVEMENTS, FINANCIAL PANICS, AND THE BUSINESS CYCLE. 325-358

Price Changes due to Two Causes, Changes in the Intensity of Desires, and Fluctuations in the Value of Money—Index Numbers as a Means of Measuring the Changing Value of Money—Kinds of Index Numbers, Simple Arithmetic Average, "Weighted" Average, the Geometric Mean, the Median, and the Mode—Fluctuations in Prices as Indicated by Index Numbers—Fluctuations in Volume of Credit Money—The Business Cycle, Four Phases of the Cycle; Recovery, Prosperity, the Crisis, Financial Panics—Relation of Banks to

Financial Panics—the Industrial Depression—Causes of the Business Cycle—Social and Economic Consequences of Cyclical Movements—The Stabilization of the Dollar—Forecasting and Regularization of Industry.

Page

CHAPTER XV

INTERNATIONAL TRADE AND SETTLEMENTS 359-383

Specialization in Relation to Trade—Differences in Productive Power the Basis of Trade—Domestic and International Trade Alike in this Respect—International Trade; the Visible Items, the Invisible Items; Extension of Loans, the Payment of Interest, Immigrant Remittances, Tourist Expenditures, Freight and Other Charges for Services Rendered, Insurance Charges—Balance of International Payments of United States in 1926—Absolute and Comparative Advantages in Trade—Some Implications of Comparative Advantages—International Dependence—International Settlements—Par of Exchange—Bill of Exchange; a Commercial Bill, Sight, and Time Bills—Gold Points—International Balances—Three-cornered Exchange—Arbitrage—Gold Settlements—Favorable Balance of Trade.

CHAPTER XVI

PROTECTION AND FREE TRADE. 384-401

The Meaning of Free Trade; of Protection—Incidental Protection—Tariff for Revenue, Principles that Govern the Selection of Commodities to be Taxed—Arguments for Protection; Infant Industries, Home Market, Diversified Industries, Labor Arguments, Equalizing Costs, Political Arguments—Difficulty in Removing a Protective Duty—Incidence of a Tariff; Consumers Generally Pay the Tariff, One Possible Exception—By-products of Protective Duties.

CHAPTER XVII

GENERAL THEORY OF DISTRIBUTION 402-415

Distribution Defined—Functional and Personal Distribution Distinguished—Evaluation and Distribution—Diminishing Productivity and the Law of Variable Proportions in Relation to Distribution—The Marginal Productivity Theory of Distribution—Economic and Contractual Shares Distinguished—Effect of Legal Institutions on Distribution—Influence of the Rights of Property and Contract—General Conclusions and Implications of the Theory.

CONTENTS

XV

CHAPTER XVIII

RENT	PAGE 416-434

Rent Defined—Rent and Capital Expenditures Distinguished—Three Uses of the Term Rent—The Nature of the Services of Land; Location and Rent, Method of Measuring the Services of Land—Diminishing Productivity and Rent; Relation to Marginal Land—Rent in Terms of Value—A Rent Formula—The Intensive Margin—The Rent Principle Applied to Urban Land—The Completed Rent Formula; Caution in its Use—Rent as Producer's Surplus—Relation of Rent to the Value of Land—Rent of Mines; Depletion Cost.

CHAPTER XIX

RENT (CONT'D.)	435-452

Growth of Population and Rent—Improvements in Cultivation and Rent—Rent and Property in Land—Rent as an Unearned Increment; Meaning of an Earned Income—Land as a Private Investment—Ripening Costs of Land Utilization—Rent in Relation to Price—Rent as a Marginal Cost—Taxation and Rent—The Single Tax and Rent.

CHAPTER XX

CAPITAL AND INTEREST	453-477

Interest Defined—The Nature of Capital—Social and Individual Capital Distinguished—Economic and Contractual Interest—The Three Major Issues—First, Why is it Necessary to Pay Interest? The Problem of the Formation of Capital and Why it is Scarce—Second, How is it Possible to Pay Interest? The Productivity of Capital—Finally, the Determination of the Rate—The Meaning and Place of Time Preference—Corporate Savings in Relation to the Rate—The Demand for Capital—The Rate on Consumption Loans—Quantity of Money and Interest Rates—The Interest Rate and Capitalization—Interest as a Cost.

CHAPTER XXI

WAGES	478-503

Definition of Wages—Two Problems; How Measure the Services of Labor?—How Determine the Value of those Services? Differences in Labor, Non-competing Groups—Diminishing and Marginal Productivity as Applied to Labor—Economic and Contractual Wages—Application of the Marginal Principle—Wages of Management; Method of Testing the same—The Meaning of Demand for Labor—Supply of Labor—Total Population and Supply—Growth of

Population and the Malthusian Principle—Immigration and Supply—Subsistence Theory—Wages-Fund Theory—Wages *as Advances*—The Discounted Marginal Productivity Theory, and a Criticism—Organization and Wages—Standard of Living and Wages—The Saving Wage—Wages as a Source of Demand.

CHAPTER XXII

PROFITS 504-522

Summary and Restatement of General Assumptions—The Entrepreneur as a Profit Taker—The Nature of Profits; Ordinary Profits, Business Profits, Pure Profits—Wages of Management as Profits—Rent, Interest, and Profits—Source of Pure Profits—Profits and Losses—Risk and Uncertainty in Relation to Profits—Profits as a Functional Share—The Profit Motive in Industry—Production for Profits versus Production for Service—Monopoly Profits—Competition versus Regulation—The Concept of a Fair Return.

CHAPTER XXIII

GENERAL SUMMARY 523-541

Restatement of the Characteristics of the Present Economic Order—An Indication of the Extent to which Economic Principles Depend upon the Structure of Society—Review of the Main Topics Treated—Relation of the Elementary Principles to some of the Important Fields of Specialization, as Business Organization, Marketing, Finance, Money and Credit, Personnel and Industrial Relations, Accounting and Statistics, Social Control and Regulation, Associations and Combinations, Public Finance and Taxation, Public Ownership, and the Readjustment of the Social Structure.

INDEX 543

PRINCIPLES OF ECONOMICS

CHAPTER I

THE NATURE AND SCOPE OF ECONOMICS

Economics Defined.—The subject of economics treats of the processes by which man gets a living. A little observation will reveal to the thoughtful person that the major part of the time and energy of mankind is devoted to this task. Let anyone ask himself how much time his business or profession requires, and why it is necessary to spend so much time in this way, and the truth will be borne in on him that the problem of getting a living is one of vital concern to every member of the human family. Economics undertakes to study this problem, not in all of its aspects, but in those that have to do with the buying and selling of goods and services. This phase of the problem is studied both in its individual and its social aspects. The phenomena which appear in connection with the production and flow of the necessities, comforts, and luxuries of life are brought under observation, and an attempt is made to formulate principles that will aid in explaining that part of human conduct that is devoted to this task. All business activity, whether within the field of industry or of commerce, is studied not only for the purpose of knowing and understanding it, but with the view of developing guides for its conscious direction. While it is not primarily the function of the economist to work out either plans of social reform or methods of improving the conduct of business, yet, before suggestions can be intelligently made for either of these purposes, it is desirable to know the fundamental nature of the problems involved. Economics undertakes to study this whole realm of human experience and may be defined as the *social science which deals with those aspects of the wealth-getting and wealth-using activities of man that are concerned with the production, purchase, and sale of goods and services.*¹

¹ ELT, R. T., "Outlines of Economics," 4th Ed., p. 2.

Economics and Political Economy Distinguished.—Many of the earlier writers treated this subject under the title of *Political Economy*.¹ This term conveyed fairly accurately the limits of the subject as it was then conceived. Interest in this field was first developed in connection with the problem of finding and of administering an income for the king's household. In the initial stages, this was primarily a problem of private economy, but in the course of time the king's revenues came to be synonymous with the revenues of the state from which the expenses of operating the government had to be paid. For this reason, the first topics to command the attention of the early writers were taxation, finance, money, trade, and such other questions as were thought to be intimately connected with the welfare of the state. When Adam Smith, who is frequently referred to as "the father of political economy," published his great work, he called it "An Inquiry into the Nature and Causes of the Wealth of Nations."² This title represents the purpose and scope of the subject as it was conceived at that time. Since the economy of the state occupied so much of the attention of the early writers, it was appropriate that the term "political economy" should have been used by them to denote the scope of their work.³

In more recent years the scope of the subject has been broadened to include the economy of individuals as well as that of the state. It has since been seen that the welfare of the state depends very largely upon the welfare of the individuals who compose it and, therefore, it becomes important to consider the problems affecting private income and private expenditure, or, in other words, private economy. There are several possible limits to the range of such a study. If the unit chosen is the family, we would have a science devoted to family economy, somewhat akin to the fields of domestic science and home economics. If

¹ Antoine de Monchrétien in a book published in 1615, entitled "*Le Traité de l'Oeconomie politique*," first used the expression *political economy*. This was an appropriate expression as it coincided with the historical development of modern states.

² Smith published the first comprehensive and formal treatise on the subject, the first edition being published in 1776. Smith did not use the term, "political economy," it is surmised, because Sir James Steuart had published his work a few years earlier, and had included this phrase in his title. The discussions of these early writers were concerned very largely with the problems of their day, as the titles to their works will reveal.

³ A modern text has undertaken to stress this point of view. Cf. CARYER, T. N., "Principles of National Economy."

attention is centered upon the business firm and its activities, we would have a science dealing with the economy of an individual business, a field that is now claimed by the familiarly used term, "business economics." In this field there has already developed a large amount of specialization as, for instance, business organization, factory organization and management, marketing and distribution, advertising, personnel administration and industrial relations, finance, accounting, etc. It is within the range of these subjects that the schools of commerce are beginning to develop their professional specialization. A science devoted to the internal organization and operation of a business in its acquisitive aspects might well be called "business economics."

But if attention is directed to the state, using this term in its generic sense to include all of the political units of government, we would have a science that dealt with national economy, i.e., with the income and expenditure of the state as here used. This is the field that is now commonly known as public finance and taxation. Finally, if emphasis is placed upon the general principles of wealth getting and wealth using, irrespective of their application to a specific field, we are dealing with social economy or the economy of the social group. The term "economics," because it is more general, has come to be used to include all of the principles that have been formulated in this more general field of wealth-getting and wealth-using activities of man. Hence, the term "economics" has been substituted for "political economy" to express this change in the content of the field studied. Economics is the more general term and implies the science of economy, whether of the individual, the business firm, or the state.

The Meaning of Economy.—The term "economics" comes from two Greek words which mean "household administration." The central idea of the subject is the administration of resources, whether by the individual, the household, the business firm, the state, or by human society. To economize is to make the most effective use of resources—to use them sparingly and in such a manner as to yield the maximum benefits. To the individual, the problem of economy in getting a living involves many choices. First, he must decide how his time and his energy will be used. The disposition which he makes of these will determine the amount of his income. Then, after an income has been received,

he must decide what use he will make of it. He may spend it for goods which he expects to consume at once, or he may save it and invest it for the purpose of increasing his income at some future date. In either case, choices must be made by the individual and his economic welfare will depend upon his decisions. But the problem of income and expenditure is not confined to individuals. In our day, many services are received through the state, so that the choices made in the use of the energy of the nation become a matter of vital concern to every citizen. The resources of a nation may be used up rapidly by the present generation or they may be used more sparingly and preserved for the future. Fundamentally, economy means the choices made either by an individual, a business firm, the state, or other collective group in the use of time and energy in securing an income, as well as in the disposition of that income when received. In other words, the administration of the resources of the individual, the household, the business firm, or the state, is the basic idea of economy.¹

Economics as a Science.—In the definition given, it is asserted that economics is a science. Doubt has sometimes been expressed as to the possibility of developing a science in this field because so many of the phenomena are under the control of the human will and, therefore, it is argued, cannot be made to conform to principle. How, it is asked, can the action of an individual be reduced to principle, when that individual is free to make choices in accordance with every whim or wish? Whether economics is, or is not, a science is largely a matter of definition.² Running through all uses of the term "science" there is one uniform element, namely, the attempt to discover causal, or sequential, relations among phenomena. Likeness appears also in the methods used, which consist, broadly speaking, of observation, classification, and generalisation. So far, then, as men seek the cause and effect relations among the economic phenomena of life, they are proceeding in the same manner as students in other fields of learning. Their methods are scientific and their

¹ CARVER, T. N. "Principles of National Economy," p. 2, has made economics consist largely of the problems of income and expenditure and has reduced these to a convenient form to show the broad relations existing between the different parts of the subject.

² According to the Standard Dictionary, science means knowledge gained by exact observation and correct thinking, especially as methodically formulated and arranged in a rational system.

classified observations and generalizations constitute a science. The purpose of such a study is to find a body of principles that will aid in explaining the day to day experiences of our economic life.

Economic Principles Express Tendencies.—One caution, however, should be noted at the outset. The principles developed in the field of the social sciences are not as absolute or as universal in their application as those in the physical and biological sciences. While it is not possible to predict results with the same degree of assurance in the former as in the latter, the difference is one of degree rather than of kind. Men regularly draw conclusions which are based on economic principles, and formulate business and social policies as the result of the acceptance of these principles. Ask any business man what will happen to the price of wheat if the next crop should fall 20 per cent short of the normal yield, and he will promptly respond that the price will rise. He has generalized and his answer rests upon an accepted economic principle. He is not able to prophesy precisely the quantitative change of price, but it is significant that he can predict the direction of that change. However, the more accurate his information concerning the existing supply and the course of per capita consumption in the past, and the more skill he has in making and interpreting observations, the more nearly will he be able to approximate the exact price change. The conclusion, then, is that generalizations are used in economics in the same way as they are in other sciences, but instead of calling them laws they are called principles. A principle in this field expresses a tendency or causal relation observable among economic phenomena.¹

Basis for Confidence in Economic Principles.—Taking up again the doubt expressed concerning the possibility of developing a science in this field because of the volitional character of human conduct, it may be asked, What basis is there for confidence in the principles formulated, if the individuals are free to choose and to direct their conduct according to purely personal and individual preferences, or even by caprice? In reply three general propositions may be laid down which give confidence in the principles developed in this subject. First, some economic principles rest upon the uniformities found in the physiological organization

¹ Cf. Gidd, C., "Principles of Political Economy," pp. 7-12, for a different point of view in respect to the existence of "laws" in the field of economics.

of man. Take, for instance, the experience of any person in consuming some commodity, say apples. The individual apples may be identical in size, color, and all the qualities that make them pleasing to the taste, but every one knows that by consuming one after another of these apples within a short period of time the desire for apples will become gratified. If the individual continues to eat them, instead of giving satisfaction they may give positive pain. This illustration is a matter of common experience and what is true in connection with the consumption of apples is also true in a greater or less degree in the consumption of any other commodity. The choices of the individual are affected by this physiological organization and the uniformity of human conduct resulting from its existence has been observed and has led to the formulation of a principle, to be discussed later, which is of great economic significance and is of wide application.

In the second place, some economic principles rest on the uniformities found in nature. Take an illustration familiar to every farmer. If he should undertake to increase the acreage yield of some crop like corn, he might find that by improved methods of cultivation and more effective fertilization he could increase the absolute number of bushels grown per acre, but soon the increase would not keep pace with the added labor and cost necessary to secure it. Here the uniformity is found in the response of nature to the productive efforts of man, and the observation of this fact has given rise to an economic principle of great significance and of wide application. Since, as stated, this principle rests upon the uniformities found in nature, it is not subject to the control of the human will.

Lastly, even within the realm of human choice, uniformities are found that are explainable upon the assumption that human conduct is guided mainly by reason rather than by caprice. Men do not ordinarily choose to do the foolish thing and their actions can usually be explained on a rational basis. Other things being equal, they prefer to buy at a low rather than at a high price, if the quality of the goods purchased is identical. Uniformity in reason implies some uniformity in human conduct, and while it is impossible to predict with precision what an individual will do, it is possible to predict with a reasonable degree of accuracy what men in the aggregate will do. No one can predict how many pounds of meat, or sugar, or pairs of shoes an

individual will buy in the course of a year, yet, on the basis of past experience, per capita consumption of any one of these articles can be estimated and, on the basis of this estimate, total annual consumption approximated. The action of the individual in such cases becomes swallowed up, as it were, in mass action, and mass action can be reduced to fairly definite rules by the law of averages.

Furthermore, underlying the uniformity of reason is the assumption that most persons in making choices will select those commodities which they think at that time will yield the most beneficial results, *i.e.*, those which are expected to gratify the keenest desire. The individual may make a mistake in his choices and later find that some other article was better suited to his needs. He may have been deceived and buy a good that will not serve his needs at all. Or, under the influence of skilful advertising or sales talk, he may be induced to buy an article for which he has no real need. While buying of this kind is common, nevertheless, experience tends to lead to more careful and discriminating choices—choices based upon an assumed or recognized beneficial result. Education and training tend to bring the choices of the individual into line with enlightened self-interest, and the more completely individual action is guided by self-interest the more accurately is it possible to predict individual conduct. In so far, then, as human action is guided by self-interest and rational choices, it is possible to reduce such action to a uniformity that will constitute a basis for the statement of a principle. Whether the uniformity which underlies the principles in this field rests upon conditions found in nature or in man's mental or physiological organization, there is ample evidence that human conduct is predictable within limits, and may in this sense be said to be guided by principles.¹

* **Economics a Social Science.**—In the definition of economics given above, a second idea appears, namely, that the subject is a social science. This distinction is of importance and signifies that the subject treats of the social aspect of economic life. It is possible to conceive of an existence in which individual, or more likely family, life is lived in isolation and hence is self-sufficient. The disposition which individuals so situated make of their time and energy could be analyzed and reduced to a scientific basis; their choices could be reduced to rules or princi-

¹ CHAPMAN, S. J., "Outlines of Political Economy," pp. 10-12.

ples. Such principles would constitute a science of individual economy, but, since this form of existence is not typical of modern economic life, these rules would have very limited application today, even though they might accurately portray the past experiences of the family when it was largely self-sufficient. In our day most persons are a part of a great socio-economic group in which each individual is mutually dependent upon every other. Most persons today are engaged in some form of economic activity that is commonly known as business. To the owner of a business, its conduct appears to be a strictly private undertaking, but in reality all business is part of a great social process.

Every business concern or establishment has both a social and a private, or individual, side. It has both external and internal relations. Its internal problems are concerned very largely with questions of organization and administration. How may goods be produced most efficiently and put on the market at the lowest possible cost? Where secure the raw material? How lay out the plant and equipment to the best advantage? How maintain an efficient group of workmen, keep them satisfied as to wages, and at the same time secure from them effective workmanship? How shall the products be marketed so that they will yield a profit, thus enabling the owner to pay all operating expenses and maintain his standing with the bankers as a basis for credit? Questions of this character affecting the internal operation of a business press upon the business executive for decision. The success of the venture depends very largely upon his initiative, energy, and foresight in rendering these decisions. The purpose of the business, viewed from the standpoint of the individual, is to "make money" for its owners, and its policies are shaped as far as possible to increase their income. From this point of view business is purely acquisitive in nature and is operated for private purposes.

Every industrial concern is a unit in a great economic structure that is social in character. No establishment, however large, is more than a step in this great social process, the object of which is to bring the raw materials of nature into a form in which they may be used by man. In performing this task, each business unit is mutually dependent upon every other. Therefore, even from the point of view of private gain, the success of a business is greatly affected by external forces beyond the control

of the managers or owners. These external forces and relations constitute the social aspects of business. Economics treats of these external forces and influences, and endeavors to reduce them to principles, and is, therefore, in this sense a social science. It deals with the social aspects of economic activity. When its principles are intelligently applied, they are an aid to the executive in shaping the internal policies of his business in conformity with the external conditions that surround it. By so doing, he promotes general welfare at the same time that he is increasing the private gain of the owners. It is not the purpose of the science to teach individuals how they may become rich. The object is to explain the causal relations found in this phase of human conduct. Nevertheless, the executive who best understands these principles is in a strategic position to adapt his policies to the external conditions and thus promote his own private gain. In this way, the individual business man, while working for himself, is making a contribution to the economic welfare of the social group of which he is a part. By thus treating of the social aspects of business activity, economics becomes one of the social sciences.

Two additional ideas are present in the definition given above, namely, that the phenomena studied are concerned with the wealth-getting and wealth-using activities of man. Wealth-getting, in its social aspects, means production in all of its varied forms. It includes the production of raw materials, the manufacture of these into a usable form, the transporting of the finished goods to the markets, and the merchandising of these to the final consumer. All of these processes, which in actual business may be subdivided into a multitude of separate business establishments or units, are brought under observation and constitute the phenomena of wealth-getting. Wealth-using means the consumption of the things produced. Consumption of wealth is the ultimate aim of all production. The finding of the means to satisfy human wants is the objective of industry when viewed from the social standpoint. The fact of human wants gives rise to industry from which the means is found to gratify these wants. This priority in causal relations gives to consumption an important place in the science of economics, but the further development of this phase of the subject will be postponed to a later place in this work.

The Relation of Economics to Other Sciences.—In the definition given above, economics is called a social science. In taking up this subject for the first time, the reader will doubtless find it helpful to know, in a general way at least, the broad outlines of the field covered with some of its relations to other sciences. Remembering that economics deals with the wealth-getting and wealth-using activities of man, he should recognize at once that these two lines of human action are dependent in greater or less degree upon practically every other science. The physical and natural sciences make a very definite contribution both to the getting and to the using of wealth. For instance, modern engineering, which is so vital to the industry of today, is based upon principles that were first worked out by the physicist and the mathematician. In like manner, industrial chemistry, which has made, and is making, large contributions to modern production, is dependent upon the discoveries in the field of "pure chemistry" where the elements of nature are isolated and the laws of chemical combinations are discovered. Similar instances of contributions to the production, or to the use, of wealth might be cited from nearly all of the natural sciences. Economics begins with the results of the application of these principles and endeavors to work out rules that will explain human conduct toward the wealth produced. The phenomena with which it deals are quite distinct from those that occupy the attention of other scientists. "Just as the geometer may consider the dimensions of bodies apart from their physical properties, and the physicist their physical properties apart from their chemical constitution," so the economist may study their relations to man and their capacity to gratify his desires.¹ While the distinction between economics and the natural sciences is clear enough, the student should recognize that each of these sciences makes its contribution to the special field occupied by economics.

A more definite scientific relation exists between economics and psychology than between economics and the natural and biological sciences. Desires and motives are psychological phenomena, the outward manifestations of which appear as stimuli to human actions, which in turn become economic phenomena. The mere desire for a good is a psychological fact, but when that desire expresses itself in the form of a purchase of a

¹ KERNAN, J. N., "Scope and Method of Political Economy," p. 115.

good, then the psychological fact is translated into an economic phenomenon. Of course, no one can measure directly the mental state of another person. No one can determine with precision whether A will get more satisfaction from a pair of shoes than will B, yet there are objective evidences that constitute an indirect basis for such comparisons. When an individual debates as to whether he should make one purchase rather than another, or whether he should buy or save, his decision constitutes an outward manifestation of his mental state that can be measured by objective standards. If he decides to put a dollar in the savings bank rather than spend it, the only logical conclusion that can be drawn from his action is that he regards a future use of the dollar of greater significance than its present use. Likewise, when we find two persons, equally situated as to their ability to buy a good, just willing to pay a definite sum for a unit of it, we can conclude that the two individuals get about the same degree of satisfaction from its consumption. These outward manifestations are the data with which economics deals. It assumes a psychological explanation of the choices of mankind, but, in fact, is concerned more with the results of choice than with their psychological explanation.

It is at this point, however, that psychology and economics come very close together. The economist presupposes psychology just as he presupposes the physical sciences, and the natural starting point for his inquiries is a consideration of the motives by which individuals are usually influenced in their economic relations. These motives are psychological facts of great economic importance, but the economist assumes them and does not attempt to establish them. He does not seek to explain or analyze them; nor does he investigate all of the consequences to which they lead. His attitude is well put by Cairnes, who observes that:

Rent is a complex phenomenon, arising from the play of human interests when brought into contact with the actual physical conditions of the soil. The political economist does not attempt to explain the physical laws on which the qualities of the soil depend; and no more does he undertake to analyze the nature of those feelings of self-interest in the minds of the landlord and tenant which regulate the terms of the bargain. He regards them both as facts, not to be analyzed and explained, but to be ascertained and taken account of; not as subject matter, but as the basis of his reasoning. If further information be

desired, recourse must be had to other sciences: the physical fact he *hands over to the chemist or the physiologist*; the mental to the *psychologist or the ethical scholar*.¹

Put in another way, the economist builds upon the existence of motives without attempting to establish a psychological explanation of them, which he regards as belonging to the province of the psychologist.

A much closer relation exists between economics, sociology, political science, and ethics than between economics and the other sciences treated. These last mentioned subjects are social sciences and it has been held by some authors that there is no justification for treating them as separate sciences and that there should be only one general science of society. It is argued, for instance, that it is "irrational to attempt an economic or industrial analysis of society, apart from its intellectual, moral, and political analysis." These phenomena are regarded as interrelated, and even though economic, political, and ethical facts can be isolated and studied separately, it is thought to be futile to attempt the development of separate sciences on account of the interrelation of the several groups of facts. Those who hold to this view would make sociology the broad, inclusive science that would embrace all phases of social research. The consensus of opinion today, however, is in favor of developing a science in each of these specialized fields. Greater results can be attained by developing along specialized lines than by attempting to take account of all influences that affect social phenomena.

In this scheme of scientific treatment economics deals, as has been indicated, with the wealth-getting and wealth-using activities of man, and all of the phenomena arising out of these lines of human action fall within the scope of its inquiry. Recognition may be taken of the influence of ethical and political forces, but these forces are studied only in so far as they affect the economic actions of mankind. Sociology, on the other hand, treats of the whole range of social processes and all of the facts and influences that affect the life of a social group. It is true that the sociologist takes cognizance of wealth and other economic facts as well as of political and ethical facts that have a bearing upon the phenomena of group life. He is likely to be concerned, also, with the results of social action and is interested in questions of social

¹ CAIRNES, J. E., "Logical Method of Political Economy," pp. 37, 38. I have drawn heavily from J. N. Keynes in the above treatment.

policy and social betterment. He recognises the significance of wealth as a necessary basis for the support of the whole social structure, but he does not attempt to explain value and price. The economist, in his scientific capacity, gives little or no attention to the question how economic relations ought to be, as he is occupied in explaining what exists. He is seeking to find principles that will explain economic phenomena and, while he may hope that his discoveries will serve mankind and become rules for guiding human conduct, yet his first concern is to explain existing economic relations. Sociology, then, as here conceived, is a general science of society, while economics is confined to an analysis of the causes affecting the production and use of wealth.

The field of political science is easily differentiated from that of economics. The former treats of all forms of government and political action while the latter deals only with wealth and its uses. Political conditions and stability of governments greatly affect the production and use of wealth. For instance, the laws of property and contract govern the economic relations existing between man and man, and all economic activity is conducted within the legal framework of the state. On the other hand, the development of law is a long and gradual social process that tends to conform to the social needs of any period. While the fundamental conceptions of law and order may remain the same, the outward manifestations are modified by a changing economic life. The economic needs of a people cannot long be bound by legal forms. The law may hold economic forces in check for a time or change the direction of growth, but, eventually, the needs of a people will cause a modification of the legal forms so as to secure a better adaptation to the existing group interests. For instance, the case of the railways may be cited. With the extension of the roads into different states, it became necessary to develop some form of general regulatory body that could deal with such questions as rate discrimination as between places, and other similar problems. The Interstate Commerce Commission was set up and its powers have been gradually extended in order to express the thought of the community on these issues. Economics is concerned with transportation and the effect of regulation on the production and movement of goods, whereas political science and the law are concerned with the authority of the regulatory body. It is evident, therefore, that there are mutual relations between economics and political science and

that the two fields may come together at many points, especially where questions of social policy are involved, yet it is clearly apparent that, as scientific fields of learning, the one specializes on the forms of government and the political action of mankind, while the other deals with the production and use of wealth.

The distinction between economics and ethics can be suggested by the relations existing between what is and what ought to be. Economics is concerned primarily with the explanation of what is, while ethics deals with the results of human conduct measured by some standard of group welfare. In ethics, an evaluation of social phenomena is made and a judgment reached as to what conditions will contribute most to the well-being of those making-up the social group. It is not always an easy task to keep these two types of reasoning separate. In practical economic problems, such as that of "fair rates" for public utilities, or "fair wages," or other similar problems, the question of what ought to be keeps constantly pressing for recognition. But as soon as this question of "fairness" is raised, an ethical standard of measurement is implied. The economist frequently has to deal with data that involve human relations similar to those suggested, and it is difficult to treat them quite apart from the effect of the conclusions resulting from his analysis. In fact, economic theory should always be tested by the light it throws upon human conduct economically considered. The economist should always endeavor to direct his investigations into those lines that promise aid in solving questions of practical, social significance and should strive to make the results of his investigations serve some human purpose. It would be extremely difficult to find an excuse for such investigations upon any other basis.

But while the ultimate objective for research in this field is the hope of finding rules that will aid in guiding human action, the immediate purpose should be the collection, classification, and interpretation of facts. The discovery of scientific relations existing among complex economic phenomena is the first step in the development of rules and principles of human conduct. No scientist has much claim for recognition and confidence who does not postpone his judgment as to these matters until after he has first examined the facts and discovered the existing uniformities affecting that conduct. The function of the economist is, primarily, the discovery of principles. When he begins to pass judgment on questions of social policy, he enters

the realm of social ethics or what the economic conditions ought to be. It is true that many of the standards for ethical judgments are derived from the analysis and conclusions of the economist.

Take for example the case of a "fair rate" suggested above. The ideal of "fairness" in such cases is obtained from the conditions that are assumed to exist in a perfect market. In such a market it is assumed that there is perfect competition and perfect mobility of the factors of production, so that as soon as any factor may receive a larger share of income in one use than in another, it will immediately flow into the more remunerative one. Under such circumstances each factor will receive an income that accurately measures its contribution to the total income produced, and this amount is regarded as a "fair" compensation for its services. It should be noted that the conception of "fairness" comes from the assumption that in a perfect market each factor gets in income a share that is equal to what it produces. As evidence that such standards are appealed to in the actual affairs of life, we may suggest that whenever courts or public utility commissions are called upon to determine what is a fair return on the investment in some public utility, they regularly resort to the conditions that prevail in competitive industry as a basis for rendering their decisions. If 6 per cent, or some other rate, is commonly earned in competitive industry, then a price for the utility service which will yield a similar rate on the investment in the public utility will be regarded as fair. Thus, while ethical standards may be derived from the results of economic analysis, it is not difficult to distinguish between the development of economic principles and the formulation of standards to determine questions of general social policy. The one lies in the realm of economic analysis, the other in that of social ethics.

•Business Economics.—In recent years the expression "business economics" has come to be widely used, both among professional economists and among the more popular writers on economic subjects. Interest in the general economic questions of the community has fostered an active discussion of a wide range of economic topics. This discussion has concerned itself with many problems that vitally affect private business and, although essentially popular in character, it has been carried on in semi-scientific terms. It has, therefore, been an easy step to adopt this new expression without much thought as to its impli-

cations. There is a need, however, for a distinction in the use of the term "economics" and the expression "business economics" with which the careful student should be familiar. This distinction is approximately the same as that implied by the terms "pure" and "applied" science. By way of illustration, take a science such as chemistry. So long as the chemist is studying the chemical elements of nature and endeavoring to discover the laws of chemical combinations, his conclusions are general in character and may be said to lie within the domain of "pure" chemistry. But when he begins to apply his knowledge of chemical laws to the accomplishment of some specific purpose, as when he analyzes the chemical properties of the soil with the view of providing proper nourishment for the plants to be grown therein, then his studies pass over into the realm of "applied" or industrial chemistry. This distinction may frequently be one of degree, yet nevertheless it is often extremely useful to have such a differentiation made.

Likewise, in economics there is need for a distinction between the formulation of general principles and the application of those principles to concrete situations. It is one thing to attempt to develop a rule that will assist in explaining general economic relations, and quite another to attempt to use this formula as a guide for the operation of a specific industry, or for determining the policy of an individual business concern. Considered as a science, economics deals with the economic activities of a social group. Although the phenomena which constitute the basis of study consist of the actions or choices of individuals, these actions are never considered in isolation, but are always thought of as a part of the general group activity. The buying and selling of an individual appears at first sight to be solely personal but, in fact, it affects the experience of the whole group. In modern society "no man lives unto himself alone," and what is true of an individual in this respect is equally true of a business establishment. Hence, any attempt to explain the economic forces operating within a social group must go beyond an explanation of the actions of an individual or the internal operations of an individual plant. Economics undertakes to deal first of all with the broad, general experiences of the social group and to reduce these experiences to common rules or principles. The formulation of these broad, general principles constitutes the domain of economics as a science.

In the present system of economic organization, the responsibility for applying general principles to concrete situations rests with the business executive or manager of a concern. While nominally free to operate his business independently, he is, in fact, dependent upon conditions beyond his control. His establishment is but a unit in the economic structure of the whole social group in which it is located and, therefore, its activities, if financially successful, must be adapted to conditions that lie beyond his control. There are two forms of adjustment that have to be made.

First, the internal policy of a business has to be adapted to the external conditions surrounding it. No manager has complete control over the purchasers of his product, hence, it is necessary for him to produce goods that people want at a price they are willing to pay. Knowledge of the general principles of value and of the facts and forces that determine these principles will assist him in making this external adjustment.

Second, the general principles frequently have to be adjusted to the peculiar problems of an industry or even of an establishment. These internal problems may pertain to the organization of plant and equipment, to industrial relations, to the financial structure, or to any one of the large number of perplexing problems that confront the executive of a modern business concern. This situation may be illustrated by two examples selected from questions of business policy that differ quite widely in character. First, we may take the question of the determination of the wage rates to be paid. Wage principles are expressed in general terms, but the wage bill has to be paid in specific amounts per hour or per piece. The business executive may be convinced that the general principle of compensating labor on the basis of the value added to the product by each worker is correct, but his problem is to measure with some degree of accuracy the contribution of each worker. The determination of the wage rate, while falling within the scope of general principles explaining wages, requires special adaptation for practically every operation within a plant. This adaptation has to be made not only with some approximation to scientific accuracy, but in such form as will satisfy the psychological attitude of the worker.

A second example of somewhat different nature will illustrate still further the adaptations that the business executive has to make in order that general principles may fit the peculiarities of

his business. As the standard for the organization of the economic activity of any social group, there has been developed the conception of an equilibrium between the capacity to produce and the capacity to consume. This equilibrium, or balance, is never actually attained, but is an ideal toward which the economic forces operating within that group are constantly pressing for realization. Since there are literally thousands of business executives who are producing in anticipation of demand on an estimate of that demand, it is not surprising that a perfect balance is not attained. Sometimes the supply is in excess of, and sometimes less than, demand, but the general economic forces constantly tend to compensate these mistakes of judgment. The fluctuating movements of business, now commonly known as the business cycle, are merely manifestations of this failure to attain a perfect balance between supply and demand. As a means of forecasting these cyclical movements, a number of agencies have developed curves based upon general statistical data. These curves are designed to show general trends. But the problem of the business executive has to do with the future demand for his products. No general curve of demand is perfectly adapted to his business. Hence, if he uses devices of this character intelligently, it is necessary for him to develop curves based upon the experience of his own concern and relate these to the general curves. In this way he can determine whether his business shows regular cyclical movements and what the relation of these movements is to the general movement. It is a matter of real significance for him to know whether the cyclical movements within his plant come before or after the general movements.

What has been said concerning wages and the adjustment of the policy of a business to the general cyclical movements may be equally true in connection with a vast number of judgments which the business man must render. Here is a field that occupies about the same relation to economics as engineering does to physics and mathematics, or industrial chemistry to "pure" chemistry. The term "business economics" may very well be applied wherever there is an attempt to adjust or adapt the policy of an individual business to the general economic conditions of the social group, or to adapt the general principles to fit the peculiarities of a particular concern.

Social Economy.—Most modern texts on the principles of economics devote considerable space to what is sometimes called "problems." These so-called problems usually fall either within the sphere of social control, as the regulation of railroads, or of monopoly, whether public or private, or that of "business economics," as here defined. The expression "social economy" is sometimes used to denote this field, and there seems to be a real need for terms that will distinguish between the adjustment of the policy of a private business to the general principles, and the application of these principles as an aid in developing social control over private industry. Business economics, and social economy seem well adapted to these two purposes.

The Method of Treatment.—From the preceding discussion, it should be apparent that the study of economics is primarily theoretical in character. The method used in its development is essentially the same as that found in the treatment of other sciences. There is observation of economic facts, comparison to discover likenesses or differences in the data observed, classification, and generalization. In so far as uniformity is found in the behavior of any group of economic facts, attempts are made to explain this uniformity and these explanations and interpretations become the principles of the science. What economics does, then, is to survey the facts of the present economic system, and arrange and analyze them with the view of finding principles that will explain them. Isolated facts are a mass of disjointed details and by themselves are meaningless, but when subjected to careful analysis they may yield uniformities that bring order out of apparent chaos. This is the task of economics as a science, and, like every other science, it progresses by discarding previous generalizations when new facts or further analysis of old facts reveal unsound or imperfect conclusions. New generalizations are constantly being made in the light of new facts. Each generation must restate the principles in the light of the economic relations of its own day. In this respect economics is like every dynamic science. The opportunity for new adaptations becomes the challenge to the best minds of each generation and this challenge adds to the interest and stimulus which students find in the subject.

Because economics is theoretical, some persons are inclined to discredit the subject and to disregard its findings. This is especially the case of some practical men who are frequently

unnecessarily skeptical of theory. Such men usually associate theory with a lack of reality and, unfortunately, it does frequently happen that theoretical statements are deficient in facts. But faulty reasoning is more likely to appear in the thought and conclusions of the practical man himself than in the generalizations to be found in economic treatises. As Marshall has well said:

It is not the function of a science to lay down practical precepts or to prescribe rules of life. The laws of economics, as of other sciences, are couched in the indicative, and not in the imperative mood; they are statements of the effects produced by different causes, singly or in combination. They are not rules made for immediate application in practical politics.¹

The critics of theory overlook the fact that the results of economic theorizing were not intended to be used as absolute rules in formulating judgments concerning specific questions of either public or private policy. Decisions of this character should be made in the light of the specific facts that surround the particular situations, whether these be historical, psychological, social, or physical.

Consciously or unconsciously everyone deliberates upon economic facts and either draws independent conclusions or uses the *generally accepted conclusions* of the group with which he is identified. Scientific method puts all conclusions to the test and accepts only those which are helpful in explaining the phenomena that are observable. Theory, then, should be recognized as an attempt to reduce economic phenomena to a body of principles that will be helpful in explaining the causal relations existing in this phase of human activity. There should never be any conflict between theory and reality as theory is the scientific attempt to explain the relations existing among the facts actually found in our day-to-day economic experience.

In studying the subject, no better practice can be followed than constant reflection upon the economic experiences which every individual has. The ordinary purchases which one makes and the choices necessary in distributing his income may become exceedingly useful as a means of furnishing concrete data to illustrate what may otherwise appear abstract and unreal. Every student has some economic experience and, if he will

¹ DUNBAR, C. F., "Economic Essays," p. 64, footnote.

bring this under careful scrutiny, he will find that this practice will assist him in grasping the generalizations set forth in the body of the text. Naturally, the discussions will be more instructive to those who are familiar with the ordinary transactions of business. Those who are deficient in this information will find it helpful to supplement their knowledge by reading the literature that is descriptive of the structure and operation of modern economic life.

Suggested References for Further Study¹

- BOUCKE, O. FRED, "Principles of Economics," Vol. II, Appendix, Pts. I, II, and III.
BYE, RAYMOND T., "Principles of Economics," Chap. I.
BYE, RAYMOND T., and HEWITT, W. H. "Applied Economics," Chap. I.
CARVER, THOMAS N., "Principles of Political Economy," Chaps. I-VII.
CHAPMAN, S. J., "Outlines of Political Economy," Chaps. I and II.
CLAY, HENRY, "Economics for the General Reader," Chap. I.
ELY, RICHARD T., "Outlines of Economics," 4th Ed., Chap. I.
GIDE, CHARLES, "Principles of Political Economy," Rev. Ed., Chap. I.
HAYES, H. GORDON, "Our Economic System," Vol. I, Chap. I.
HENDERSON, HUBERT D., "Supply and Demand," Chap. I.
KEYNES, J. N., "The Scope and Logical Method of Political Economy."
MARSHALL, ALFRED, "Principles of Economics," Book I, Chaps. V and VI.
PIERSON, N. G., "Principles of Economics," pp. 1-9, 15-33.
SELIGMAN, EDWIN R. A., "Principles of Economics," Chap. II.
WALKER, FRANCIS A., "Political Economy," Chap. I.
WOLFE, ALBERT B., "Functional Economics," Chap. XIII, in "The Trend of Economics," edited by R. G. Tugwell.

¹ The references mentioned at the end of each chapter are not intended to be a complete bibliography on the topics covered but are incorporated as an aid to the reader who wishes to pursue the subject matter further.

CHAPTER II

THE CHARACTERISTICS OF MODERN ECONOMIC SOCIETY

The Meaning of an Economic Order.—The body of principles which constitute the science of economics has been developed from an analysis of the economic relations of mankind. The starting point in this analysis is the recognition of the existence of human desires. Economists accept this fact but make no attempt to analyze the psychological character of desires, a task that belongs to the psychologists. The existence of recurring and unsatisfied desires tends to stimulate mankind to the activity necessary to produce the means of gratifying them. All industry has developed in response to this stimulus. If we wish to obtain a picture of the economic mechanism of any period, we may ask, How do the people of that period secure the food, clothing, shelter, or other goods which they consume? A little reflection on this question will reveal that the methods differ as we go back from generation to generation. Originally there was man with such mental and physical powers as he possessed and the resources of nature. By slow, laborious steps he has developed a comprehensive and complex industrial structure. This whole structure, which takes the raw materials of nature and brings them through a long line of complicated processes into a form capable of serving the purposes of man, has been developed because of the imperative demands of human desires. The organization existing among these various processes is called an "economic order."

Before one can understand economic principles and be able to think independently concerning economic facts, he should be familiar with the structure of the period to be studied, because economic principles are merely an attempt to explain the relations obtaining within that structure. One should be able not only to visualize the economic mechanism in its outward characteristics but he should comprehend the delicate adjustments that exist within that mechanism and the complicated factors that influence the operation of its various parts as well as the structure as a whole. The purpose of this chapter is to set forth

some of the chief characteristics of the economic society in which we live as a background for the study of the principles operating therein.

Four Systems of Economy.—In satisfying the desires of mankind different systems of economic organization are possible. By way of contrast and as a means of securing a better understanding of the present order, four systems of economy will be briefly reviewed.

First, the Autonomous Order.—It is possible to think of a system in which every individual by his own unaided efforts produces the goods necessary to gratify his own desires. Robinson Crusoe has long stood as the stock illustration of such a system. While here and there an isolated individual may produce the things he consumes, this is by no means typical of modern economic life. During the pioneer days of our country, however, the settler and his family approximated this method of producing the things which they consumed. Food, clothing, and shelter were largely the product of the efforts put forth by the members of the family. The family unit was, in effect, self-sufficient. An economic system in which the individual, or as in this case the pioneer family, produces all the things that he consumes has been aptly called an "autonomous economic order."¹ It is an order in which the individual or family is self-sufficient.

Second, the Communistic Order.—At the opposite extreme from this order, a system of economic organization may be conceived in which the community is completely organized for the purpose of producing and distributing the things which its members consume. In such a system the individual would have little freedom of initiative and enterprise. He would perform the tasks assigned him and consume whatever was allotted to him. There would be a high degree of organized, cooperative effort in such a society, but the responsibility for enterprise and industry would be assumed collectively. An economic organization of this character would constitute what is usually understood as "communism." In a society of this character no responsibility for the operation of the system would be left to individual initiative. Both production and distribution would be col-

¹ TAYLOR, F. M., "Principles of Economics," (The Ronald Press Edition), p. 17. I am indebted to Professor Taylor for several terms and ideas that appear in this chapter. These can be found in his admirable chapter, entitled, "The General Survey of the Existing Economic Order."

lectively organized and operated. Such a system may be called a "communistic economic order."

Third, the Socialistic Order.—Between the autonomous organization and the extreme centralization of all the economic processes, as under communism, stand two other systems, both of which are cooperative in character but very unlike in organization and operation. In both, the organization is less complete than in communism. First, there is a system in which the individual would have a large degree of freedom in connection with personal consumption, but not in connection with production. He would be paid wages or a salary for his services and be free to buy from socially owned and operated markets the goods which he consumes in much the same fashion that the individual makes his purchases now. But the productive processes would be collectively owned and operated. Farms, mines, and workshops would be socially owned, and all persons engaged therein would be in the relation of employees. All industry would be organized in much the same way as the post office or the public schools are today, and labor would be employed under conditions similar to those which now obtain in public institutions. Roughly speaking, this is the proposal of the socialist, and industry organized in accordance with this plan has been called "socialism," or a "socialistic economic order."

None of the three systems thus far described is in operation on a scale of any size at the present time, although experiments have been, and are being, tried along one or other of these lines. The most conspicuous experiment of our day is that now in operation in Russia. The soviet organization of industry involves a high degree of collective ownership of all the productive processes, and the individual is paid a salary or wage—conditions similar to those proposed under socialism. There are many people in the world who believe that the welfare of mankind lies in the direction either of socialism or communism.

Fourth, the Individual Exchange Order.—Lastly, the system of economic organization under which most civilized peoples are living today also lies between the two extremes, the autonomous and the communistic orders. In this system, large freedom is left to the individual both in determining what is produced and what he individually consumes. Within broad limits, he produces what he chooses, sells his product on the market for what it will

bring, and, with the proceeds buys those things for which he has need. The responsibility for the organization of production and for the determination of both the character and quantity of the goods produced rests mainly with the individual. For this reason the system may be called an "individual exchange order." While it is individualistic in character, it differs from the autonomous order in that production is for the market, rather than for one's self. Instead of self-sufficiency, there is a high degree of mutual dependence operating through the market.

Individual Initiative and Private Enterprise.—Turning more specifically to the characteristics of the present system, the feature that distinguishes it from the socialistic and communistic orders is, as has just been suggested, the place occupied by the individual in its operation. Instead of collective responsibility, the system is built mainly, though not wholly, on individual initiative and private enterprise. Reliance is placed upon the motive of self-interest as the means of enlisting the productive energies of the individual and of promoting the economic welfare of the social group. The individual is left free, within rather large limits, to choose his occupation or work, to sell his services or products, and to buy those things for which he has a desire. Through this freedom and private initiative, the individual tends to conform to a social purpose. Under the guise of producing for himself, he is in reality producing for others. For instance, the farmer is growing grains or other food products; the manufacturer is producing machinery and tools, or is preparing consumers' goods of one kind or another; the builder is erecting homes, factories, or stores in which to conduct business; and in each case the products are made to sell. Broadly speaking, the larger the volume and the better the quality of the goods brought to the market by these producers, the greater will be their command over other goods and at the same time the more completely will human desires be gratified.

It should be apparent that the amount of wealth that the individual secures for his own personal use depends both on the quantity and the quality of the goods which he brings to the market. If he idles his time, or brings only those goods that few persons desire, his bargaining power will be small and he will get few goods in return. On the other hand, if he produces a large volume of goods of the kind keenly desired by others, he will be able to secure a large amount of wealth for his own purposes.

Thus, by permitting the individual the right to enjoy the fruits of his own efforts, this system places a high premium upon the use of all of the creative powers which he possesses and stimulates him to make the largest contribution to the economic welfare of society of which he is capable.

The weakness of the system is not in its appeal to self-interest but in the possibility of a conflict developing between individual and social welfare. So long as the economic interests of the individual can be kept in harmony with those of society, the well-being of mankind, economically considered, will be promoted by private initiative. But unrestrained liberty may create a conflict of interests. Not every commodity which can be produced and sold to private advantage will contribute to general welfare. An individual, or more likely a group of individuals, may secure a special privilege of one kind or another, commonly known as monopoly, and exercise this in such a manner as to cause a conflict between his personal interests and those of the balance of the community. Similarly, one may produce some commodity, as a harmful drug, which he can sell for a profit, but the use of which is clearly detrimental to mankind. In all such cases, society may limit the freedom of individual initiative and private enterprise in the interest of what is accepted as the common welfare. Within the limits thus set, however, freedom of enterprise and self-interest are powerful forces calling forth the productive energies of mankind. The large volume of wealth that yearly flows from the factories and workshops of the world bears testimony to the effectiveness of this appeal.

Division of Labor and Specialization.—In its outward aspects, one of the most striking features of the present system is the extent to which division of labor and specialization are carried. On every side are seen men specializing in a single industry or occupation. They are farmers, factory workers, miners, carpenters, bricklayers, lawyers, doctors, ministers, etc. Indeed, specialization does not stop at the division of occupations, but practically all occupational groups are subdivided. The farmers of the South specialize in growing cotton; farther north the grains are the chief products, and so on. In manufacturing, the processes have frequently become so minutely subdivided that the workmen are required to perform a simple operation while the material passes them on a conveyor, as in the case in the slaughtering houses and in many of the automobile and other factories.

The same tendency toward specialization is observed in the professions. The family physician, who was once a general practitioner, has given way to the specialist; the lawyer also devotes his attention to patent, criminal, constitutional, or to some other specialized branch of the law. In fact, wherever we examine the character of modern economic life we find this tendency to specialization present as one of its most significant features.

The direct result of the division of labor is the development of exchange. Specialization necessitates an exchange of goods and services as a means of gratifying our desires. The market, then, becomes the organized agency to which the individual may bring the products of his own specialized effort for the purpose of trading them for the products of other persons equally specialized. It is for this reason that exchange is emphasized in the description of the present order. It is an exchange order in which the individual assumes responsibility for production, but he produces for the market rather than for personal use.

A Cooperative System.—Broadly speaking, the present system is cooperative in character. At first this term may appear a contradiction of the statement that the system rests upon individual initiative and enterprise. How, it may be asked, Can a system be cooperative, when the individual decides what shall be produced? The term "cooperation" is generally used when a group of persons associate for some common benefit, as in the case of a cooperative store, or creamery, or some agency for the marketing of farm products. For instance, consumers may establish a store that is jointly owned and operated for the benefit of the patrons, and, in like manner, the dairy men of a given district may establish a cooperative creamery for the purpose of obtaining for themselves the gains from this further step in production. In this usage, cooperation implies consciously organized, collective action on the part of the interested parties. But as here used, cooperation refers to the general social process by which mankind as a whole gets a living. The tasks of individuals may be highly specialized, as indicated above, but when a survey is made of the whole productive process it is seen that men after all are working with each other and producing for each other. Viewed in this light, the farmer, who produces wheat is cooperating with the miller and the baker to furnish society with bread. What is true in this instance is generally true in all other

lines of industrial endeavor, *i.e.*, the individual produces some specialized commodity or service which he trades in the market for the goods or services of others. It is, therefore, correct to conclude that there is, in the existing system, a large degree of mutual dependence in finding the means of gratifying human desires. In this sense, the present system may be said to be cooperative, the cooperation being effected through the market.

By way of illustrating how cooperation works in the present economic organization, think of any common article of consumption. Trace back the different processes of production through which the good has gone in preparation for the market. Clay uses the purchase of a woolen shirt to illustrate this point.¹ The purchase is made at a local store. The local storekeeper bought the shirt from a wholesaler or direct from the shirt manufacturer. The shirt manufacturer bought the materials from the manufacturer of flannels. The yarns used in making the flannel had to be spun from the wool, which may have been grown on another continent. Thus, the material that has gone into the shirt has passed through several factories and doubtless through the hands of many middlemen, and may have traveled half way round the globe before becoming a finished product.

But the woolen industry is not the only industry involved in the production of the shirt. Each factory through which the materials have passed required machinery. The production of machines involved the iron and steel industry. The oil and leather industries were likewise called upon to furnish materials. The coal industry had to furnish fuel in the development of power; the building trades constructed the factories and warehouses; the transportation systems were necessary to move the materials to points where the various processes of production took place. At each successive step in the movement of the materials, the credit and banking system was called upon as a necessary means of facilitating the operations. "It would hardly be too much to say that the apparently simple transaction of purchasing a shirt was the completion of a process in which the whole modern economic system was involved."² This simple illustration can be multiplied out of the experience of any individual who makes

¹ CLAY, H., "Economics for the General Reader," pp. 1-4; BURNS, A. R. and E. M., "The Economic World," pp. 1-11.

² CLAY, H., *op. cit.*, p. 3.

purchases. The point here emphasized is that there is a high degree of cooperation in the modern economic order.

Spontaneous Cooperation.—While it is apparent that the present system when viewed as a social process, is cooperative in character, yet we should note the way in which this cooperation is effected. In socialism and communism, cooperation is *conscious* and centrally determined, but in the present system it is *unconscious* or *spontaneous*, and to a large extent *automatic*. It is effected through the free exchange of commodities produced by individuals and sold by them in the market.¹ The individual in managing his business is not ordinarily aware that he is cooperating with others in producing a flow of wealth for the satisfaction of human desires. He may at times be compelled to face his dependence upon outside forces, as when a shortage of raw materials or fuel may cause a stoppage of his production, or conditions in the market may prevent the sale of his goods. Such experiences are common enough to cause the business manager to recognize his dependence for the success of his enterprise upon forces beyond his control, but, nevertheless, he is not ordinarily concerned with the social results flowing from his operations. His attention is centered on securing a private income. If his accounts show a surplus above costs, the undertaking is a success from his point of view. The pursuit of gain is the economic motive which actuates him and directs the course of his business decisions. While he is thus impelled by the desire for private gain, the results of his action contribute to the general economic welfare by the additions which he makes to the flow of wealth through the market. All producers are manufacturing, as it were, for a great common reservoir, the market. Each brings his own product which he exchanges for goods made by others. In this way each industry is making its own contribution to the sum of goods available for the gratification of human desires. Hence, there is a high degree of cooperation in this process when it is considered from the social, as distinguished from the private, point of view. The individual business man, pursuing the motive of private gain is, in fact, harnessed to a general, social purpose. Producing, as he sees it, an income for himself, he is in reality producing for others. Therefore, it may be said that the cooperation in the present system is brought

¹ TAYLOR, F. M., *op. cit.*, pp. 17-21.

about through the market by means of unconscious or spontaneous influences.

Private Property.—The foundation of the present system is the institution of private property. As here used, property is a legal concept and means the right of exclusive control within the legal structure of society over valuable things. This right, as will be shown later, is derived from the state and the powers of the state may be invoked to protect it. Its significance in the operation of our economic life is the protection it affords the individual in the enjoyment of the fruits of his own efforts and enterprise. For free enterprise to function effectively, the individual must either be strong enough to protect himself and his property, or he must be protected by some agency outside of himself. How long would goods continue to flow regularly into the markets if the rights of ownership were not respected and physical force were the only test of the right to use? To raise the question is to answer it. It is perfectly evident that modern industry could not exist if the law of banditry were the only law of the land.

In order to call forth the most productive energy of the individual and to secure to him the benefits of his own initiative and his own enterprise, the institution of private property came into existence. Through private property the individual is given the authority to determine what shall be produced, how it shall be produced, and what disposal shall be made of the finished goods. In other words, control over the use of the factors of production rests mainly on this right. The owners of land and capital particularly are the persons who decide what use shall be made of these factors and what goods shall be brought to the markets. So important and fundamental is private property in the present economic organization that we scarcely realize that it is not only the creation of man but that it has evolved in accordance with the experiences of past generations. However fixed it may seem, it is important to recognize that this right has not always existed in its present form and is still subject to change and adaptation in accordance with the needs of any period. Enough has been said to indicate that private property is fundamental in the operation of modern industry, and must, therefore, be regarded as one of the chief characteristics of the existing economic order.

Bequests.—Closely associated with the right of private property, and in fact flowing from it, is the right of bequest.

The individual is protected not only in the ownership and use of his wealth, but also in his right to bequeath it to others. The distinction between the right of private property and that of bequest is the fact that the former gives exclusive control over property, while the latter is confined to the *transmission* of property from one generation to another. This right is peculiar to the present economic order, for under socialism and communism there would be scarcely any private property to be transmitted, as property would be collectively owned. The social significance of the right to bequeath property lies in the assumption that it acts as a stimulus to the productive efficiency of the individual. It is assumed that among the motives which actuate the individual is the desire to provide for his children, or for future contingencies of one kind or another. In so far as this motive serves to stimulate the productive energies of the individual, its enlistment through inheritance coincides with the social purpose of a large flow of wealth. Only to the extent that the individual is stimulated to make a larger contribution to wealth can we say that the state is justified in establishing and protecting the right of inheritance.

Right of Contract.—Next in significance to the right of private property is the right of free contract. Contracts are the connective tissue, as it were, that ties the various and separate units of industry into one complete system. By this right the state permits men to bind themselves and, when the agreement is voluntarily entered into, it compels them to abide by its terms.¹ Underlying all the operations of modern industry is some sort of contract. Business men place orders for goods, and this act gives rise to an oral or a written contract, depending upon the practice in the industry. In many industries these orders have not been regarded as binding (that is enforceable), hence cancellation has been, and is, a common practice. During a period of rapidly declining prices, this practice is keenly felt by business men and causes much confusion particularly between the manufacturer and the merchant. However, it is by means of contracts that members of the industrial system are bound together and through their use much of the cooperation described above is made possible.

¹ While this statement is substantially correct, some important qualifications should be made, as the state will not recognize as valid all forms of contracts, even if freely entered, as for instance a man may not sell himself / into slavery.

Control in the Present System.—Another important feature of the present economic order is the way in which control is exercised. By control, as here used, is meant the planning and the directing of the operations of modern economic life. The criticism is not infrequently heard that the existing system of economy is planless, and the clear implication is that it is without order. Our attention is directed to the duplication of plant and labor to be found in all lines of economic endeavor. Why, it is asked, is it necessary to have so many separate milk-distributing companies, with wagons covering the same ground and delivering milk to the same apartment buildings?¹ Why have so many middlemen, each adding his private toll to the costs of producing the things we consume? It is asserted that economy in the use of both labor and capital would lead to the centralized operation of such industries and to the elimination of the wastes entailed by the duplication of plant and equipment. Other instances of waste and duplication are cited as evidence that the present system of industry is planless. Further evidence is also found in the frequent periods of industrial depression and unemployment. During such times there are as many factories and workshops in a country as in prosperous times; there is just as much raw material to be worked up; and that there are thousands of unemployed men and women willing and anxious to work but unable to find jobs. Facts such as these are marshalled as evidence of the lack of plan or system in the operation of the existing economic order.

Granting that there are many evidences of waste and duplication in the operation of modern industry and many instances of imperfect adaptation of production to consumption which result in recurring periods of industrial depression, is it true that the system is without plan and is chaotic in its organization and operation? In other words, is there no control exercised over its operation? In considering this question it may be asked, How often does the consumer, when he has need for goods or commodities, fail to find them waiting on the shelves of the store-

¹ During the war a special committee sat in Chicago to fix the price of milk between the producers and the distributors. A citizen's committee organized by the City Club collected and presented a large body of facts concerning the organization and costs of distributing milk. It was found that the milk wagons operating in the city traveled roughly 25,000 miles in making daily deliveries. This distance was about three times the combined length of all streets and alleys in the city.

keeper? In normal times, he is seldom disappointed when he is in the market as a purchaser. Not only are the goods there before him waiting to be purchased, but he is besieged on all sides by skilful advertisements and sales slogans designed to induce him to become a purchaser. The fact that the goods have been produced in anticipation of the consumer's demand is evidence that there is some plan present in the operation of the system. Some control must be exercised in order to cause the system to conform to the primary needs of mankind.

One further illustration may be cited. Within the metropolitan area of a great city, like New York or Chicago, reside millions of people. Just imagine the great volume and multiplicity of commodities that must enter those areas daily in order to satisfy the wants of these people! Think what would happen if this steady flow of the necessities of life were to stop for only a few days! Hunger and famine would very soon begin to show their ravages. Yet, day after day, the residents of these areas lie down to peaceful slumber with full assurance that the break of a new day will bring no calamity and that the ends of the world will contribute to their wants. Reflections of this character are very convincing that there is order and plan in the operation of our economic life. While there is no human authority presiding over this complex system, consciously organizing and directing it, nevertheless, the people have confidence in it and the basis for that confidence is to be found in the remarkable precision with which it works. The production of the vast volume of goods that flows daily into the markets in anticipation of demand is such convincing evidence as to the presence of order in the operation of our economic mechanism that it cannot be set aside by reference to the mistakes in calculation and maladjustments that occur in the attempts to balance production and consumption.

Even though the evidence of plan in the existing system does seem convincing, it may still be asked, Where does this plan originate, and how does it operate? How is the nature and quantity of goods produced determined? These questions raise the whole problem of the exercise of control in modern industry. Under socialism and communism, the control would be conscious and exercised by central authority. The determination of what would be produced and the operation of the whole productive process would rest in the hands of officials elected for this purpose. But in the economic system of today,

control may be either *unconscious* or *conscious*, i.e., *spontaneous* or *predetermined*. In the main, control like cooperation, is spontaneous. The individual, guided by the normal motive of private gain, is led to conform to a general or social purpose. He looks about and sees the needs of mankind; he studies the market and decides to produce those things which he thinks he can sell at an advantage to himself, and he continues to produce as long as the market price fully covers his costs of production, with possibly some surplus for profit.¹ But the determination of the commodities produced is left mainly, though not entirely, to the initiative of the individual. If his estimates for production are too low, that fact will be registered in the market by a rise of the price of the commodity or commodities affected. Profits will then increase and, in order to take advantage of the more favorable conditions, producers will increase their output. If too much has been produced, the reverse process will work itself out through market influences similar to those just described. The effect of the operation of these influences is to bring goods onto the market when needed and to check production as the demand for them falls off. In spite of the maladjustments and miscalculations which occur, the result is a remarkably orderly procedure in the operation of our economic processes. The regulating influences, or in other words, the control thus far described, operate unconsciously and are guided by spontaneous forces.

While spontaneous and unconscious forces are important in regulating the present system, no study of it would be complete that failed to take account of the *conscious*, or *predetermining* influences. Conscious control may be either privately or publicly exercised. When privately exercised, it may vary in completeness from a joint collection and exchange of statistics of production, of supplies, of costs, or of any other business facts that may serve as a basis for developing an intelligent production program to an exclusive control over the entire industry. The complexity of modern business has given rise during the past decade to an increasing demand for the compilation of business statistics to be used for the purpose of forecasting the trend of business. The desire on the part of business executives for information of this character has called forth a large number of statistical agencies that compile business facts and undertake to forecast the course

¹ This subject is more fully treated under Profits, Chap. XXII.

of business and to advise men how to shape their production policies more intelligently.¹ Whenever public or private agencies furnish information of this character that can be used as a guide in the operation of our economic mechanism, the control to that extent becomes conscious. The more accurate the information and the more scientifically it has been collected and interpreted, the more effective does this method of control become.

Should control over production become exclusive and give to an individual or group exercising it the power to dominate the productive processes, then, there is developed what is commonly known as "monopoly." Wherever monopoly exists there is a conscious effort made to adapt the processes of production to the advantage of the monopolist. This type of control is most effectively applied where there is a possibility of limiting supply. Whenever an individual or group of individuals get sufficient power over the production of some commodity or in rendering some service as to enable them to dominate the whole industry, the control is conscious and the business may be said to be a private monopoly.

Conscious control is publicly exercised when the regularly constituted governmental agencies, as a municipality or a state, undertake to render some service directly, like the operation of street railways, gas and electric plants, or any other similar kind of a service. There is a steady increase in the number of services that are rendered through some form of public agency. Another form of conscious control that is semi-public in character is that found in the organization of consumers, as in the case of the great cooperative enterprises of England, where the attempt is made to govern the productive and distributive processes for the benefit of the cooperators. Activities of this character are types of conscious control for the benefit of the particular group that maintains the enterprise. We may conclude then that conscious control is present in our economic system, operating through both private and public agencies. Like spontaneous control, it is a force operating through the market.

The Place of the State in Modern Industry.—In the operation of our system it has been shown that we depend mainly upon individual initiative and private enterprise, but our understand-

¹ Cf. SEARIS, H., "An Introduction to Statistical Methods," Rev. Ed., Chap. XVI, for a comparison of the various index numbers now in general use as aids in forecasting business conditions.

ing of the system would be incomplete if we failed to take cognizance of the influence of the state upon our economic life. The term "state" is here used in its generic sense and is intended to include national, state, and local governmental agencies. In this sense, the state affects our economic system in two principal ways. First, as has been indicated, it renders some services directly, as the maintenance of order, provision for public highways, and other services similar in character. The state undertakes especially to provide various types of developmental services, such as the provision for public schools and other educational institutions; and the promotion of scientific investigation and experimentation, such as that carried on by the scientific branches of the government. Of like kind is the provision for forestry, irrigation, and other similar projects; also the postal system and other means of fostering communication. In all of the instances mentioned, it should be noted that the nature of the undertakings is such as to make them extremely difficult to operate as private enterprises, either because of the costs involved or because of the peculiarities of the industry itself. Besides, in many instances, the work can be done more effectively when undertaken by the state. By rendering these services directly, the state is exercising conscious control over a part of the operations of our economic system.

In the second place, the state both protects and limits the freedom of individual initiative. It protects individual initiative, as was shown above, through the right of private property. A very large part of the activity of state and local governments is devoted to this task. The survey and description of real estate, the recording of deeds, the activities of courts in settling conflicting claims to property are among the numerous ways that the various governmental units are occupied with property rights. While the state thus undertakes to protect the individual in the exercise of his freedom of enterprise, it also sets limits on private initiative. In some instances, the individual is prohibited from producing things which he could make and sell at a profit, such as harmful drugs, alcoholic beverages, and the like. The state has either prohibited the production of such articles or it has regulated their production and sale because of the social consequences that flow from unlimited freedom of individual initiative and enterprise in such lines. In other instances, the state limits freedom in one direction as a means of increasing it in another.

Laws regulating the hours of labor or other working conditions limit the freedom of the employer and employee to make a labor contract in excess of the permitted hours, but increase the liberty of the worker to enjoy leisure. Limitations of this character are enacted where a public purpose is recognized, and it is through the police power of the state, *i.e.*, the indefinite power residing in government by which the freedom and initiative of individuals are limited in one direction as a means of expanding them in another, that these ends are accomplished.

Regulation of Private Control.—One further instance of state activity may be cited. When private enterprise, following the motive of private gain, has resulted in exclusive control over the production and sale of some good, in other words, a condition of monopoly, the state has usually undertaken to prohibit such action or endeavored to regulate it so that the benefits will become general rather than accrue to the particularly favored individual or group. One of the primary functions of government is to limit the strictly selfish aspects of private enterprise. Such conditions have arisen in connection with the operation of railroads, electric railways, gas and electric companies, and other similar industries. In all industries of this character, the state has undertaken to limit private control so that it will conform to general or public interests. Thus, while our economic system is mainly one of individual initiative and private enterprise, yet in its operation it is greatly affected by the state, as defined above, which both limits and supplements individual enterprise with the view of developing the welfare of all. Running through the whole system we find a large degree of cooperation effected through the market. It seems appropriate, therefore, to describe the economic organization under which we live as an *individual exchange economy*.

Competition as an Organizing Influence.—In the description of the structure given above, emphasis has been placed on the fact that the forces that control the system operate through the market. As here used, the market means the area within which the influence of buyers and sellers operates. As will be shown in greater detail later, markets may vary widely in scope, but in all cases the same general characteristics appear, *viz.*, the endeavor of buyers and sellers to effect an exchange of commodities or services. The chief force which affects both the buyers and the sellers in any market, and tends to direct them in the pro-

motion of a social purpose is competition. So important is this force in our economic life that we frequently hear the structure referred to as a competitive system. There can be no doubt that competition is a very real force, yet it is difficult of exact definition. In its broad aspects, it is the endeavor of two or more persons to gain the same end. In economics, it is the endeavor or striving of two or more persons acting independently to secure the custom or trade of a third party. This rivalry takes place in the market and, aside from a comparatively small number of idlers and economic parasites, individuals are everywhere striving to discover the most urgent wants of other people and are endeavoring to serve them more effectively than any one else in order to secure and retain their patronage. The motive behind this rivalry is selfish rather than altruistic, for experience has shown that, where competition is most active, service to customers is a necessary condition for the retention of their trade. As a result of this mutual striving to serve, individuals are led to make the largest contribution to the total volume of wealth available for the satisfaction of human desires. As a market phenomenon competition has two sides, namely, the rivalry to sell a good, or render a service, and the rivalry to purchase the same. It follows that the individual takes part in the competitive struggle both as a buyer and as a seller. The effect of competition is to coordinate the buying and selling influences operating in a market and, as a result of the play of these influences, the prices paid for commodities are determined. The prices, in turn, tend to determine the amount of production and thus to adapt it to the wants of man. Competition, then, is an all-pervading influence and one of the chief forces in the operation of our economic system.

Fair Competition.—While the present economic order relies on competition as an operating influence, it has been found by experience that the rivalry may take forms that are antisocial. In order to secure trade, sellers may deceive their customers; they may misrepresent their goods, or take advantage of those who are unable to bargain on equal terms. As a means of limiting this type of rivalry and of keeping competition in harmony with the prime social purpose, the state sets limits upon the competitive struggle. It prevents fraud and brute force, protects children and women from the evil consequences of unrestrained competition, prevents libelous and fraudulent

statements about a competitor's business, prevents individuals from inducing customers or employees from breaking their contracts, renders illegal the competitive struggle that has the definite purpose of running a rival out of business, and other similar antisocial activities. In other words, the state acts as an umpire to see that the competitive struggle is made to conform to the general welfare and is not allowed to serve private or acquisitive purposes only. The state endeavors to lay down rules for the conduct of competition which will put a high premium upon the use of those faculties and energies which will contribute most toward increasing the flow of wealth to the market.

Competition in Relation to Economic Theory.—In its historical development, the science of economics was based upon the assumption of unimpeded natural forces. Among these, free competition and mobility of the factors of production were the most important. Traditional economic theory assumed that there was perfect freedom of competition among producers and consumers, sellers and buyers; that each producer was striving to bring to the market those commodities which the consumers desired most keenly, and for which they were willing to pay a price that would yield the producer a profit. The evidence of the desire on the part of the consumers was the price they were willing to pay for the commodity. The buyers, it was held, would bid against each other and would not permit a purchase at a price lower than that prevailing for every other identical commodity. Likewise, sellers competed with each other for the trade of the buyers. As a result of this interplay of forces, the conclusion was reached that in a competitive market there could be but one price at a given time. In addition to fixing the price, the rivalry in the market tended to distribute the goods where they would yield the most satisfaction.

For competition, as here conceived, to be effective there must be perfect mobility of the factors of production. Land, labor, and capital must move readily and quickly from one use to another, the direction of the movement to be determined by the returns to the factor in the particular use to which it is put. If there were perfect competition and perfect mobility of the factors of production, then, according to the assumption above mentioned, market price would be a perfect measure of wants, and would serve as a perfect guide in the operation of our economic

system.¹ Through the operation of these forces, each factor in production would receive compensation in accordance with its importance in the productive process. In other words, distribution of wealth would be determined by these forces. Broadly speaking, these are the assumptions that underlie the writings of the early economists, and the conclusions drawn therefrom still occupy a large place in modern treatises on economics.

Only superficial observation is required, however, to demonstrate that competition and mobility do not operate with the degree of perfection which the earlier economists assumed. There are both natural and artificial obstacles to the perfect operation of these forces. Buyers and sellers do not in some instances, and cannot in others, have the knowledge of the supply of goods, nor of the needs of consumers requisite for the perfect operation of competition. Production for the market in anticipation of demand involves a time element. No business man, however well informed and shrewd, can forecast with precision the course of demand. Desires change with changing fashions, and the psychological reactions of a social group cannot be predicted with accuracy. Since all production requires time, the greater the length of time elapsing between the beginning of the productive process and the utilization of the finished commodity, the greater the risk of mistakes. The existence of failures of this kind is a further illustration of the impossibility of attaining a perfect balance between production and consumption. All that can be said is that competition works in the direction of that balance. In addition, producers may consciously undertake to control the processes for their own private advantage by limiting supply and fixing price, and thus prevent the attainment of perfect competition. The conclusions, therefore, based upon the assumption of perfect competition must be qualified to the extent that it does not work as assumed.

Because competition does not work perfectly, shall we cast aside all the reasoning and conclusions of the past on this subject and begin anew? The answer is in the negative for the following reason. No one will deny the existence of some competition in modern economic relations; nor is there any evidence that in

¹ This statement leaves out of account the fluctuations in the value of money and its effect on price, a topic which receives consideration in Chaps. XI, XII, and XIV.

the near future competition will cease to be an important force in our economic life. We may, therefore, conclude that so long as, and as far as, competition operates, its effects are as indicated, or, to put it in other words, the pull of competition is in the direction of the conclusions reached by the early economists. It tends to make market prices the measure of wants, and to compensate each factor in accordance with its importance in production.

In recognizing the force of competition, the modern economist does not need to assume individualistic conditions that are unnatural. He is dealing with men and women who are born into, and are living in, normal social conditions and whose actions are molded very largely by the social life of their group. By way of illustration, take intercollegiate sports. There are few fields with which college students are familiar in which competition and rivalry are keener than in that of football. The breaking of bones, the gouging of eyes, or the slugging of opponents are not essential conditions for the existence of competition in this sport, even though unrestrained competition might lead to such practices. In fact, intercollegiate competition takes place within the rules of the game that have been accepted in advance of the play. The players strive to win but must conform to these rules. In like manner, the economist, in advancing competition as a force affecting economic life, does not need to assume that it is unrestrained, but rather that it operates within the limits imposed by the social life of the time.

Similar statements can be made concerning mobility of the factors of production as have been advanced concerning competition. While the factors of production—land, labor, and capital—do not, and in fact cannot, flow perfectly from one occupation or use to another, yet no one will deny that there is some transference of factors between industries. For example, land may be used for agriculture or manufacture, and when it has once been utilized in one way it tends to become specialized; yet there is some shifting, as from corn to wheat, from grazing to grain growing, from farming to manufacturing, residential or mercantile purposes. When any of these changes in the utilization comes, the determining factor is the return in the alternative use. If land will bring a higher return in grain growing than in grazing, the pressure of the economic motive will be in the direction of a shift of the uses. In other words, some mobility is found in the use of land.

In the case of labor there is even greater possibilities for mobility. Labor in our own country is comparatively free to pass from one occupation to another, but there is by no means perfect mobility. The section hand does not readily and normally become a railway president. Where skill in an occupation is a factor, the movement from one occupation to another is limited. Then age is also a factor. Few men change their trade or occupation after forty-five years of age. The mobile elements in the labor supply are found in the rising generation. Those entering industry for the first time have a chance to select their trade or profession with comparatively little hindrance. In making their choices, opportunity for gain, though not the sole motive, is an important one. While, therefore, mobility of labor is not perfect, it is present and the effect is to shift labor in the direction of the highest economic gain.

Likewise, in the case of capital, when it is considered as a factor in production, there is only a limited sense in which it can be said to be mobile. Capital, like labor and land, becomes specialized. A machine adapted to a cotton mill could scarcely be shifted to an automobile factory. It has a specialized use, and though some machines may be shifted to similar processes in another industry, this kind of shifting is narrowly limited. The Eighteenth Amendment to the Constitution of the United States forced large industrial readjustments. Buildings formerly used for breweries have been converted into food factories, canneries, or soft-drink manufactories. Other instances of specialized capital having been shifted to a different use might be cited, yet, generally speaking, it tends to become specialized and, hence, its mobility is limited. Mobile capital, however, appears in the form of capital value, or savings seeking investment. These funds will flow, other things being equal, into the industry yielding the highest returns. In this form capital is very mobile and its flow follows the conclusions from economic analysis, *viz.*, toward the largest gain.

The actuating motive which lies back of these forces is self-interest. The system of industrial liberty as an economic philosophy is based upon the idea that the individual should be left free to follow out his own economic interest. This theory of free, competitive industry is conceived as the most effective method of satisfying human desires and of promoting economic welfare. An economic society, whose actuating impulses arise

out of the motive of self-interest, and whose operating forces are competition and mobility of the factors of production, may be said to be governed by unconscious forces. While admitting the existence of these forces, there are those who argue that conscious control, privately operated, has become more important in the conduct of our economic life, and on this account economic principles should be developed from this point of view. They claim that the principles based upon the assumption of a competitive system are unreal.

But, since no one will deny that competition is still a force to be reckoned with in our analysis of economic phenomena, the question resolves itself into one of expediency. Is it wiser to proceed on the assumption of conscious control and make qualifications for the existence of the unconscious forces, or should the principles be developed from the point of view of spontaneous and competitive forces, making due allowance for the existence of conscious control? A logical statement of the principles could be developed from either point of view, but, since the literature of the subject has been developed on the assumption that competition and mobility are the controlling forces in the present system, it seems wiser to adhere to this approach. No one, certainly, can deny that some competition exists among business men and that there is some striving on their part to promote their own pecuniary advantage. So long as, and to the extent that, these forces are at work in the operation of the system, their effect is to make market price the measure of human wants and the indicator for the adjustment of the productive processes to human needs. In this manner competition serves to organize industry in harmony with a general, social purpose and to compensate each factor in accord with its significance in promoting that purpose.

In conclusion, it may be said that the present economic order is one based upon individual initiative and private enterprise; that the actuating motive is self-interest and that these forces operate through the market in the sense that goods are produced primarily to sell. Furthermore, the system rests upon the institutions of private property and free contract, by means of which the individual is required to respect all of the obligations which he has freely assumed and is protected in the enjoyment of the fruits of his own productive energies. Then, in its outward manifestations, the system is characterized by division of labor

and specialization which become integrated in the market. In its broad, social aspects, the system is cooperative in the sense that the individual, while engaged in producing a private income, is, in fact, contributing to the sum of wealth available for others. This cooperation may be said to be spontaneous rather than predetermined. Likewise the control or the direction of the operation of the system is mainly spontaneous, though conscious control is present in the form that is known as monopoly. The authority for the exercise of control flows from the existence of the right of private property, which gives to the owners of wealth the right to determine what use shall be made of it.

But while the system depends primarily upon private enterprise, the state, expressing the will of the whole social group, limits individual initiative in two directions, either by rendering some services directly, or by establishing rules governing the conduct of the individual. Finally, operating through the market as an organizing influence, is the force of competition, which is not only active in determining the lines of economic endeavor, but has been the basis for the historical development of most of the principles that make up the science of economics. The recognition of the characteristics of the present economic order that have been set forth above is an essential condition for an understanding of the principles of the science.

Suggested References

- BOUCKE, O. FRED, "Principles of Economics," Vol. I, Chap. IV.
 CLARK, JOHN M., "Social Control of Business," Chap. IX.
 CLARK and GIDDINGS, "The Modern Distributive Process," Chaps. I and II.
 CARVER, THOMAS N., "Social Justice," Chaps. I-V.
 ———, "Principles of National Economy," Chaps. V and VI.
 CLAY, HENRY, "Economics for the General Reader," Chaps. VI and VII.
 ELY, RICHARD T., "Outlines of Economics," 4th Ed., Chap. II.
 GIDE, CHARLES, "Principles of Political Economy," Rev. Ed., pp. 335-348.
 HAYES, H. GORDON, "Our Economic System," Vol. I, Chaps. II and III.
 MILL, JOHN STUART, "Principles of Political Economy," Book II, Chaps. I and II.
 SELIGMAN, EDWIN R. A., "Principles of Economics," Book III, Chaps. IX and XII.
 TAYLOR, FREDERICK M., "Principles of Economics," Chaps. II and III.
 VEBLEN, THORSTEIN, "The Place of Science in Modern Civilization," pp. 56-251.

CHAPTER III

FUNDAMENTAL CONCEPTS

Desires as Motives.—The starting point of all human activity which has any economic consequence is the existence of unsatisfied desires. The idea of desire is common enough in human experience to be generally understood. For the purpose at hand, desire means a feeling or longing for some object or end to be attained. This longing may be for food, clothing, ornaments, recreation, companionship, leadership, or for any one or more of the great multitude of ways in which the innate longings of the human being find expression. The absence of the object of the desire is what is commonly understood as want. The economist accepts the existence of desire as a fact and leaves to the psychologist the analysis and explanation of its origin and exact nature. While the psychologists do not agree perfectly in their explanations, some things seem reasonably well established as a basis for the structure which the economist has to build. It is known that all human beings are born into this world endowed with certain common instincts, or with predispositions to act in a definite way as the result of definite stimuli. All experience the appetite of hunger and have the capacity to undergo such emotions as love, fear, and anger. In addition to the psychological equipment with which man is endowed at birth, every individual is surrounded from infancy by a social environment that greatly modifies and develops this primary equipment. Racial customs stand at the head of the list of these influences. What man thinks and what he desires are powerfully affected by the customs and traditions of the social group into which he is born. If one compares, for instance, the habits of dress of Orientals with those prevailing in the United States, he will be convinced that social custom greatly modifies the desire for clothing of persons born into these different social groups. Within the group, much depends upon parental training, the influence of teachers and books, the association of

friends, and the multitude of contacts that play upon daily life as to what will be desired.¹

As just stated, the economist is not concerned so much with the explanation of the origin and nature of desire as with the fact of its existence. The effect of human desires is a problem of importance in economic analysis. We know, by observing the behavior of others and by reflecting upon our own individual reactions, that the existence of ungratified desires is an impelling force to human activity. The pursuit of food, clothing, and shelter occupied most of the time of primitive man and is still a powerful motive force among the civilized peoples of our day. But with the development of civilization, primitive wants have been expanded to include not only the comforts and luxuries of life but also the longing for things beautiful, whether of sound, form, or color. Love of achievement also has developed as the elemental desires became satisfied. With the expansion of desires has come a greater stimulus to the activity necessary for finding the means of gratifying them, and these are the driving forces to economic activity. As economic motives, desires are recurrent and persistent. They impel men to exert effort, to endure hardships, and to sacrifice leisure in order that their wants may be satisfied. In this respect, desires differ from impulses, which may have some economic consequences, but which are usually temporary and transitory in character. Impulses do not give rise to persistent human action such as is necessary for the organization and operation of a modern, complicated business establishment. The complex economic structure of our day has been built up in response to powerful and persistent motives, the basis for which is found in the existence of human desires.²

¹ BRY, R. T., "Principles of Economics," p. 61-70; cf. also DICKINSON, Z. C., "Economic Motives"; PARKER, CARLETON, "Motives in Economic Life," *Am. Econ. Rev. Sup.* March, 1918, pp. 212-238; WATSON, JOHN B., "Behaviorism"; THORNDIKE, E. L., "The Original Nature of Man."

² Marshall makes a slightly different statement. He says, "Speaking broadly, therefore, although it is man's wants in the earliest stages of his development that give rise to his activities, yet afterwards each new step upwards is to be regarded as the development of new activities giving rise to new wants, rather than of new wants giving rise to new activities." "Principles," 4th Ed., p. 164. There can be no denying the force of this interpretation of human experience. However, mere activity is not sufficient to explain expanding wants; or the large volume of goods made to gratify our

Classification of Desires.—Desires may be classified into three main groups. First, there is the desire for *material goods*. In primitive society the wants of mankind were limited in character and were confined mainly to the physical necessities. The pangs of hunger drove the primitive man to search for food and the severity of climate made him strive for clothing and shelter as protection against cold and rain. The desire for these goods and others like them is primary and may be classified as desire for material goods.¹ In this condition man had not progressed much beyond the lower animals and, like them, he was dependent upon his environment for the gratification of his desires. But with the passage of time man has proven his superiority over the lower animals by shaping his environment to his own needs. He has progressively adapted natural conditions to suit his purposes. In this process the desire for expensive and luxurious material goods has played a large part in shaping his efforts.

Following close upon the desire for physical necessities is the longing for the admiration of others. The primitive man's love for decoration finds its modern counterpart in the fondness for display. The *desire for esteem*, or the regard for the opinion of others, is a powerful motive not only for the development of a common social type, but as a force to promote economic action. This desire may take the form of a wish to stand well intellectually, socially, or economically. A young man or woman may wish to give a good report of himself in his classroom or in his chosen profession. He may wish to be recognized as the leading physician or lawyer in his community, and this desire will spur him on to the accomplishment of his purpose. In social relations one may wish to be known as the most genial host or hostess in the neighborhood, or to be a member of the exclusive

desires. Activity must result in a good that is closely enough related to human experience in some of its many manifestations as to call forth a desire for it. It is true that fads, or fashions, do gain sufficient momentum to create within the individual the desire to conform to the conventions of the hour. In this sense it is correct to say that the desire is created by the producer, or whoever starts the fad. But even when this concession is made and full recognition given to created desires, the fundamental truth still holds, that desire is the basis of our economic system, motivating individuals and stimulating them to productive activity.

¹ CARVER, T. N., "Principles of National Economy," Chap. II; cf. also FETTER, F., "Principles of Economics," Chap. II.

social group, and this desire will stimulate the individual to strive for this recognition. Others may wish to be known as the wealthiest persons in the community, and strive to accumulate large holdings. The desire for esteem is ordinarily a mixed one, as esteem itself usually arises from a variety of causes. The applause of an audience may please the actor or opera singer but, if it were not for the salary or share in the receipts, this applause would soon lose its pleasing effects. In such cases the stimulus to action arises from the combined desire for esteem and for the money income with which material goods may be purchased.

As soon as physical needs have been gratified there arises a *desire for action*. Action undertaken for its own sake is called "play." The play spirit is deep seated in man, and is probably a common heritage with the lower animals. Who has not watched with interest a well-fed animal display its feelings for action in playful antics? The love of sports is man's expression of his desire for action. While this desire may be considered separately as a motive, it is frequently combined with an aspiration for esteem, as when there is an opportunity to show physical prowess. In football or other athletic sports, in addition to the love of the game, is the desire for the applause from the grand stands and for the glowing reports on the sporting page of the daily press. Many other instances of a combination of desires as motives could be cited. In fact, motives usually are mixed, but "love of the game" even in business plays an important part as a stimulus to economic activity. By way of illustration, one financial house or railroad group will try to outplay another, and gain supremacy in the industry. In this contest there is more than the pursuit of money profits acting as a stimulus and the game spirit is certainly a potent influence.

If human desires could be gratified by the means secured through the play motive, there would be no large and complicated economic structure. No compensation would be necessary other than the pleasure from the action itself. But such is not the case. As soon as action is carried beyond the point of the pleasure derived from the activity itself, it ceases to be play and becomes work. Work, then, may be defined as any action which requires some form of compensation other than the pleasure derived from the action itself. However, the play motive is present in many forms of work. The pleasure which a man gets from his trade or profession is frequently an important

stimulus to activity, but this motive must in all cases be supplemented by the hope of a money income or other valuable remuneration. Economics deals principally with that type of action which requires some form of compensation outside of the pleasure of the action itself. The stimulus of unsatisfied desires, whether elemental or those associated with the higher ranges of mental pleasures found in the pursuit of learning within the fields of literature, history, science, philosophy, or art, has driven man to build the complicated industrial structure which is apparent on every side and which occupies the major portion of the time and energy of mankind.

Goods Defined.—In the preceding paragraphs, desires have been shown to be the actuating motives in the development of economic activity. Economic principles likewise can be traced to the same starting point. The order of development is first, the desire; second, the effort required to secure the means of gratifying the desire; and lastly, the enjoyment itself. Thus it will be seen that economic analysis starts in the subjective reactions of the individual and in the realm closely associated with the consumption or use of goods. It then proceeds into the objective realm, or that of the production of the various means of gratifying our desires, and finally comes back to the subjective realm in the consumption or enjoyment of the goods produced. The analysis of these relations has given rise to the use of certain terms which have come to have specialized meanings which the reader must master before proceeding further into the subject. We have the term "goods," which may be defined as any means of gratifying a desire. As here used, "goods" is an inclusive term embracing both material and immaterial things. A service that is enjoyed directly, as the pleasure derived from a well-rendered piece of music, is a good in the economic sense just as much as a concrete, material object, as a pair of shoes. In both cases, the means of gratifying the desire is called a good; the first an immaterial, and the second a material, good.

Utility.—All goods have one common quality, namely, utility. By utility is meant the capacity to gratify a desire. While this term is frequently used in another sense, it should be distinguished at the outset from the specialized meaning here given to it. When we speak of a public utility, or a utility, we are using the word in a substantive sense. A street railway may be properly referred to as a public utility. But as a concept in

economics, utility expresses an attribute of a good. The quality of a good which enables it to gratify a desire is the idea contained in the term "utility." For instance, a pair of shoes may possess a number of qualities as pleasing appearance, comfort, durability, etc., which give to them the ability to gratify the desire for shoes. The quality, whether it is a single element or a combination of elements that gives to a good the capacity to gratify a desire, is known as "utility." Briefly stated, then, utility may be defined as the capacity of a good to gratify a desire.

This definition seems to imply that utility is an inherent property of a good, but, rightly understood, the term is used to express a relation between a good and the state of our desires, which change from time to time.¹ For instance, a second pair of shoes, like those just mentioned, may be made from the same material and on the same last and still have less utility for the purchaser who has just bought a pair, because his desire for shoes for the present has been partially gratified. Hence, the capacity of the additional pair of shoes to gratify his desires has declined, not because of any change in the good itself, but because his desire for shoes has changed. Utility, then, as an economic concept must be regarded as an attribute of a good which is reflected from the state of the desires of consumers. We may think of the utility of a good increasing or decreasing when, in fact, the change is not in the good itself but is in the purchaser and represents a subjective change affecting the intensity of his desires. Every business man has seen the demand for his products change because of a sudden variation either in style or in method of use. Instances of such shifts in demand are frequent and help to emphasize the fact that utility, after all, emanates from the desires of the purchasers rather than constituting an inherent quality of the goods themselves.

Classification of Goods.—Goods may be classified into two principal groups, *free goods* and *economic goods*. A free good is one which exists in quantities sufficient to gratify all known

¹ GIDE, C., "Political Economy," pp. 40-44. Pareto suggested the Greek word, "ophelimity," which expresses the relation of fitness between an object and a desire, but this term was not accepted. Gide suggested "desirability" which places the emphasis on the subjective condition rather than on the object. There is much to be said in favor of the latter term but it has not been widely accepted. The student, therefore, should understand that utility as used in economics implies a relation between the state of the desires and the goods evaluated.

human desires for it. Under ordinary circumstances air is a free good. No one is willing to put forth effort to secure any particular unit of air, for if this one unit is not available, some other one equally satisfactory can be secured. Free goods may possess a high degree of utility, but they exist in quantities sufficient to satisfy all known desires for them. There is, therefore, no economic problem in connection with free goods. An economic good is one which does not exist in quantities sufficient to gratify all known human desires for it. It has the common element of utility with free goods but has the distinguishing feature of *scarcity*. Scarcity does not mean that the good is rare, but only that there is not enough to go around for those who desire it. All economic problems have grown up around this latter group of goods.

A free good may become an economic good if the quality of scarcity should develop. In the case of air, which is usually a free good, scarcity may arise, as in large buildings, in the basement of department stores, or in many other places that will readily occur to the reader. When the supply of fresh air becomes inadequate to maintain health, expensive ventilating machinery is installed. While the air itself is still free, in order to get it supplied where needed, a cost must be incurred. In this sense it may be said that air has become an economic good.

Another condition that requires some distinction is where a good is furnished free to the individual but involves a cost to society as a whole. For instance, a public park is free to the individual. He is free to make use of its facilities as he sees fit so long as he observes the general regulations governing its use. But, while free to the individual, the furnishing of such goods involves a cost to the governmental agency that supplies it, and it is, therefore, an economic good from the point of view of society. Such goods have been called *public goods*. It should be noted that public goods are a special form of economic goods.

Goods may be classified also by the use to which they are put. There are those which are destined to gratify desires directly, such as food, clothing, and shelter. Such goods are called "consumers' goods." Then there are those which never gratify desires directly, but which are always used as aids in producing consumers' goods, such as machines, tools, factories, warehouses, etc. They are said to yield utility indirectly. These are called "producers' goods." A question may arise as to the classifica-

tion of raw materials which eventually become consumer's goods. By applying the test given as the basis for distinguishing consumers' and producers' goods, namely, whether the utility is derived directly or indirectly, it seems clear that, until the productive processes are completed, such goods should be regarded as producers' goods.

Satiability of Desires.—Three observations can now be made concerning desires and the use of goods. As an individual consumes successive units of any good within a given period of time, the intensity of his desire for it declines. By way of illustration, take some simple but familiar article of food, say apples. If one consumes within a short period of time one apple after another, very soon his desire for apples will diminish even though subsequently his appetite may revive and be as active as before. What is true of apples is true in a greater or less degree in connection with his use of any other good. The tendency here described is called "diminishing intensity" of desire. Following this illustration further, should the individual continue to consume apples, his desire for them will become completely satisfied. Here again, it may be said that what is true of apples is at least relatively true in the case of his use of any other commodity. This observation may be called the "satiability of a particular desire." But, ordinarily, consumption is not carried to the point that no further satisfaction can be obtained from additional units of a good. What actually happens is that an individual will continue to consume units of one good until his desire for some other one becomes more intense. At this point he will shift his consumption to the good for which the desire is then most urgent. It seems also to be a characteristic of human nature that as one desire becomes gratified a new desire arises. To put this idea to the test, can you think of any one whose desires are completely satisfied? An examination of the behavior of others or reflection upon one's own experience will reveal that desires rise faster than the ability to find the means for gratifying them. This observation has led to the statement that *desires in general* are insatiable. From this analysis these three general observations may be recognized; namely, the tendency for the intensity of desires to diminish as the result of continued stimulus, the satiability of any particular desire at a particular time, and the insatiability of desires in general.

Balanced Desires.—The purpose of economic activity is to provide a flow of economic goods. The larger the volume of goods available for the satisfaction of human desires, the greater will be the contentment of mankind so far as human contentment is dependent upon economic considerations. To put the same idea in other words, the more completely desires are gratified, the greater the economic welfare of man. Maximum welfare is attained by having balanced desires, that is, by carrying the gratification of desires in one direction until the satisfaction derived is equal to that received in other directions. Wisdom in the choice of economic goods counsels that income be so distributed that there will be equal degrees of satisfaction of all human desires, or in other words, that there will be an attempt to maintain a balance in the intensity of desires for the various kinds of goods.

Wealth and Services.—Another term that is commonly used in economics is *wealth*, which may be defined as a stock of economic goods existing at an instant of time. It will be apparent from this definition that wealth is a collective term and is always used in connection with economic goods as distinguished from non-economic, or free goods. The term is synonymous with economic goods, but is used when the emphasis is placed upon a stock of those goods. Care should be exercised to distinguish this usage from that appearing in the writings of the early economists where wealth usually meant concrete, material goods.¹ The same problem is confronted here as was met in connection with goods, namely, that wealth may be either material or immaterial in form. Material wealth includes tangible economic goods, such as land, wheat, flour, bread, etc., while immaterial wealth includes intangible economic goods, such as the services of a physician, lawyer, artist, or any other person whose services are enjoyed directly. The problem presented by this distinction is whether personal services should be regarded as wealth. Some modern writers have distinguished between wealth and services, applying the former term to material goods and the latter to immaterial goods. The fact that services disappear with the act

¹ For an exposition of this concept of wealth, cf. MILL, J. S., "Principles of Political Economy," Book I, Chaps. II and III, par. 1-3; FISHER, I., "Elementary Principles of Economics," Chap. I, par. 1. For a different treatment, cf. MARSHALL, ALFRED, "Principles of Economics," 4th Ed., Book II, Chap. II.

that gives rise to them has led many authorities to confine the term "wealth" to tangible goods. But when we take the point of view of an instant of time, we can at least conceive of the existence of a stock of services. Lawyers, doctors, teachers, and all others whose efforts result in a direct gratification of a desire are rendering services which may be regarded as a stock of intangible wealth, even though the stock is a most perishable thing and does not remain fixed except for an instant of time. While this distinction is a problem of classification, and a logical treatment can be presented from either point of view, in this text wealth will be used in the inclusive sense embracing personal services as immaterial wealth. At times for the sake of clarity and ease of presentation, the term personal services will be used in place of the expression immaterial wealth.

While this definition of the term departs somewhat from the ordinary usage, it has the advantage of including all things of value. We regularly pay for services of various kinds that do not result in concrete tangible form, and these services are frequently very highly prized as in the case of movie stars, opera singers, great surgeons, lawyers, and persons serving in other similar lines of human endeavor. Such services contribute very greatly to the well being of the human family and should, therefore, be included as a part of the wealth of society. Even from the point of view of an individual, the right to receive such services might be regarded as a part of his wealth, as in the case of contracts with star performers, whether in opera, baseball, the movies, etc. However, since such services do not become embodied in concrete goods, it is necessary to take an instant of time as the basis of classification if services be thought of as a stock. By so doing wealth can be made a general term and include both material and immaterial goods that possess value.

Characteristics of Wealth.—There are three characteristics of wealth that should be noted. First, to be wealth the good must possess utility. It must have the power to gratify a desire. Second, it must be scarce, that is, it must exist in quantities less than sufficient for known desires for it. These two qualities, it will be observed, are the same as those given for economic goods. Third, a further distinguishing feature is that of externality. Every man has faculties, mental and physical, that may enable him to contribute to the supply of economic goods. By the exercise of these faculties he not only produces goods

but secures command over those which he himself consumes. The question then is, Are these inherent faculties wealth? To put the case concretely, Is the skill of a carpenter wealth? We know that by the exercise of his skill the carpenter earns his living, but does this fact prove that the skill itself is wealth? By definition wealth is made to refer to a stock of economic goods existing at an instant of time. The question whether personal qualities, such as skill, should be included within the concept of wealth turns upon the classification of economic goods, and we have seen that an economic good is anything, material or immaterial, that has the capacity to gratify human desires. Do we, as a matter of fact, consume skill, or do we consume the products or services rendered by the aid of skill? Is it possible to distinguish between the possession of a faculty, like the skill of the carpenter, and the service rendered by the exercise of this faculty? What we buy when we employ a carpenter is not his skill but the commodity or the service which his skill enables him to produce. When we call a physician we do not buy the man, nor his knowledge of diseases and cures. What we pay for is the service which his knowledge, skill, and experience enable him to render. He still possesses as much skill and ability to serve as before we employed him. In fact, his capacity to serve may have been increased by his experience with the case at hand.

The point that is being stressed is that personal faculties, whether mental or physical, are inseparable from the individual and are, therefore, not economic goods, but the services which the possession of these faculties enables one to render are economic goods and are, therefore, wealth. The distinction here made is not only in accord with practice, but is necessary as the basis for sound social policy in connection with certain economic relations. Man is not wealth, except in a state of slavery. He is the end for which economic endeavor is undertaken and he should not be regarded as a means. The recognition of this distinction in the fundamental concepts will help to remove a great deal of confusion in our thinking on questions of social policy. Man in a state of freedom should be eliminated from the concept of wealth, and the acceptance of this classification will remove also his personal faculties as a part of wealth because they are inseparable qualities of the man himself. Externality, then, must be included as one of the important characteristics of wealth.

Individual and Social Wealth.—In the treatment of economics there are two points of view that must be kept constantly in mind, namely, the individual and the social. From the point of view of the individual, there are some things which would be regarded as wealth that would not be included from the point of view of society. An individual would reckon not only his stock of material goods as wealth but also his rights and claims to receive wealth. For instance, the individual may own a bond which he would regard as a part of his wealth, not because of the paper on which the contract is printed, but because the bond gives him the right to receive money income with which to purchase economic goods. The terms of the contract convey valuable rights, which to the individual are as much a part of his stock of economic goods as any other valuable things which he may possess. But to society the bond represents a debt on the one hand and a credit on the other. The bond could be completely destroyed without affecting the total stock of economic goods then existing. The loss suffered by one person, the creditor, would be gained by another the debtor. It thus appears that from the social standpoint a bond is not wealth at all, but only a legal claim of an individual to a part of the social wealth. Whenever there are reciprocal obligations similar in character to the creditor and debtor relations, as represented in the ownership of a bond, this distinction between individual and social wealth will appear.

Wealth and Income.—When we look upon economic goods as a stock existing at an instant of time, we use the term wealth, but it is frequently desirable to think of the flow of goods, ripening, as it were, during an interval of time. When emphasis is placed upon a flow rather than a stock, the term "income" is ordinarily used. However, this term has a variety of meanings, the more important of which should be carefully distinguished. First, we have the subjective use. Wealth exists for the purpose of gratifying the desires of mankind, and it is frequently necessary to distinguish between the goods and the benefits derived from them. The benefits or pleasurable sensations arising from the consumption of wealth are psychic reactions, and for this reason the term "psychic income" may be used to denote these benefits. Income in this sense is an intangible, psychological magnitude, but is none the less real. We will later see that the only measure we have of the subjective benefits that an individual derives from the use of wealth is an objective test which consists of what he is

willing to exchange in the form of services or other economic goods in order to obtain it. Some forms of wealth, such as personal services, yield income in a single use, while other forms, such as durable goods, are capable of yielding income through a period of time. If the idea of flow is expressed in connection with income used in the subjective sense, it is tacitly assumed that the wealth is durable in character. For instance, we might refer to the flow of income from the use of a house. Here the house is a durable economic good, capable of yielding satisfactions during a period of years and these satisfactions, regarded as psychic income, may be thought of as a flow. One further thought is that psychic income does not vary directly with variations in the quantity of economic goods, as doubling their quantity will not ordinarily double the subjective income derived from them. This conclusion will be readily recognized as depending upon the principle of diminishing intensity of desires as the quantity of any good consumed is increased.

In the second place, income is much more commonly used in an objective sense to denote the flow of economic goods during an interval of time. Income in this sense may be considered either from the social or the individual point of view. From a social view point it would mean the net addition to the stock of wealth made during an interval of time, as a year. In this sense the term "social income" or national dividend is used to denote the aggregate of goods, material and immaterial, that have been produced during a definite period of time.¹ It is evident that social income is a flow of goods, in contrast with wealth which is a stock, whether this stock be measured by means of an inventory of the resources, or by a monetary expression of its value. In one case, there would be an itemized list of the available economic goods, while in the other there would be a statement of their worth in terms of money. The same condition obtains in the case of an individual. His wealth consists of the total stock of economic goods in his possession, whether listed as an inventory or expressed in terms of value, while his income would consist of the net addition derived from the use of his wealth during an interval of time. In the United States, the habit prevails of expressing the wealth of an individual by means of the value of his possessions. A man is worth

¹ MARSHALL, ALFRED, *op. cit.*, p. 594.

\$100,000 when the value of his possessions is \$100,000. In England, it is customary to express wealth in terms of income, and in this case a man's wealth might be said to be worth \$5,000 per year. From both the individual and the social point of view income has here been used in an objective sense to denote a flow of economic goods.

In this sense income appears in two forms, as has already been implied in the previous discussion, namely, *money income* and *real income*. The most familiar form is that of money income. All persons, except those depending upon others for their support, receive money income either from products sold or services rendered. Some are engaged in business and get profits; others may furnish capital and get interest; still others furnish land and get rent; and those who labor get wages or salaries. With this money income, from whatever source it may arise, they purchase the economic goods which they desire. The volume of goods which their money will buy constitutes their real income. It is obvious that these two forms of income are not identical. Money will buy more goods at one time than at another, hence real income fluctuates with the fluctuation in the value of money. *Psychic income*, on the other hand, varies only in accordance with the urgency of the desires, which implies a relation between the quantity of the goods available for consumption and the intensity of the desires for them.

In ordinary business operations the term "income" is used in the sense of money income, and refers to the receipts which a business firm derives from its transactions. The total receipts from the operation of the business are called *gross income*, while *net income* is what is left from gross income after all expenses are paid. There are many accounting problems that arise in connection with the use of these terms which need not be discussed at this stage of our treatment. The important point to note here is the fact that the owners of a business receive no income until after the expenses of operation have been paid.

Finally, income is used in connection with the shares which individuals receive as owners of the factors of production. The division of the products of industry among those who have participated in their creation is a problem of distribution. The amount of the total product which the owners of each factor receive is regarded as a share of the social income or national dividend. At this point, we get back to the relation between

the subjective and objective uses of the term. The mere ownership of economic goods, whether in the form of money income or real income, is not the ultimate end in view. The real purpose of all economic endeavor is the satisfaction derived from the use of the goods. True income is subjective, but in the ordinary transactions of business we are usually concerned with income in its objective sense as money income. This fact should not obscure the more fundamental economic relations in connection with income in its subjective sense as the benefits derived from the use of wealth.

Wealth and Property.—When wealth has been produced a very important question arises as to who shall enjoy the benefits flowing from its use. The right to enjoy wealth is known as a property right, and, as distinguished from wealth, property always refers to the legal right to use wealth. It implies the

Case	Wealth on which the property right is based	Benefits of that wealth	Character of the property right	Evidence of ownership
Manufacturing plant	Factory	Products	Exclusive right in perpetuity	Deed
Street franchise.....	Street	Use of same for specified term	Right to run cars on same	Charter
Lease or hire.....	Store	Use of same	Right of tenant till fixed date	Lease
Industrial bond.....	Plant and equipment	Payment of interest and principal	Right to same and contingent right to foreclose	Bond certificate

legal relation established by society between man and wealth, which we call "ownership." To own anything implies the right to enjoy the benefits flowing from its use and this privilege is established by the legal rights of property. This distinction between property and wealth is of great significance in understanding the operation of our economic system. Business men are engaged in transferring property rights to wealth. Every

sale or trade involves a transfer of property rights, and the determination of the use of wealth is based upon these rights. It is evident, therefore, that our economic system is held together by means of property rights, but the legal relation should not be confused with the thing itself. The main objective is to secure wealth, and property establishes the right to that wealth. Again, the right to use should not be confused with the evidences of that right. The relation of wealth, rights, and evidences of ownership is seen in the preceding examples on page 59.¹ The reader can work out additional examples illustrative of the relations existing between wealth, property rights, and the evidences of ownership. He should observe that all business is engaged in transferring these rights.

Property may be either private or public. When the right to wealth rests with the individual we call it private property, or a private property right. The most evident forms of ownership are of this character. Farms, workshops, stores, and residences are owned by individuals. The growth in size of modern economic establishments has necessitated the development of a method of subdividing the ownership. This has been accomplished through the corporation, which is a form of collective ownership that enables a large number of persons to own an enormous business enterprise by means of shares of stock. In this case the property of each individual is limited to the number of shares of stock which he owns. The importance of private property in the operation and development of control in our present economic system has already been discussed. In addition to being privately owned wealth may be publicly owned. The ownership may be vested in the nation, the state, the county, or the municipality. The evident examples of public ownership are the postal system, public parks, public highways, public schools, municipal water works, and many others similar in character. There is constant demand for an extension of public ownership of such industries as the railroads, the mines and the municipal utilities, such as street railways, and gas and electric-light plants. The advisability of such an extension is a problem of social expediency, which turns on the question as to whether the service will be rendered better and at a lower cost than by private

¹ The suggestion for this table is found in FISHER, I., "Elementary Principles of Economics," p. 33.

ownership. Public ownership is by no means insignificant in our economic life, but when the whole industrial system is under consideration, it is evident that private property is still the fundamental form of ownership.

Wealth and Value.—In the preceding discussion, wealth has been treated as a stock of economic goods at a given instant of time. In all business transactions it is necessary to have some means of measuring the amount of wealth. There are two principal ways by which wealth can be measured. We may, for instance, use a unit of weight or length, or some other common standard of measurement. Thus, we can say that there are 10,000 bushels of wheat, or 5,000 suits of clothes, either of which conveys an idea of quantity and is the result of a measurement of the wealth in question. But this method of measurement will not serve as the basis for exchanging one good for another, hence it has been necessary to develop a second method of measurement. We might have for consideration 10 bushels of wheat and 10 pairs of shoes. In quantity there are 10 units of each, but this does not establish a basis for exchanging a pair of shoes for a bushel of wheat. The question of measurement here involved is, In what ratio will a unit of wheat exchange for a unit of shoes? How many bushels of wheat will be required to exchange for one pair of shoes? To answer this question requires the discovery of some quality by which the social significance of units of wealth can be compared. That quality is known as value. Value is the power which one good has in commanding other goods in exchange for itself. Value, then, is always a phenomenon of the market where exchanges are taking place. The reason why goods have value is because they are useful, *i.e.*, they possess utility. Their value depends on how keen the desire is for the goods in question relative to the supply. If traders are willing to give up many units of one good for a unit of another, we say the value of the latter good is high. For instance, if one pair of shoes will exchange for 5 bushels of wheat, the value of shoes in terms of wheat is large, but the value of wheat in terms of shoes is small. Value, then, as a measure of wealth is always a comparison of the utility of different kinds of goods.

Value and Price.—It is customary to quote the exchange power of economic goods in terms of some one commonly accepted good, which has come to be known as money. When value is quoted in terms of money that quotation is called price. The

price of a pair of shoes, we say, is \$5, while the price of wheat is \$1 per bushel. By this process we have expressed the exchange power of two commodities in terms of a third, thus establishing a basis of comparing the relative utility of the two commodities and also of establishing a basis of exchanging one for the other. Value and price are very important concepts that run throughout all economic analysis. Further treatment of these subjects is postponed to a later part of this text.¹

Caution in Use of Terms.—In the preceding paragraphs attention has been given to the content of some of the terms most commonly used in this field of learning. While it is highly desirable to seek precision and especially clarity in the use of terms, caution should be exercised in order that definitions may not become too rigid. The danger, on the one hand, is that words will be used so carelessly as to cause confusion of thought, and, on the other, that their meaning will be so narrowed as to render the presentation of the subject unreal. Economic phenomena constitute a continuous flow of human experience, and a too rigorous classification may cause a breach in that continuity. Most distinctions in economics are differences in degree rather than differences in kind, and this fact should be kept constantly in mind by the reader. Yet in the presentation of the subject, especially for beginners, a degree of definiteness and precision in the use of terms is perhaps justified as an aid in developing correct habits of thinking in this field. Qualifications and fine distinctions can be made much more effectively after one has developed an understanding of the complex relations existing in our economic life. The reader should accept the terms as defined and watch for two things, namely, the shades of meaning that may appear from the context and by means of qualifying words or phrases, and the effectiveness of the terms in describing the facts and experiences of life. Unless they assist in explaining and interpreting these experiences, they fail in their main purpose. As Marshall puts the case:

We shall thus generally find that there is some use of each term which has distinctly greater claims than any other to be called its leading use, on the ground that it represents a distinction that is more important for the purposes of modern science than any other that is in harmony with ordinary usage. This may be laid down as the meaning to be given

¹ Chapters VIII, IX, and X.

to the term whenever nothing to the contrary is stated or implied by the context.¹

If the reader exercises due caution in following the distinctions made, he should experience no serious difficulty with the usage of terms, even though it may depart somewhat from that with which he is familiar.

Suggested References

- BYE, RAYMOND T., "Principles of Economics," Chap. IV.
CARVER, T. N., "Principles of National Economy," Chaps. II-IV.
DICKINSON, Z. C., "Economic Motives."
ELY, RICHARD T., "Outlines of Economics," 4th Ed., Chap. VII.
MARSHALL, ALFRED, "Principles of Economics," Book II, Chaps. I-IV,
Book III, Chaps. I-VI.
PARKER, CARLETON, "Motives in Economic Life," *Am. Econ. Rev.*, Sup.
March, 1918, pp. 212-238.
RUFENER, L. A., "Principles of Economics," Chap. I.
TAUSSIG, FRANK W., "Inventors and Money Makers."
THORNDIKE, E. L., "The Original Nature of Man."
WATSON, JOHN, "Behaviorism."

¹ MARSHALL, ALFRED, *op. cit.*, p. 22.

CHAPTER IV

THE NATURE OF PRODUCTION

Production Defined.—The starting point of economic analysis, as has been pointed out, is the existence of unsatisfied desires. Since nature does not freely supply the means of gratifying these desires, mankind has undertaken to fashion goods and to make them available for this purpose. This preparation of goods is what is known as "production." The processes employed might be considered from a technical point of view and an examination made of the physical and mathematical principles involved. But while these technical principles are important in developing productive methods and are of primary significance in the field of engineering, they contribute nothing to the science of economics, as we are concerned in this subject with the results flowing from the application of these principles rather than with a consideration of the principles themselves. [We are dealing with the relations existing between men and goods and not with the natural forces as forces or the manner in which they may be utilized in the production of those goods. In the preparation of goods, it must be evident that man creates no new elements or substances. What he actually does is to take the raw materials of nature and adapt these to his uses. To be useful, goods must possess utility, hence, production may be defined as the creation of utility.

Methods of Creating Utility.—A cursory examination of business will show that utility may be created in more than one way. On the one hand, we see men engaged in fashioning the raw materials of nature into a form that is better adapted to gratify our desires. For instance, the carpenter with his kit of tools proceeds "to make" a house out of the raw material—lumber. He changes the form of the raw material, puts it into a shape that will gratify the desire for shelter, or any other desire that may be gratified by the use of a house. All who are thus engaged are said to be producing form utility and are changing the form of things so that they will be better adapted to the uses of man-

kind. The significance of production in this sense can be realized when it is remembered that all of the processes in the factories and workshops of the country are employed in changing the form of things, and may, therefore, be said to be producing form utility.

We note also that some men are engaged in transporting goods from one locality to another, from the place in which the desire for the good is less keen to one in which it is more keen. This type of activity is called the production of place utility. As an illustration, the railroads carry wheat from the harvest fields of the Dakotas, where wheat is plentiful, eastward to the flour mills and other markets, where wheat is relatively scarce. The railroads, we say, create place utility. Then, we have those persons who hold goods from one season to another, thus adding to the want-gratifying power of the goods in question. In winter, ice is plentiful enough and few people desire it at that time, but during the hot summer months the desire for ice is much more urgent. Hence, we say that those engaged in holding goods from a period in which the want-gratifying power is low until such time as the desire becomes active, create time utility. The storage business particularly is devoted to the creation of time utility.

By this classification, it is not intended to imply that all forms of productive activity can be brought easily into one or the other of these groups. Often a business concern will produce both form and time utility, as, for instance, when agricultural implements are made in advance of the season of active sale and use. Careful analysis of business activity will reveal other combinations, and in fact, most business establishments are likely to contribute to utility in more than one way. Then, there are some forms of business, such as that of merchants, brokers, bankers, opera singers, teachers, etc., that may not be easily classified into the above groupings, but a little careful discrimination will show that practically all production results in the creation of utility in one or more of these three ways.

Factors of Production.—The creation of utility involves the combination of a number of factors. For instance, to produce any simple farm product, such as corn, requires an area of land which contains the chemical elements necessary to feed the plants. Then, there must be rainfall, sunshine, and labor and tools to cultivate the plants. All of these agencies, or elements used in

production, are called "factors" of production, and for the purposes of this study we may classify these factors into land, labor, and capital.¹

Land.—By land we mean all the natural forces which assist man in his productive efforts. Not only the soil and its inherent fertility, but the configuration of the surface, the moisture, the sunshine, and any other elements that contribute to its productiveness are included in the concept of land. In addition, the term includes the power developed at water-falls, the electrical energy that is harnessed for man's purposes, coal and mineral deposits, and all other natural resources that contribute to the production of wealth. Land, then, in an economic sense, includes the superficial area with the appurtenances thereto, such as rainfall, sunshine, temperature, etc., and in addition, all natural forces, elements, or conditions which make it possible for man to add to the supply of the goods available for his use. From this definition it is apparent that land is a most fundamental factor in all production. The materials of nature out of which all goods are made; all the natural power that is used to supplement human energy in the productive processes, whether steam, electrical or water power; the space necessary for the location of industrial plants, whatever may be their character, are all included in the concept of land as here defined. The significance of land as a factor of production has led one author to say, "In a sense there are only two agents of production, nature and man."²

From this brief statement it will be evident that there are many kinds of land. Among these we may mention agricultural, mineral, forestry, fisheries, power, whether electrical, water or steam, and urban with its many and varied uses. This classification is intended to be suggestive rather than exhaustive, but it

¹ Some writers make the entrepreneur a separate factor of production but I have classified his services as a form of labor. A different classification can be found in A. R. and E. M. BURNS, *The Economic World*, p. 22 ff. These authors list four factors as follows: Land and natural resources, labor, material equipment, and nonmaterial equipment. They object to the use of capital as a factor for two reasons. The term sometimes means money, which does not take part in production, and secondly, it is not usually used to include "all knowledge that has been accumulating since the history of man began." They stress the significance of this knowledge as a very important factor in production.

² MARSHALL, ALFRED, "Principles of Economics," 4th Ed., p. 214

helps to emphasize how important land is as a factor of production.¹

Labor.—Labor, as a factor in production, includes all human energy, other than play, whether mental or physical, that goes into the creation of utility. Very little wealth is available for man that is not the result of his thought and energy. Nature supplies very few goods in the form in which they are consumed. Before the raw materials of nature can be used by man, human effort must be spent on them in order to give them the shape or form that will enable them to gratify human desires. All such human effort, physical or mental, is included in the concept of labor.

Productive and Unproductive Labor.—The earlier economists held that only labor which resulted in a concrete material good could be considered as productive. Teachers, lawyers, physicians, and all others whose efforts did not become embodied in a material good, were regarded as unproductive. This view was held even though the work of the latter group was recognized as highly important for the welfare of society. The following quotation may be taken as illustrative of this line of argument. Mill says:

I shall, therefore, in this treatise, when speaking of wealth, understand by it only what is called material wealth, and by productive labor only those kinds of exertion which produce utilities embodied in material objects.²

In the application of this idea, Mill was liberal, and included exertion which did not of itself become embodied in a material object so long as it was a necessary step in the production of material wealth. Thus, with him the acquisition of skill was productive labor, if the ultimate purpose was the production of a material good.

This idea of productive labor has persisted in the writings of economists and is still held by many persons. The acceptance of this view leads to erroneous conclusions which frequently are made the basis of unwise economic and social policies. It is an easy step from the acceptance of this concept of production to the assertion that only physical or manual labor is productive,

¹ ELY, R. T., "Characteristics and Classification of Land," pp. 19-34; 42-49, 1922. (Mimeographed by Edwards Brothers, Ann Arbor, Michigan.)

² MILL, J. S., "Principles of Political Economy," Book I, Chap. III, par. 3.

and, with this latter notion in mind, the conclusion is often implied, if not definitely asserted, that the manual laborer is exploited if any of the income from industry goes as compensation for other than manual work. The error in this reasoning can be traced to a concept of wealth that is too restricted for the purpose of explaining the economic phenomena involved. By expanding the term "wealth" to include both material and immaterial goods and by defining production as the creation of utility, the difficulties which the early economists encountered can be avoided. The test for productive labor is to inquire whether there has actually been a contribution to the supply of economic goods. If the supply of economic goods has been increased as the result of the human effort put forth in bringing them into existence, then the labor is productive. A net addition has been made to the utility available for the gratification of human desires. But if there has been no addition to the supply of economic goods, then the labor is unproductive.

The acceptance of this definition of productive labor enables us to dispose of the question concerning the work of teachers, lawyers, physicians and other mental workers. So long as these persons actually contribute to the gratification of human desires, their efforts are productive. They add to the stock of economic goods. It follows, then, that the ~~compensation~~ given for this effort is a payment for a service rendered. But is the work of the manager and office force of a business productive labor? It is quite common around factories to use the terms productive and unproductive labor to distinguish between labor that can be definitely assigned to the product, as the work of a machinist, and that which cannot be so assigned, as the work of the clerical force. In the one case, there is a direct relation between the effort and the output of the factory, while in the other, the relation is indirect. This indirect labor is usually referred to in the plant as "overhead" or "burden." This "overhead" is always included as a part of the costs of operating the plant, the total costs being divided into *direct* and *indirect* costs. Whatever may be the wisdom of this usage of the terms "productive" and "unproductive" in the plant and for the purposes of the plant accounts, from a social point of view the work of the office force is as essential in the operation of the business as that of the man at the machine and should, therefore, be regarded as productive.

Is there, then, any unproductive labor? As stated above, if effort does not add to the stock of economic goods, it is unproductive. Applying this test, it is evident that the gambler, the swindler and the thief, while they may exert both physical and mental energy, do not contribute to the supply of economic goods. Their efforts are predatory, *i.e.*, they take from others without giving in exchange an equivalent service.¹ Such cases are clear enough. But there are many lines of business activity that require careful discrimination. A stock illustration is the manufacturer of a patent medicine that is harmless but possesses no remedial powers. Suppose by widespread advertising a demand is created for this medicine, is this labor productive? Does not the advertising create a demand for this compound and, therefore, is a desire not satisfied by its use? If the conditions are as stated, the consumer is deceived and, because of his ignorance of the contents of the compound, he buys an article that can contribute no satisfaction. The exchange does not involve a *quid pro quo*, hence all the effort required in making and selling the so-called medicine, including the work done in connection with the advertising campaign is, therefore, unproductive. No equivalent service is rendered. A little reflection will convince the thoughtful reader that there are many lines of business activity which approximate the patent-medicine illustration. The effort put forth to sell securities in fake business undertakings, to promote mining and oil companies that can never be developed into paying businesses, all such activity is clearly predatory, and is, therefore, unproductive from the social point of view, even though a handsome income is derived by the promoters. To be productive from the social point of view, there must be a net addition to the supply of economic goods. If the ultimate end of any business is not productive as tested by this standard, then all effort devoted to the undertaking, whether direct or intermediate, is unproductive.

A careful examination of the competitive methods of modern business will show many lines of activity that are of questionable economic significance when tested by the above standard. The discriminating student will find that frequently there is a conflict between the social and individual points of view when he

¹ This statement is true except where the professional gambler is contributing to the love of games of chance, which is a deep-seated desire in human kind.

observes certain well-recognized forms of business practice such as advertising.¹ From the point of view of the individual business, it is easy enough to show that increased sales result from skilful advertising. Yet, a gain by an individual business does not of necessity coincide with a general gain to society. It may happen that what one firm gains another loses, or the gains of one industry may be at the expense of some other industry. If all firms in a given industry conducted advertising campaigns of equal intensity, or if the advertising of all firms in all industries were equally effective, it does not follow necessarily that there would be a net addition to all firms.

Take a concrete case that will show the limits of the productivity of such business effort. Daily we see merchants appealing to consumers in the most alluring manner, urging them to buy now, while financial institutions at the same time are appealing to the same persons, emphasizing the importance of saving and portraying the merits of stocks, bonds, or other means by which saving can be accomplished. The fact of advertising does not enable the consumer to follow both suggestions with equal force. What the consumer spends now, he cannot save and what he saves, he cannot spend now. The only way by which he can both spend and save more is by producing more. Advertising may serve to whet his desires and thus stimulate him to greater productive activity. But the net addition to the supply of economic goods comes from the greater productive activity rather than from the advertising. At best in this case the advertising can only be regarded as a secondary cause of the additional economic goods. For advertising to be productive from a social point of view, it must comply with the standard set above and contribute a net addition to the sum total of economic goods available for the gratification of human desires. Much of the advertising that appears in the operation of modern industry is productive only from an acquisitive or individual point of view. The gains are individual rather than general and would appear much less significant if advertising were of equal intensity and equally effective for all firms and in all industries. To the extent, then, that advertising activity contributes only to acquisitive gains, it may be concluded that it is socially unproductive labor.

It must not be understood from the discussion above that all advertising effort is unproductive labor. Whenever it contrib-

¹ Other forms of business practice could be chosen for illustrative purposes.

utes directly or indirectly to the supply of economic goods, it is productive. By making known to the consumer where he can get the goods that will satisfy his desires and by giving him information concerning them that will enable him to make an intelligent choice, the advertiser is serving both the consumer and the producer. Again, by means of advertising the market may be so expanded as to make possible production on a scale that will enable the unit costs to be lower than if the market were more restricted and, hence, the total costs of operation, including the advertising costs, will be less per unit of product. Advertising may be a less expensive method of marketing a product, in which case the total costs of placing goods in the hands of the consumer would be less than by any other known method of selling them. Under all such circumstances the contribution is indirect, as the savings permit the diversion of some capital and labor into other fields from which a net addition to the supply of economic goods can be made. In these and other ways advertising does contribute to the sum of economic goods and is, therefore, to this extent productive labor.

What has been said concerning advertising as productive labor may be applied with equal force to many other forms of business activity, and it is frequently difficult to draw a sharp line between that which is and that which is not, productive. It is sometimes asserted that the special costs arising from purely acquisitive practices in business are an integral part of the competitive system and are, therefore, a part of the price that must be paid for the operation of the system itself. There is some merit in this statement. In all social and economic experimentation there will of necessity be many losses, but those losses will be more than offset by successful experiments. Otherwise, society would be better off from an economic point of view if the experiments had not been undertaken. An expenditure of effort of any kind that results in a loss does not of itself contribute to economic goods. This is true of any wasteful methods of production. Sound principles of economy will tend to direct all productive energy so that the largest possible net addition will be made to the supply of economic goods.

Careful discrimination must at all times be exercised in the thought concerning productive and unproductive labor. This is a problem not alone for the student of economic life, but for the managers of business as well. The business man must decide

what combination of the factors of production best contributes to the supply of economic goods. Unfortunately, in rendering his decisions he is not at all times guided solely by the social utility created. He may increase his personal income and thereby his power to consume wealth by methods that are purely acquisitive. For instance, if a business man, because of his strategic position, can gain more by limiting his production, his private interest will incline him toward restriction even though a different policy would add more to the total supply of economic goods available for the gratification of human desires. A conflict between individual and social economy arises here, which is currently voiced in the expression, "Production for profits instead of production for service." The reader should recognize this conflict and get clearly in mind that the fundamental purpose of production and industry is to create utility. This distinction is significant also as a basis for formulating a social policy toward business. All regulations of economic activity should aim to harmonize this conflict and to encourage the business manager to make his decisions in accordance with general, social welfare.

Capital.—The third factor in production is capital, which may be defined as that part of wealth, other than land and personal services, that is used as an aid in further production. In the development of the economic life of mankind, capital did not appear as a primary factor. Originally, man and nature were the primary factors of production. Capital has resulted from the application of man's ingenuity in adapting natural forces to his own needs. When considered from the point of view of the evolution of economic life, capital should be regarded as a secondary factor, but if it is considered from the standpoint of its importance in modern industry, it will be apparent that it is of primary significance in all forms of production. As here defined, the most typical form in which capital appears is that of a machine or a tool, but the definition includes also all producers' goods, such as raw materials and partly finished products.

The purpose of capital is to render labor more productive. In fact, the use of capital in production may be thought of as different ways of using labor, and it is evident that labor is more productive when used in some ways than in others. For instance, a workman with a spade could prepare within a day or week much less ground for agricultural crops than if he were equipped with a modern tractor and a gang plow. But before

the tractor and the gang plow can be made available, there must be an expenditure of human effort in manufacturing these implements, and in addition there is a time element involved in the form of waiting. Some one must be in a position to wait, i.e., he must be able to postpone present consumption during the time the machine is being made and put into use. Waiting, then, is a necessary condition for the existence of a machine or a tool. This time element is a very important factor in determining whether the machine is truly productive. It must be possible not only to turn out more units of product per unit of time, but the additional amount must be sufficient to induce those who provide the waiting to postpone present consumption. The waiting involved in connection with the development of capital together with the indirect methods of production have been referred to as "round-about" processes of production. Speaking generally, the "round-about" processes are more productive than direct ones. This fact is perfectly evident when we think of the work that can be done by means of a spade and by a tractor-drawn gang plow. The additional work that can be done by the use of implements will more than pay for the cost of manufacturing them, including compensation for the waiting necessary to bring them into existence.

Capital and Land Distinguished.—The question may be asked by the business man, why land is not included as a part of capital? He may be accustomed to classify both land and machinery in his accounts as equally a part of his capital and, therefore, be puzzled by the separate treatment found in most economic texts. There are two important reasons for the distinction between land and capital. First, land is not to any appreciable extent a produced good, while capital is always the result of human effort and always involves waiting. It is true that some land is made, as when a swamp is drained, or a river or lake filled in. The area thus made available for use may be thought of as produced in much the same sense as a tool or a machine. The form of the material has been changed from an unusable to a usable shape. However, in the case of "made" land, there are some features that distinguish it from capital instruments. The situs of such land, and the climatic conditions, as temperature, rainfall, etc., that surround it, are not the result of human endeavor. These are natural conditions and are, therefore, not made by man. In addition, this "made" land does not wear

out with use as is the case with a capital instrument. Once a swamp has been drained, or a body of water filled in, it thereafter behaves like natural land. It does not wear out nor does it have to be rebuilt. While we may recognize that some land may be made in approximately the same sense as are machines and tools, the amount of such land in relation to the total area is insignificant. Originally most land was a free gift of nature to mankind.

Sometimes it is argued that the uses of land, as distinguished from the land itself, are produced in precisely the same sense as the uses of any other good. It is asserted, that without adequate transportation facilities, land of any section of the country would have very little use, and that the building of a railway creates the utility of the land that is tributary to it as truly as the manufacturing processes employed in transforming leather into shoes. There is, of course, an element of truth in this general statement, but, like many similar statements, it is only partially true. A little reflection will make it apparent that the land itself must possess certain definite properties quite independent of the transportation facilities. The building of a railway into a desert will not create utility in that land. There are qualities which the land must possess that will enable it to serve some human purpose or the building of a railroad would be a useless undertaking. These are the properties of the land that are to a very large extent the gifts of nature to man and, therefore, it is correct to say that, in the main, land is not a produced good.

The second reason is that the value of machines, tools, and other forms of capital cannot long depart from their cost of reproduction, while the value of land can rise to any figure, depending entirely upon the intensity of its use. The reader can easily verify the truth of this statement. Let him take a plot of urban or city land and follow the growth of its market value through a period of years. If the population has increased and the desire for land has thus been intensified, its market value will have risen. Frequently it will be found in a rapidly growing city that the annual rental value exceeds the purchase price paid by the present owner of the land. The extremely high front-footage value of land in the central districts of any of our large cities is evidence of the truth of the statement that land values continue to rise with the increase in the demand for space.

While it is true that this same influence may intensify the demand for the services of capital instruments and thereby cause an increase in their market value, this increase will, in most instances, be checked not only by the wear and tear of the instruments but also by the fact that their supply can generally be increased. Their value will not depart very far from the cost of reproducing them, whereas, in the case of land, its value is not limited to any appreciable extent by cost, but tends to rise as long as the desire for it increases. These distinctions give rise to important social problems which further enhance the desirability of separate treatment. Questions concerning the form of ownership and the burden of a tax on land, as contrasted with the same problems in connection with capital as will be seen in subsequent treatment, require careful distinction. For instance, it is generally held by tax authorities that a tax on land value cannot be shifted, while a tax on capital tends to become a part of the cost of production and is, therefore, shifted to the purchaser. If for no other reasons, these practical considerations would warrant a separate classification and, therefore, a separate treatment.

Capital Goods and Capital Value.—The use of the word "capital," as defined above, runs somewhat into conflict with the ordinary usage of this term. Among business men the term capital usually means a quantity of value expressed in terms of money. Ask a business man how much capital he has, and he is most likely to respond by stating a capital sum as \$10,000, \$50,000, or some other similar amount. This usage, of course, appears in the accounts of all business firms. Should you ask the same man what he has his capital invested in, or what is the nature of his capital, he would most likely reply by enumerating the concrete forms of wealth used in his business. These two replies call attention to two separate concepts that must be distinguished. Where the attention is centered on the character of the wealth as an aid to further production, we have in mind machines, raw materials, or some form of productive agent. In such circumstances, as an aid to clarity, it frequently helps to use the term "capital goods." This expression emphasizes the material character of capital as an aid in production. At other times, it is desired to express a quantitative measure of capital. This quantitative measure could be expressed in terms of an inventory of the concrete forms of wealth used in production, such as the

number of machines, tools, volume of raw materials, etc., but, if it is desired to emphasize the significance or importance of this wealth, this end can be accomplished by expressing the value of the capital. Again, for purposes of clarity, it is customary to use the expression "capital value," when the emphasis is to be placed upon the quantity of capital in the sense of the social significance of that capital. In this text, capital will be used as instruments of production, but whenever it is desirable to distinguish the goods from their value, the terms "capital goods" and "capital value" will be used.

Division of Labor and Specialization.—No phase of production is of greater significance than the extent to which labor has been subdivided. A trip through any large industrial plant will impress one with the extreme subdivision of labor that prevails. It is said, for instance, that in the manufacture of the small, triangular section—the cutting blade—used on the mower and reaper, there are thirty-three distinct processes, and that this small piece of metal passes through the hands of that number of workmen. Similar examples can be multiplied many times by observing the productive processes in any factory or workshop of importance. If any one wishes to obtain an impression of how significant division of labor is in our modern life, let him make a list of the articles which he consumes during the course of a day, and then ask himself two questions. First, how much did I do toward the production of the goods which I consumed? Second, how long would it have taken me to make the same quantity and quality of goods? Very little reflection is necessary to convince one that both the quantity and quality of the goods used are greatly increased as the result of the division of labor and that human society as now constituted is dependent upon a highly subdivided system of production. Thus, when production is surveyed from the standpoint of society, it appears as a great, cooperative process, an organization by which society combines all of its productive forces and performs the various tasks necessary to furnish its members with a continuous flow of wealth. But closer observation will reveal the fact that the productive processes are subdivided into a multitude of industries, occupations, and professions. From the social viewpoint, therefore, division of labor appears as a cooperative process, but from the viewpoint of the individual it means specialization. His produc-

tive energies are specialized and restricted to a single trade, or, it may be, to a single operation within a trade.

Specialization as it appears in modern industry is possible only in a highly organized social group. In primitive conditions human wants were few and consequently the economic functions involved in supplying them were simple. However, an individual or members of a family living in isolation would have the problem of dividing their time among the diverse lines of work that were necessary in producing the goods that they consumed. There would be some division of labor as between members of the family but, relatively speaking, each member would perform a great variety of jobs. The frontiersman must of necessity be a jack of all trades. In the division of labor of today each workman specializes in a single trade or even in some minute process within an industry, and exchanges his money income for the products of other workmen. It is possible for him to devote his time exclusively to the production of a single commodity or parts thereof, as, say, clothing, because others are producing food, shelter, and other necessities of life. Division of labor, therefore, depends for its existence upon an organized social group and operates by means of the exchanges that take place in the market.

Limited by the Area of the Market.—So long as the division of labor rests upon the direct exchange of commodities, as when one craftsman exchanges the goods that he produces for those produced by some one else, it cannot be carried very far. The beginning of economic specialization takes place when goods are made for sale rather than for barter. The development of trade transformed production for the household or for a local community into production for a general market. The change was first noticed in the woolen industry. In the beginning, no doubt, just the surplus over domestic use was offered for sale in the town or at the fairs where a large number of buyers were present. In this way, gradually, production for domestic use became incidental to production for a general market.¹

As the market area expanded, specialization tended to increase in two directions. First, growth in population in towns and cities increased the number of people who had need of specialized services. Second, by means of the development of improved and

¹ JOHNSON, A. S., "Introduction to Economics," (Rev. 1922), Chap. VII, gives a very satisfactory treatment of specialization.

cheap transportation, the wants of a larger group of people could be reached. The importance of transportation in the development of markets can be seen from the fact that industry and trade began in seaport towns, where a supremacy was maintained until the coming of the railway which enabled inland cities to compete on an equal basis with the centers situated on water routes. It is, perhaps, not an exaggeration to say that the area of the market and, hence, the extent to which specialization can be profitably carried, depends upon the existence of cheap transportation. The revolutionary changes in the organization of industry during the last 150 years have been greatly affected by this cause. Prior to 1800, the market area was relatively very limited. Consequently, we find that at this time the laborer was not only a craftsman but generally a trader also. He had to buy his own supplies and find a market for his products. But with the expansion of the market area, merchants, or specialized traders, began to operate. They soon relieved the craftsman from the task of assembling the raw materials and of marketing the finished products. During the intervening time, transportation facilities have been improved, the market area has been greatly expanded, and specialization has taken on its modern form. The recognition of these interrelations warrants the conclusion that the extent to which division of labor is carried in any period is dependent upon the character of transportation then in existence.

Forms of Division of Labor.—There are three principal types of division of labor to which attention should be given. The simplest form may be called *occupational* division of labor, i.e., when a person or group of persons specialize in an industry or occupation. When men become carpenters, tailors, farmers, lawyers, etc., we say that they have specialized along an occupational line. The term "occupational" is here used in an inclusive sense and embraces not only what is known as an occupation or trade but also what is commonly regarded as a profession. It is customary to think of professions, but from an economic point of view, the work of the lawyer or doctor should be regarded as an occupation. All separation of employment along such occupational lines would be included in this form of specialization. It should be noted that these occupations or professions are carried on contemporaneously and that each person, or group of persons, performs all of the operations of his trade or profession. Each

specialized group completes a productive process and turns out a complete product or renders a complete service.

In dealing with the problem of occupational classification, the United States Bureau of the Census in 1920 made out nine occupational groups as follows:

- Agriculture, Forestry and Animal Husbandry
- Extraction of Minerals
- Manufacturing and Mechanical Industries
- Transportation
- Trade
- Public Service (not elsewhere classified)
- Professional Service
- Domestic and Personal Service
- Clerical Occupations

The purpose of the Bureau of the Census in the grouping used was to emphasize the occupational divisions of labor. This purpose is expressed in its own terms, as follows:

"After careful study of the classifications in use in a number of the principal nations, it was decided that the proper basis for a classification of occupations is the worker and his work, and, hence, that occupations should be classified with respect to the kind of work done or service rendered rather than according to the article made or worked upon, or the place where the work was done. Therefore, it was thought that the best form of classification for the United States would be an occupational classification with an industrial framework. Such a classification, it was believed, would give the most vivid picture of the occupational position of each and every worker and would show the specific services rendered, work done, or processes performed by each worker."¹

When industry as a whole is considered, five main classifications can be made by using some characteristic of the industry as the basis for the classification; namely, extractive industries, genetic industries, manufacturing and mechanical, trade and transportation, and service industries. The extractive industries are those whose raw materials exist in nature, such as mining, lumbering, fishing, hunting, and the utilization of water power. Genetic industries are those that grow their products, such as agriculture, forestry, and fish culture. Manufacturing and mechanical industries are those that take the products from the two previously mentioned industries and transform them into a more usable shape. Any factory or work shop is illustrative of this type. Trade and transportation are those engaged in the moving

¹ Fourteenth Census of the U. S. Vol. IV, p. 10.

of goods from the points of production to those of use. All forms of merchandising and all transportation agencies fall within this form of specialization. Finally, service industries are those engaged in rendering services directly, as domestic and professional service. Here lawyers, doctors, preachers, teachers, and maids, housekeepers, porters, etc. would be grouped. This classification will be found useful when it is desired to contrast or compare industrial occupations in their larger groupings.¹

In the second place, specialization may take the form of a *division of processes*. Within any trade or profession there may be a high degree of specialization. The carpenter may specialize in the laying of hardwood floors, or in some other branch of the trade; the lawyer may become a specialist in the law of patents, in constitutional or international law; the physician may specialize in the diseases of the eye, ear, and throat. All such specialization falls under the division of processes within an occupational group. This form of division of labor is particularly prevalent in modern factories and other industrial establishments where an individual worker may perform only a single, routine operation. In the instance cited heretofore, the small section of the reaper blade passes through the hands of thirty-three workmen. Each workman performs some simple operation

¹ Professor John D. Black, in his *Production Economics* (pp. 65-66) has made the following classification:

Under the following classification:			
Extractive.....	Mining	} Primary production	
	Lumbering		
	Fishing		
	Hunting		
Genetic.....	Water-power utilization		
	Agriculture		
	Forestry		
	Fish culture		
Manufacturing and mechanical			
Transportation and communication			
Storage			
Merchandising—trade			
Household production			
Professional service			
Other personal service			
Public service (not elsewhere classified)			

Professor Black claims for this classification that "it takes up commodity production in logical order, starting in with the farm, the mine and the factory, and ending with the household." He recognizes that the last three classes do not fall strictly within this characterization.

and then passes the material onward until it finally comes out a complete product. While the processes are contemporaneous, they are, in fact, successive steps or stages in the production of a complete product. This form of specialization is typical of modern industry. It is carried to its extreme in such a plant as the Ford automobile factory, or in the packing industry, where the parts travel on a conveyor and each workman performs his task as the material passes in front of him. Much of the productiveness of modern industry is accounted for by the efficiency that is obtained through subdividing the processes in this manner.

Finally, specialization may appear as *territorial division of labor*. Owing to climatic conditions or the character of natural resources, one section of a country or part of the world will specialize in the production of one or more commodities. The specialization that takes place in geographical areas is known as territorial division of labor. We are accustomed to speak of a corn belt, a wheat belt, a cotton belt, or some other specialized region. These phrases are in fact simply condensed expressions to indicate that specialization in the growth of particular crops has taken place in the sections of the country referred to. We know also that tropical products are confined to definite geographical areas; that tea and coffee can be grown more readily in some parts of the world than in others; that mineral products are more generously deposited in some localities than in others. From a variety of causes there is created a division of labor and a localization of industry within definite territorial limits. All such specialization is included in the term territorial division of labor.

The classification of the division of labor given above is reasonably inclusive and will assist in placing most lines of production that will come under observation. For some purposes, it may be desirable to classify division of labor into contemporaneous and successive. At a given time we may find one group of men producing breadstuffs, another meats, another clothing, and so on, each group carrying the raw materials through the various productive processes until they are ready for consumption. Specialization viewed in this way may be called contemporaneous division of labor. But when different sets of men work upon the same material, each set bringing the product one step nearer its final form, the process may be called successive division of labor. This is the form of specialization

that is found in all modern factories of any size. The different operations in successive division of labor are carried on contemporaneously, but instead of each operation resulting in a completed commodity, it serves only as one further step in the production of a product or a service. In fact, the terms "contemporaneous" and "successive," as forms of classification, are almost synonymous with the "occupational" division of labor and the "division of processes," given above. When it is desired to emphasize the time element in production, the terms "contemporaneous" and "successive" are useful but in this text the classification first given will be the one generally followed.

There is still another kind of specialization within industry, which, though of a different character from that just considered, is of sufficient significance to warrant special treatment and has some claims for being regarded as a fourth type of the division of labor. In modern production, to an increasingly large extent, one group of persons provides the labor, another group the land, still another the capital, and finally the initiative and responsibility for the organization and direction of production is assumed by a fourth group.¹ To the extent that different groups of persons are furnishing these separate productive factors, they may be said to specialize. Professor Taylor, who has directed attention to this form of specialization, calls it *functional specialization*.² While this is a significant characteristic and should be considered, its importance lies in its bearing upon the problem of ownership in industry and in the part that ownership plays in laying the foundation for the distribution of income rather than the light it throws on the phases of production which are covered by the three types of specialization treated above.

Advantages of Specialization.—The advantages of a division of labor have long been recognized, but Adam Smith was the first to analyze them carefully and to put them into a form that was generally accepted by his contemporaries and, for that

¹ While the specialization here indicated is taking place, there is a counter-acting tendency observable. Through the widespread sale of corporate securities in this country to workmen, customers, and the general public, a diffusion of ownership of industry among all classes in the community is being effected. While this movement is perceptible and should be recognized the general statement concerning the ownership of the different factors is essentially true.

² TAYLOR, F. M., *op. cit.*, pp. 24-25.

matter, his statement of the case is accepted as essentially correct even to this day.¹

1. *Specialization results in greater skill and dexterity.* According to Smith, by reducing a man's work to a single operation, he not only increases his dexterity and improves his skill, but by so doing he improves the quality of the output. The hand, eye, and other parts of the body gain in coordination, thus enabling the laborer to produce with greater accuracy and speed and with less fatigue. Thus division of labor tends toward greater dexterity and skill, which in turn tend to improve the quality of the goods produced.

2. Smith says that a division of labor results in a saving of time. If one man undertakes to do several different kinds of work, there will be a loss of time in changing from one operation to another. In passing from job to job the kind of tools will vary and he will lose time in their manipulation, in addition to being less skilful in their use. As the character of the job becomes simpler through specialization, it will require less time to learn it, and by devoting his time exclusively to the simpler operation the workman will save time performing each process. Thus specialization results in a saving of time in both of these ways.

3. By subdividing the processes into their simplest operations, division of labor fosters invention and the substitution of machinery for human labor. As productive processes have approached routine and mechanical operations, it has become comparatively easy to devise a mechanical method of performing the work. In this way specialization stimulates invention.

These three advantages of the division of labor which were pointed out by Smith still obtain, and modern writers have added little to his treatment of this subject. But during the latter part of his life and throughout the nineteenth century, power machinery became increasingly important in the processes of production.

4. This extension of the use of machinery may be regarded as a fourth advantage of specialization. The most delicate and complicated processes, if reduced to routine operations, can be performed by machinery. Any one who has observed the manufacture of watches will recognize that very delicate operations can be performed by mechanical means. The following quota-

¹ SMITH, A., "Wealth of Nations" (1776), Book I, Chap. VII.

tion will indicate the minute character of the work actually done by means of machinery:

By the aid of special machines in these watch factories one man can make 1,200 fine screws per day, some of which are so small that more than 100,000 are required to weigh a pound. One of the finest pieces made is a "pallet-arbor" or pivotal bolt, which, for a small-sized watch, has a thread of 260 to the inch, weighs 1/130,000 of a pound, and undergoes 25 operations, and costs but 2¼ cents. Measurements are gauged to 1/25,000 of an inch.

The balance wheel, after being machined, weighs only 7 grains and when fitted with 16 gold screws weighs 7.2 grains; there are 80 separate operations upon a balance wheel, 66 of them being drilling, threading, and counter-sinking holes; the drills revolve at a speed of 4,800 turns a minute and one operator can drill upwards of 2,200 holes for the balance wheels per day.¹

The effect of such extreme specialization and the use of machinery is greatly to increase output. An investigation of watch manufacturers in Switzerland and in the New England states showed that the workmen in Switzerland, without the aid of machines, turned out 40 watches each per annum, while the output in New England, where machinery was used, averaged 150 per man per year.

Machinery will be used when there is a demand for the repetition of the same operations. The advantages of the machine over the hand-operated tool are that it makes possible the use of power greater than can be supplied by a human being and that it can be driven at a speed greater than a hand-operated tool. When driven by some natural force like steam, there is no work so heavy nor none so delicate that cannot be performed by machinery.² In addition, the use of machinery makes possible the exact duplication of parts, a difficult task with hand processes. This exact duplication of parts is an enormous economy in the production of goods in large quantities. While it is true that the introduction of machinery depends upon the division of labor, its use is a tremendous stimulus to further specialization. Before machines are available, some labor must be specialized in their production, which in turn requires further specialization of labor in the production of the tools required

¹ OUTERBRIDGE, Jr., A. E., "Specialization in Manufacture," *The Annals*, Vol. XXV, pp. 47-48, 1905. Quoted from MARSHALL, L. C., "Industrial Society," p. 280.

² CARVER, T. N., "Principles of National Economy," p. 170.

for the manufacture of these machines and in producing the raw materials from which they are made. In this way the use of machinery is both a result and a cause of specialization.

5. As a result of the division of labor, both labor and capital in the form of equipment can be more advantageously used. Work that requires unusual strength, physical or mental, can be assigned to those who possess these capacities. The possibility of making use of the special ability and aptitudes of men is particularly important in modern industry which calls for a high degree of initiative and ability to organize and direct large industrial enterprises. The use of qualities of leadership in an effective way results in large gains in the volume of wealth produced. In like manner, specialization makes possible an economical use of tools and machinery. When one man performs many operations, part of the capital equipment must be idle. But by specialization all of the various operations can be carried on simultaneously. The tools and machines can in this way be kept at work efficiently all the time. The gains from such a method are obvious.

6. Division of labor makes supervision of production easier. The simplification of the task of each laborer makes it comparatively easy to determine how much work is being turned out and to measure the workman's efficiency. It is obvious that the problem of supervision is thus greatly simplified through a division of labor.

7. Finally, specialization tends to reduce fatigue from labor. By subdividing the processes, each operation becomes easier to perform. By repetition of the same operation both the mental and physical effort becomes less. Everyone can recall from his own experience how much mental and physical effort was required to learn some new operation like playing the piano or operating a typewriter. When repeated sufficiently to become more or less mechanical, such operations can be performed with little or no perceptible attention or effort. In this way, division of labor tends to reduce fatigue. However, this advantage should not be overstated, as division of labor may contribute to fatigue in the mere monotony of the work, as will be shown in the treatment of the evils of specialization below.

Evils of Specialization.—Unfortunately, not all of the effects of specialization are beneficent. We can hardly overstate the importance of the division of labor in increasing the volume of

wealth produced, and thereby increasing the standard of living of mankind. But when specialization is carried to the extent that is found in many modern industrial establishments, some of its effects become detrimental.

1. There is the reduction in the importance of trade or craft skill, which usually accompanies a division of labor. If the man who has spent several years learning a trade finds a change in the processes of industry which depreciates the value of his services, he and all other workmen in his trade lose in their ability to secure an income from the sale of their services. As members of a craft they lose, even though the balance of the community may gain by getting a larger volume of wealth at lower unit costs. It thus appears that such a change affects members of the social group differently. In addition to this economic loss to the particular class or group affected, there are general considerations that have both economic and social consequences, such as the effect of a lower standard of living upon the mental outlook of those who have suffered by the change. While such loss is difficult of measurement, it is none the less real and may contribute to social unrest.

2. Extreme specialization takes away from the workman a large part of the interest which he has in a completed product. So long as the journeyman was a skilled mechanic, he had an intellectual interest in his work and a pride in the results of his efforts. This fact was a significant aid in maintaining the quality of the output, in dignifying work, and thus giving the laborer greater economic and social stability.

Frequently, specialization has resulted both in a loss of skill and in the social standing of the workmen—losses that have usually been obscured by the general benefits that have followed the division of labor. The extreme specialization found in modern factories has tended to take all of the intellectual interest out of work and has left only the labor. The routine and mechanical character of a highly specialized job has a deadening effect upon the worker. Specialization of this kind tends in the direction of monotony and monotony deadens the intellectual powers of the laborer. Thus, his initiative and adaptability are greatly reduced. Specialization may also take away the educational element in industry, the element that develops the best faculties possessed by man. "The medieval craftsman, like the farmer of today, was constantly making choices between alter-

natives and exercising his judgment and initiative."¹ An exercise of these faculties develops will and judgment. In a modern industrial plant there is little opportunity for the average workmen to exercise their initiative or to assume responsibility for organizing and planning. Much of this work which was formerly done by the journeyman has now been taken over by the specialized planning department. Here the work is laid out in great detail, so that the laborer has only to follow the specifications given him on his work ticket. While the volume of wealth has in most cases been greatly increased by these methods, yet the economic gains have been at the expense of the initiative of the workmen themselves. The evil arising from this cause is both economic and social. It tends to depreciate the productive powers of those coming under its influence and to cause discontent, unrest, and instability. This is a positive loss to society that must be charged against the extreme division of labor that is found in many industries.

3. The loss of skill has also reduced the number of workmen who, by training and general knowledge of the industry, are fitted for the position of foremanship. Every large factory has had difficulty with this problem and some of the larger establishments have undertaken to solve it by providing special training for their most promising workmen. Further evidence of the effect of extreme specialization on skill is found in the common complaint among factory managers of the difficulty of finding competent and efficient workmen. Reduction in the importance of skill makes for routine operation and tends to increase the amount of direction necessary to secure results. This condition can be reckoned as one of the results of specialization.

4. In addition to the direct economic losses just mentioned, there have been other effects which have had both social and economic consequences of significance. Loss of skill has caused greater instability in workmen and has contributed to social and industrial unrest. Reduction in the importance of skill has made it easier to pass from one job to another, thus fostering labor turnover. With no ties of interest in his work and assuming no responsibility for quantity or quality of the product, the unskilled workman is more likely to center his attention upon unimportant or assumed grievances and to shift from job to job. Restlessness in his economic relations tends to breed restlessness in his social relations and the net result is greater instability.

¹ CLAY, H., "Economics for the General Reader," p. 40.

5. Another evil of specialization, which is also social before it is economic, is the increased risk which workmen run as specialization increases. The man who has become highly specialized narrows his opportunities for employment and increases the likelihood of unemployment. By becoming specialized, he is less adaptable, hence his range of opportunities for work is restricted to his own special field. To secure a job he may be forced into another trade at a lower rate of compensation. By introducing the greater risk element, specialization contributes to unemployment which is both a social and an economic problem. Idle labor is a loss and anything which prevents the full utilization of the labor power of a community or a nation lessens to that extent the full volume of production.

Lastly, specialization is responsible for an evil more largely social than economic. Through specialization the individual workman is likely to be regarded as a means rather than an end of industry. When labor is highly subdivided and especially when there is no conscious attempt made to bridge the gap between the office and the shop, the workman is likely to be treated in an impersonal way. *Men tend to become numbers, in fact, rather than numbers as a convenient means of keeping records.* No more fundamental fact should be held in mind in studying production than that man is the end of all economic activity and that he should never be regarded primarily as a means. The preservation of the laborer's personality and individuality is of greater significance from a social point of view than the production of wealth, and while labor power, so far as we can see, will always be required as a factor in production, the necessity for maintaining the supremacy of the social over the economic interests, as they affect the welfare of the individual, must not be overlooked. The recognition of this fact has led many enlightened managers of industry to create within their organizations a special department to deal with the problem of industrial relations. The importance of this problem in the mind of modern managers can be seen by the attention that is given to it in the literature on industrial relations. While this is to a large extent a social issue, it nevertheless has important economic consequences. The volume of production that can be secured from a group of workmen depends very greatly upon the morale of the group. The problem itself is primarily the result of a division of labor and specialization and, in the absence of an effective method of dealing

with it, may become a very serious evil affecting both social and economic stability.

Thus, we must set over against the gains from specialization the evils which it creates, such as the loss to particular workmen that results from the reduction in the importance of trade or craft skill, the loss of intellectual interest in the work, the decline of initiative on the part of the workmen, the reduction of the all-round, competent workers, the tendency to increase labor turnover, the increased risk of idleness borne by the laborer, and the tendency to regard the laborer as a means rather than as the end of industry, with the attendant result of industrial unrest and social instability. These are social and economic problems brought about by specialization. The recognition of their existence does not argue that specialization is primarily an evil and should therefore be prevented as a matter of social policy. But to recognize that these problems are a product of specialization should aid in ameliorating the evil consequences that flow from them, and should help to develop a wise social policy that will aim to retain all of the benefits arising from the division of labor and to attempt to reduce its evil effects to a minimum.

Suggested References

- BLACK, JOHN, "Production Economics," Chaps. II, VI, and VII.
 BYE, RAYMOND T., "Principles of Economics," Chap. III.
 CARVER, THOMAS N., "Principles of National Economy," Chap. X.
 CLAY, HENRY, "Economics for the General Reader," Chaps. II and III.
 ELY, RICHARD T., "Outlines of Economics," 4th Ed., Chap. VIII.
 FAIRCHILD, FURNISS, and BUCK, "Elementary Economics," Vol. I, Chaps. II and V.
 GIDE, CHARLES, "Principles of Political Economy," (Rev.) Part II, Chap. III.
 HAYES, H. GORDON, "Our Economic System," Vol. I, Chaps. VI-X.
 JOHNSON, ALVIN S., "Introduction to Economics," pp. 19-34.
 MARSHALL, ALFRED, "Principles of Economics," Book IV, Chaps. I-II, and IX.
 MARSHALL, LEON C., "Industrial Society," pp. 373-403.
 MARSHALL, WRIGHT, and FIELD, "Materials for the Study of Elementary Economics," pp. 181-224.
 MILL, JOHN STUART, "Principles of Political Economy," Book I, Chap. III.
 SEAGER, HENRY R., "Principles of Economics," Chaps. VIII-X.
 SELIGMAN, EDWIN R. A., "Principles of Economics," Chaps. XVIII-XXII.
 TAUSSIG, FRANK W., "Principles of Economics," 3d Ed., Vol. I, Chaps. I-III.
 TAYLOR, FREDERICK M., "Principles of Economics," Chaps. IV, VI-VIII.
 VANDERBLUE, HOMER B., "Economic Principles," Book I, Chaps. I-IV.

CHAPTER V

LOCALIZATION OF INDUSTRY AND LARGE-SCALE PRODUCTION

Localization of Industry.—In addition to the forms of specialization treated above, consideration should be given to the tendency of industries to become concentrated in particular localities. This tendency is known as the *localization* of industry, and is, in fact, a phase of territorial division of labor. Because of its significance in production, this topic is worthy of special treatment. A very brief survey of the industries of the country will show the extent to which localization has taken place. It will be found that the textile mills are centered in New England; silk mills at Patterson, New Jersey; shoe factories at Brockton and Lynn, Massachusetts; steel mills in the Pittsburgh and South Chicago districts; the furniture factories in Grand Rapids, Michigan, and Chicago; automobiles in Detroit, and soon. So characteristic is this tendency to localize that, while applicable to all industry, it may be regarded as typical of modern manufacturing. There are a number of causes which account for this tendency, and, in any particular instance, there is likely to be a combination of causes responsible for the existing localization. But the major causes may be set forth separately, even though they do not operate independently.

1. *Nearness to Raw Materials.*—This influence is important in the location of many industries. The production of paper is usually in close proximity to the spruce and poplar forests; slaughtering and meat packing near the stock-raising areas; flour mills near the grain-growing regions; the pottery industry near the kind of clay required. Additional examples could be given but these are sufficient to call attention to the significance of this influence.

2. *Nearness to Markets.*—Since production is primarily for sale, nearness to the market is an important influence in the location of all industries, and especially is this true in the case of manufacturing. According to the census figures, the center

of manufactures in the United States has moved westward, following roughly the center of population.¹ Nearness to materials and to markets means more than mere distance. It implies general accessibility which may be obtained by efficient transportation at low costs. In the early history of industry in this country, water transportation was an important cause for localization. Many of the New England manufactories were first developed at points where cheap water transportation was easily accessible, and they have been maintained by the advantages of an early start and easy access to markets, even after cheaper transportation has removed the original causes of localization.

The importance of these two influences varies greatly among industries, depending upon the ease and cheapness with which the products can be transported.² If the transportation costs are relatively heavy, the industry will tend to be located at points where these charges will be at a minimum. A cheap and heavy material like clay will be transported only a short distance, while an equal weight of pottery made from the clay may, because of the increased value of the commodity, be shipped for long distances before the transportation charges become prohibitive. Hence, the pottery industry tends to be located near clay deposits. In the case of some other commodities nearness to the market is practically controlling. For instance, in the construction of houses or other buildings, the work usually has to be performed at the point where the owner wishes to use the building. The only exception to this general statement is the so-called "ready-made" houses, which, because of the economies in connection with the cutting and fitting of standardized units on a large scale, may be manufactured and shipped ready to erect at the point of use. While some construction of the completed building takes place, it does not bulk large in the total building industry.³ Other instances of the location of industries being controlled largely by the influence of the market could be cited. Fre-

¹ Twelfth Census (1900), Vol. VII, pp. CCX-CCXIV. An adaptation of the Census treatment is found in MARSHALL, WRIGHT, and FIELD, "Materials for the Study of Elementary Economics," pp. 189-197.

² MARSHALL, WRIGHT, and FIELD, *op. cit.*, p. 191.

³ In the erection of the modern skyscraper, both the structural steel and the stone used is fabricated at a distance in accordance with specifications, and erection becomes the putting together of fitted parts. In a sense this work is also an exception.

quently subsidiary industries, like the manufacture of fittings for shoes, or accessories for automobiles, locate near the site of the main industry. Enough has been said to show that the relative significance of these two influences on localization varies from industry to industry.

3. *Nearness to Cheap and Effective Power.*—A third cause of localization is nearness to cheap and effective power. In the past, water power was an important influence in determining the location of industries which tended to group themselves along waterways where such power could be developed. Dams were constructed and the force of gravity furnished the power requisite for the operation of the machinery. The existence of water power had much to do with the localization of the manufacture of hoisery and knit goods in the early history of these industries. Water power as a motive force had its greatest significance before the days of steam, but it is becoming of much more consequence in our day through its use in generating electricity. The development of electrical power at Niagara and elsewhere and its effect upon the concentration of industries in centers that can make use of such facilities are again calling attention to the significance of water as a motive force. The introduction of steam as a source of power revolutionized industry. However, with the development of steam a cheap fuel became an important consideration also, and hence the site of the coal fields had a significant bearing upon the placing of many of our basic industries. For instance, the steel mills at Pittsburgh were formerly near both ore and coal deposits. At Gary, Indiana, the steel mills are midway between the rich ore deposits of the Lake Superior region and the coal fields of Ohio, Indiana, and Illinois and, in addition, they are favorably situated for transportation. The influence of rich coal fields on localization of industries is evident, but, at some future time, if a source of power should supplant steam generated by coal, as, say, electricity generated at waterfalls, this change would have a tremendous influence on future localization of industries which would again tend to move toward the sources of power.

4. *A Supply of Competent Labor.*—From one point of view, the supply of labor is only a phase of power, for human labor is a part of the power required to operate industry. But the human factor in industry is so essential to its operation and so different in its bearing on the success of any line of pro-

duction, that it should be separately treated as a cause for localization. The supply of labor is only relatively mobile. The attachments of home and friends retard the movement of labor from one locality to another, so that it is easier to establish a new industry near large centers of population than to attract the labor to a new place. Often an industry may choose a location in order to gain access to a surplus of labor that is available in some particular community. If, as sometimes happens, the industries of a community use primarily men, another industry which can use women effectively can find available a surplus supply of labor. As soon as localization has begun in a particular community, it tends to develop greater specialization among the laborers. The technical skill required in many operations can be developed by the worker with a minimum of risk of unemployment from such specialization. In other words, the specialized worker has a better market for his specialized skill and the employer has less difficulty in finding the kind of labor that he needs. Manufacturing industries tend, therefore, to become established in a section where a supply of labor is easily accessible.

5. *Momentum of an Early Start.*—After an industry has once become localized, there are certain forces which tend to continue the specialization already begun. Among these is the momentum of an early start. The forces already described account for the possible area within which localization may take place. They do not explain why a particular place within that area was chosen rather than some other. The establishment of the first factory in a given locality is likely to be the result of chance. In the case of the making of shoes at Lynn, Massachusetts, it was due to the fact that a skilled shoemaker by the name of John Adams Dagyr settled in Lynn in 1750 and began to teach his apprentices the art of fine shoemaking.¹ It soon became known that the shoes made in Lynn were nearly as good as those made abroad. In this way localization of the shoe industry was started at Lynn. It is conceivable that had the Welsh shoemaker settled in some other town, the center of shoe manufacture would have been differently situated.

Similar instances of chance as a cause of localization of other industries could be cited. The decision of some one man may determine the history of the industry for a long period of time. In the beginning, the town in which the founder happens to live

¹ MARSHALL, WRIGHT, and FIELD, *op. cit.*, p. 194.

is chosen without much thought on his part, even though the question of proper location may subsequently become an important issue in the development of the industry. But even at the beginning, the wise founder of any industry will take into account the various forces which have been treated above. Some have been known to time-study the location of the plant, taking account of nearness to raw materials, markets, labor supply, etc., in the same way that time-study methods are used in laying out the equipment within the plant and in directing the operations of the employees. Such methods, it goes without saying, belong to the modern period of scientific management.

To be an influence on localization, the original establishment must succeed. Once the industry has been successfully started, this fact in itself is a tremendous impetus to further localization at this place. The reasons are evident. First, there is the tendency to imitate. The success of one man in an industry will stimulate the ambition of others and they are more likely to embark in the line that has already proven successful. Employees in the successful plant may start out for themselves, and the knowledge gained in its operation will be a powerful influence in the choice of their own private ventures.¹

Then, the capital needed to finance a new establishment can be found more readily if it is to be used in an industry known to the lender. Less familiar industries will create in his mind a fear of risk involved, and hence he will ask a higher rate of return to compensate for the uncertainty of success which he feels. In addition, development of specialized skill among the laborers soon follows the localization of industry. The contact of men working in the same line and thinking of the same problems will stimulate thought and result in improved methods of production, business organization, and administration. Finally, subsidiary industries also are likely to develop, thus furthering the tendency toward specialization. Firms will specialize on the manufacture of parts, as they specialize in the shoe centers, on counters, shanks, heel stiffeners, etc. Soon the manufacture of special machinery and tools will develop and the nearness to supplies of this character will cause the least possible delay in the operation of the machines. Success tends to breed success and specialization will expand by the mere fact of the original

¹ EPSTEIN, R. C., "Leadership in the Automobile Industry, 1903-1924," *Harvard Business Review*, April, 1927.

success, even to the extent of overcoming greater advantages of location possessed by other communities.

The momentum of an early start accounts for many of the industrial advantages that exist between countries, as, for instance, the textile mills of England and New England. The English mills have to import their raw cotton, which comes primarily from the United States, and they are often able to sell the finished goods in this country as cheaply as the New England manufacturers, notwithstanding the additional freight charges involved. For much the same reason, New England has been successful in maintaining her supremacy over the southern textile mills, although the nearness to raw materials is much more favorable to the south. The influence of a competent labor supply and the advantage of the market rest with New England, but the fact that New England mill towns had first become specialized accounts in large measure for the advantages they enjoy.

The advantages of localization once obtained may be maintained by artificial influences after other centers have shown their superiority in location. This may be accomplished by marketing practices, by governmental influence as the levy of protective duties, or by other similar means. The "Pittsburgh Plus" was a good example of the influence of a marketing practice on localization. The "Pittsburgh Plus" was a method of pricing steel products, which was in more or less continuous existence from 1901 to 1924 when the Federal Trade Commission ordered the practice abolished. It consisted of charging for all rolled steel products except rails, no matter where made or where sold, the Pittsburgh mill price plus the freight from Pittsburgh to destination. By this method the Pittsburgh manufacturer was placed upon an equal competitive basis with the producers in any other section of the country.¹ The effect of the method was to retain for the producers in the Pittsburgh district the advantages of location which were formerly enjoyed. There was a time when the Pittsburgh mills were favorably situated with respect to ore and fuel, but those advantages have probably been shifted to the South Chicago district, and, except for the existence of this system of marketing the products, the location of the steel industry

¹ This statement holds so long as and as far as the Pittsburgh Plus was operative. The Colorado Fuel and Iron Company and the manufacturers in the Birmingham district did not follow the system strictly.

would have shown a tendency to shift toward the more favorably situated centers of production. Other trade practices are doubtless in existence which have a similar influence upon the localization of industry.

Large-scale Production and the Limits of Specialization.—One of the results of specialization has been the development of large-scale production. By large-scale production is meant the tendency for production to be carried on in establishments of increasing size. The figures in Tables I, II, and III have been compiled from the census publications of the United States and indicate the growth of some of the larger manufacturing industries from 1850 to 1925.

TABLE I.—AGRICULTURAL IMPLEMENTS¹

Year	Number of establishments	Wage earners (average number)	Capital (in millions)	Product (in millions)
1850.....	1,333	7,220	\$ 3.6	\$ 6.8
1860.....	1,982	14,814	11.5	17.6
1870.....	2,076	25,249	34.8	52.1
1880.....	1,943	39,580	62.1	68.6
1890.....	910	38,827	145.3	81.3
1900.....	715	46,582	157.7	101.2
1905.....	648	47,394	196.7	112.0
1910.....	640	50,551	256.3	146.3
1915.....	601	48,459	338.5	164.1
1920.....	521	54,368	336.9	304.9
1923.....	312	30,962	²	151.2
1925.....	303	48,459	²	169.4

¹ The figures from 1850-1915 have been taken from TAUSIG, F. W., "Principles," Vol. I, p. 409; those for 1920 were compiled from the 1920 census; those for 1923 and 1925, are from the "Biennial Census of Manufacturers," published by the U. S. Dept. Com. (1925), p. 984.

² Not called for on the schedule.

These figures will serve to illustrate this movement toward large-scale production. In all three instances, it will be observed that the total capital, the total value of product, and the total number of persons employed have increased rapidly, while the number of establishments has tended to decline or to increase at a much slower rate than the growth in size of the plants. These figures do not tell the whole story, for the census regards a factory in any one place as an independ-

TABLE II.—IRON AND STEEL¹
(Blast furnaces, steel works, and rolling mills)

Year	Number of establishments	Wage earners (average number)	Capital (in millions)	Product (in millions)
1850.....	468	24,874	\$ 21.9	\$ 20.4
1860.....	542	35,189	44.6	52.8
1870.....	808	77,555	121.8	207.2
1880.....	792	140,798	209.9	296.6
1890.....	719	171,181	405.8	478.7
1900.....	668	222,490	573.4	804.0
1905.....	605	242,640	936.3	905.8
1910.....	654	278,505	1,492.3	1,377.2
1915.....	587	278,072	1,720.7	1,263.3
1920.....	695	416,748	3,458.9	3,623.3
1923.....	658	424,913	"	4,161.9
1925.....	595	399,914	"	3,711.0

TABLE III.—COTTON GOODS¹

Year	Number of establishments	Wage earners (average number)	Capital (in millions)	Product (in millions)
1850.....	1,094	92,286	\$ 74.5	\$ 61.9
1860.....	1,091	122,028	98.6	115.7
1870.....	956	135,369	140.7	177.6
1880.....	1,005	185,472	219.5	210.9
1890.....	905	218,876	354.0	268.0
1900.....	1,055	302,861	467.2	339.2
1905.....	1,154	315,814	613.1	450.5
1910.....	1,324	378,880	822.2	628.4
1915.....	1,328	393,404	899.8	701.3
1920.....	1,496	446,852	1,914.9	2,195.5
1923.....	1,642	495,197	"	2,010.1
1925.....	1,638	468,352	"	1,819.8

¹ The figures from 1850-1915 have been taken from TAUSIGN, F. W., "Principles," Vol. I, p. 49; and those for 1920 were compiled from the 1920 census; those for 1923 and 1925 are from the "Biennial Census of Manufacturers," published by the U. S. Dept. Com. (1925), iron and steel, pp. 429, 441; cotton goods, p. 230.

" Not called for on the schedule.

ent and separate establishment. In all three industries, and especially in iron manufacture and in that of agricultural implements, combination and large-scale operation have been going on in a form that the census figures do not reveal. Hence, the tendency toward concentration is more marked than the figures indicate. This drift toward large-scale production shows itself in all advanced countries and may be said to characterize modern industry. To understand this tendency toward enlarging the size of the producing plant, it will be necessary to examine the causes which make large-scale production possible, and the advantages and economies that are thereby obtained.

Causes of Large-Scale Production.—A number of influences have been at work simultaneously to cause a growth in the size of the business establishment.

1. There has been the expansion of the area of the market. Before large-scale production can be carried on economically, there must be a large demand for the commodities in question. This demand may be increased either by the growth of population as in cities, or by the development of cheap transportation which enables the producer to reach a larger number of buyers, even though they be widely separated. Both of these influences have been at work to increase the area of the market during the last 150 years. Cheap transportation is the more important of the two, however, for it not only enlarges the market extensively, but it enables a large territory to be drawn upon for supplies and thus acts as a very potent force in increasing the density of population. If it were not for cheap transportation, large aggregations of people could not be fed. For this reason, cheap and efficient transportation, by increasing the area of the market, is one of the most fundamental causes of large-scale production.

2. As the demand for a commodity expands, it becomes increasingly advantageous to apply the principle of specialization both in labor and in capital equipment. In the large plant, full advantage can be taken of the gains from the division of labor. The highly specialized laborer can be economically used and can normally be kept at work at good wages. The work in the office can be subdivided and specialists used where needed; buyers thoroughly familiar with the character of materials can be employed; a selling force, skilled in the art of selling, can be maintained, and other forms of specialization that result in

handsome gains can be utilized by the large concern but which the small plant can not afford to undertake.

3. The large plant can also make more effective use of mechanical methods of operation. The same principle of specialization holds in respect to the use of machinery as in the case of human labor. The larger the establishment, the greater the opportunity for applying elaborate tools and machines, which in themselves are expensive and often confined to very simple operations. For example, an illustration may be taken from the manufacture of agricultural implements. The International Harvester Company has a machine for the shaping of poles used on wagons and other farm implements. The cost of this machine is \$2,500 and it saves 1 cent per pole.¹ Since poles are used by the thousands each year, this small saving justifies the investment. Then, the large plant can install powerful machinery capable of performing operations which could not be undertaken by manual labor. The work done by steam or electric cranes in manufacturing plants, or by machinery used in the erection of bridges or the modern skyscrapers, could not be carried on by hand apparatus and manual labor; but with the aid of power machinery these operations can be performed easily and economically. Thus, large-scale production makes possible the use of both more elaborate and more powerful machinery than could be used in smaller establishments.

4. Another cause is the advantage which the large plant enjoys in its mercantile operations. In the buying of the materials used and in the sale of the products, the large plant has an advantage over the small one. The gain from large-scale buying is twofold. The buyer can become a specialist and know the quality of the materials and market conditions as well as the producer. Then the large buyer usually secures a better price than when smaller quantities are taken. Competition among sellers of the materials for the trade of the large plant usually results in a lower price. In fact, the expenses involved in making a large sale are little or no greater than those of a small one, hence the sellers can afford to give the customer who buys in quantity a more favorable price. Likewise, there is frequently an advantage in the disposal of the product. A large plant can by means of advertising find its own market and economize in the sale of its goods. There has been a tendency in a great

¹ This illustration is taken from TAUSSIG, F. W., "Principles," Vol. I, p. 52.

many lines of manufacture to integrate the selling of goods with their production, whereas in the early history of most industries the manufacturer was concerned primarily with the problems of production. At this stage he relied upon middlemen, jobbers, and brokers to dispose of his product. But as the production problems are solved, the manufacturer begins to look for his own market and endeavors to become independent of middlemen. He may take over entirely the selling function by establishing a selling organization, even to the extent of running retail stores, as in the case of the original American Tobacco Company which established the United Cigar Stores Company to market its goods; or, he may merely determine the selling policy for the disposal of his products as in the case of a firm like Hart, Schaffner and Marx, or the Ford Motor Company. Integration of this character is feasible only where the scale of production is large.

The canning industry is a good example to illustrate how these forces operate. The canning factory must ordinarily be located near the supply of materials to be preserved and often these factories have been started locally. At this stage, the finding of a market for the product is the work of a wholesale grocer or other general distributor. But a Van Camp or a Heinz company may find it profitable to create a market through the advertising of its own brands, and to integrate the production by the purchase of the local canneries. Such companies rely upon the regular channels of merchandising, as the wholesaler and the retailer, for the disposal of their products. In this industry, there now exists in active competition both the small independent cannery, dependent upon middlemen for marketing its goods, and the large firm which has created its own market. Whatever method is used to secure trade, such as traveling salesmen, trade catalogues, advertising, and the like, it is ordinarily less expensive per unit when the cost is spread over a large volume of sales. Hence, the large producer enjoys an advantage in both buying his materials and in disposing of his output.

5. Another important stimulus to large-scale production is the utilization of by-products. In almost every industry much of the material is not fitted for the main product and in the small plant this material is wasted. The large plant can afford to install equipment and make use of this waste material, which otherwise would be a loss. The packing industry is a stock

illustration of the utilization of by-products. From the slaughtering of animals primarily for meat products, come leather, fertilizer, athletic goods, pharmaceutical supplies, etc. Slaughtering for local use would result in a loss of much of this material. Another instance is the woolen industry. In a very large woolen factory it becomes advantageous to install equipment to make use of the fatty matter that is attached to the wool as it is clipped from the sheep.¹ The grease has to be scoured out of the wool before it is made into cloth, and in a small mill it is waste material, but in the large plant it becomes a source of income. Many similar examples could be mentioned from most of the important industries of our day, but enough has been said to indicate how the large plant may add to its profits by utilizing what would otherwise be wasted.

6. A large enterprise can afford to experiment with new devices and with new methods in a way that would be ruinous for the small establishment. The successes of such experiments may be expected more than to offset the failures, and in this way the large plant gains over the small one. It can employ the best possible technical skill and spread the expense over a greater volume of output. The results of such practice are that many important scientific discoveries have been made and new devices invented. Such experimentation is not confined to technical operations, but may consist of more effective systems for determining unit costs or the application of other methods of improving the organization and operation of the business. The basis for effective competition is knowledge of unit costs, and in any complicated business the finding of these costs requires special consideration. The small plant cannot afford to maintain an expensive accounting department, hence the manager of such a firm must rely upon approximations of his costs.

In like manner, a large plant can experiment with its industrial relations in a way that would scarcely be feasible for a small one. The plan of negotiation now in use in the men's clothing industry in Chicago consists of a board, composed of representatives of the laborers and of the firms together with an impartial

¹ TAUSSIG, F. W., "Principles," Vol. I, p. 54. A lucrative business has been established in the manufacture of small implement handles from the waste material resulting from the manufacture of farm implements at Moline, Illinois. The founder of this business gained his idea from the study of the utilization of by-products in a

chairman, which undertakes to settle disputes over wages and working conditions in this industry. This method was first worked out and in operation for approximately ten years in the firm of Hart, Shaffner and Marx before it was extended to the balance of the manufacturers in that market. It would have been very much more difficult to develop such a system if all the conflicting interests of the market had been present during the early stages. In like manner, firms such as the Colorado Fuel and Iron Company, the Standard Oil Companies, and the International Harvester Company, can afford to experiment with special plans for dealing with labor that could not be undertaken by small concerns. The expense of maintaining a labor department adds to the overhead costs as does also the arbitration machinery. Not until such systems have been proven effective can the small firm undertake them, and even then, it requires one of considerable size to be able to deal with its employees in this way.

Another advantage enjoyed by the large plant is that it can experiment with plans for stabilizing its operations. One of the big problems of the production manager is concerned with the forecasting of the demand for his product. Efficiency in production requires that the equipment be kept continuously in operation but, since most commodities are produced in anticipation of demand, it is extremely difficult for any manager to determine in advance how many goods he can sell. The inability to make these estimates accurately is a contributing cause of what is commonly known as the *business cycle*, that is, recurring periods of productive activity and of depression. The large plant can afford to study this problem and to experiment with methods for estimating demand more accurately and for adding supplementary lines of production as a means of keeping both machinery and labor busy. The opportunity to study this whole problem and the feasibility of adjusting the policy of the concern accordingly are much greater in the large than in the small plant. In fact, the large plant enjoys an advantage in all lines of experimentation, as the costs involved can be spread over a larger volume of product and the successes of such experiments can serve as insurance against the risks of failure.

From the discussion above, it appears that large-scale production results in many economies. An examination of these will show that they can be classified into internal and external economies, or, to put the idea in another way, into economies

which result from mere growth in size of the producing unit, and those which result from conditions outside the establishment itself. Some of the internal economies are as follows: The large establishment can make more effective use of the principle of division of labor; it can make maximum use of by-products; it can utilize machinery to better advantage; its buying and selling can be at lower prices because it is dealing in large quantities. For similar reasons, the large plant can usually save on freight rates by buying in carload lots; it can afford to experiment with new methods and devices; and it usually enjoys an advantage in its credit and financial relations. On this latter account, it can borrow at lower rates and take advantage of all discounts allowed on cash purchases of materials used. Thus, the large plant enjoys many economies due primarily to the scale on which production is carried.

In addition to the internal economies, there are many outside conditions of an industry which result in economies in the general conduct of a business. Any improvements which lower transportation charges, as larger locomotives or freight cars, will reduce the costs of operation. The same effect will follow from all technical improvements in production methods. Then, as the volume of any commodity increases and more machinery is required in its production, the gains from large-scale production of this machinery will likely be reflected in lower costs of operation in all industries that make use of it. Other similar examples might be cited where the reduction in the expenses of operating an industry is due to economies that take place outside of the industry itself. Enough has been said, however, in support of the general statement that the economies of large-scale production may result from either internal or external influences.

Limits of Large-scale Production.—The advantages of large-scale production are so apparent and so great that one might expect that the small firm would be driven out of industry. But such is not the case. The small firm has persisted in the great majority of industries, as an examination of the following facts will reveal. The census of 1920 reported 290,105 manufacturing establishments. Of this number 235,884 employed less than 20 persons per establishment, and of these 37,934 employed no wage earners at all. Eliminating those in which there were no wage earners employed, the remaining 197,950 establishments

employed 1,772,167—an average of about 9 persons per plant. The small establishments persist notwithstanding the increase of the size of manufacturing plants, whether size is measured by the average number of wage earners employed or by the value of products produced. Since competition between the large and the small firm is not for a fixed volume of trade, the large firm may increase its growth without a direct encroachment upon that of the small firm. The fact that the small firm can, in a great majority of industries, maintain itself suggests that there are limits upon the size of the efficient plant, even though these limits are variable as between different industries.

The first limit on the size of the producing plant is the area of the market. The relation between the volume of goods that can be sold and the size of the plant that can be efficiently operated is of fundamental importance. But modern means of transportation have so cheapened the carrying charges that this influence, while always fundamental, does not operate to any considerable extent to limit the scale of production. In most industries, the area that can be served permits the development of a plant which can attain maximum efficiency from a technical point of view. Freight charges, however, do have a positive influence on the localization of industry as they tend to limit the distance which goods can be profitably shipped.

One of the most positive forces that limits the size of producing plants is that found in human nature itself. Most men do not work as energetically and with as much enthusiasm for an employer as they do for themselves. In large-scale operation it is necessary to rely upon hired labor, hence there is less appeal to self-interest than when those who carry on the work have a direct, financial interest in the success of the undertaking. The larger the scale of operation, the weaker this appeal becomes. The organization of a business becomes more complex with growth in size. Orders passing from the chief executives through minor executives lose force, so that by the time they reach the man who performs the work much of the original impelling force has been lost. Scarcity of executive ability and leadership capable of inspiring the whole working force with enthusiasm is a very positive influence in limiting the size of business enterprises. A business may become too big to be managed effectively.

A phase of this influence seems to develop when a business gets so large that there is a division of functions within the

structure of the organization and the financial management becomes separated from the technical operation of the factory. As business is now commonly organized, those who determine the policy of the concern devote most of their time and energy to the financial and credit problems, while those trained in technical operations are in charge of the factory and the manufacturing processes. It not infrequently happens that the decisions as to policy are not well founded from a manufacturing and technical point of view. Whenever such conditions are found, they set limits on the size of the producing plant, because a mistaken production policy will eventually result in a loss of business because of high unit costs. This limit on large-scale operations is not necessarily inherent in large-scale production but it is more likely to be found in extremely large organizations than in smaller ones.

In the small plant, the owner is in closer personal touch with his employees and is in a better position to instill in them some of his enthusiasm and energy. It can be laid down as a general rule that the greater the distance between those who make the decisions concerning business policies and those who carry them out, the less will be the efficiency in operation. The loss in productive force due to a lack of direct and positive interest in a business is extremely difficult to measure, but a little study of the problem will convince one that the loss is large and that this is an important limit on the size of the producing unit. Some progress has been made in overcoming the inertia of large-scale production by systems of organization in which responsibility is more definitely centralized upon departmental heads. Plans for profit sharing, for sale of stock on favorable terms to employees, and the attention given to motivating the working force are all evidences of the problem and indicate the methods which management has devised for endeavoring to deal with it.

The nature of the industry also is a limiting cause on the size of the producing unit. Agriculture, as contrasted with manufacturing, is conducted on a small scale. Personal supervision and attention are of so much greater significance in agriculture that farming has not shown a very great tendency toward large-scale production.¹ The difficulty of superintendence which

¹ There are doubtless other influences that have prevented large-scale operations in agriculture, but the ones mentioned are significant.

limits the size of plants in all industries operates much more quickly in farming than in many other lines. Canning factories also are relatively small, due to the difficulty of assembling the raw materials to be preserved. Most of the fruits and vegetables that are preserved are perishable and deteriorate very rapidly. Hence, the factories are usually situated so that farmers can make daily deliveries. This situation prevents the canning factory from becoming a very large establishment. Similar influences may be found in other industries.

In the merchandising field there has been a question as to the advantages of large-scale production. The evidence for testing the effectiveness of the large store over the small one has until recently been unsatisfactory, and even yet it is inadequate to establish the case perfectly. It has been assumed by most writers that the principle of large-scale production held in the mercantile field, and particularly in wholesale operations.¹ But recent studies of the operating costs of merchandising raise some questions as to the advantages of the large store over the small one. The following figures throw light upon the subject.²

COST OF DOING BUSINESS AS A PERCENTAGE OF SALES

Volume of Sales	Department Stores, 1921 ¹	Percentage
Less than	\$ 250,000	26.6
\$ 250,000	to 500,000	27.6
500,000	to 1,000,000	28.3
1,000,000	to 2,500,000	28.7
More than	2,500,000	28.5

¹ Harvard figures, quoted by L. D. H. WELD, *American Economic Review*, Vol. XIII, No. 1, Supplement, p. 187.

² TAUSSIG, F. W., "Principles," Vol. I, p. 53.

³ Nystrom, P. H., writing in 1915 ("Economics of Retailing" p. 334) says: "Retailing is different from most other businesses in this respect, that the small store can usually be conducted at a lower cost than a larger store. Unless other advantages, such as in buying or in advertising, may be gained to offset the increasing expense of the large establishment, there is nothing to be gained by increasing the business of a store beyond a certain point that keeps a small institution busy."

CLARK, F. E., "Principles of Marketing," Chaps. XI, XII, and XXV, 1922, discussing this question (p. 535) says: "Again it is not at all evident that large-scale retailing is more efficient in so far as the costs of operation are concerned. In fact, the costs of selling in department stores and large stores generally, with the possible exception of chain stores, are thought to be greater than the costs of small competing stores when efficiently operated."

Retail Jewelry Stores

Volume of Sales		1920 ²	Percentage
Less than		\$ 25,000	34.1
\$ 25,000	to	50,000	32.6
50,000	to	100,000	30.1
More than		100,000	32.8

² *Ibid.*, p. 187.

1921

Less than		\$ 20,000	47.1
\$ 20,000	to	50,000	41.6
More than		50,000	40.6

Retail Shoe Business

Volume of Sales		1920 ³	Percentage
Less than		\$ 30,000	25.0
\$ 30,000	to	50,000	24.3
50,000	to	100,000	23.5
100,000	to	250,000	26.6
More than		250,000	29.6

³ *Ibid.*, p. 187.

Retail Grocery Stores

Volume of Sales		1919 ⁴	Percentage
Less than		\$ 20,000	14.6
\$ 20,000	to	50,000	10.6
50,000	to	100,000	9.8
More than		100,000	13.1

⁴ *Ibid.*, p. 188, quoted from MACLEIN and McNALL, "What the Retailer Does with the Consumer's Dollar."

Retail Meat Stores

Volume of Sales		1919 ⁵	Percentage
Less than		\$ 25,000	18.11
\$ 25,000	to	50,000	17.41
50,000	to	100,000	15.33
100,000	to	200,000	15.55
More than		200,000	14.47

⁵ *Ibid.*, p. 188, quoted from *Preliminary Report* of U. S. Department of Agriculture.

Retail Clothing Stores

Volume of Sales		1919 ⁶	Percentage
Less than		\$ 40,000	19.00
\$ 40,000	to	80,000	18.12
80,000	to	180,000	20.52
More than		180,000	24.03

⁶ *Ibid.* quoted from SECHREST, "Costs, Merchandising Practices, Advertising and Sales, in the Retail Distribution of Clothing," Vol. III, p. 247.

According to Weld, the most striking features of these figures are:

1. Department stores show practically a continuous increase in expense as the size of the store increases.
2. All classes of stores except department stores show an unmistakable decrease in expense as the size increases from very small stores to medium-sized stores.
3. All classes of stores except meat shops show a definite increase in expense for very large stores, as compared with medium-sized stores. The turning point from decreasing expense to increasing expense appears to be when sales exceed \$100,000. For clothing stores the turning point is reached in a smaller size group.
4. In only one trade, the retail meat trade, do expenses decrease without interruption as the size increases.¹

After further detailed consideration of these figures and similar ones for the wholesale trade in hardware, clothing, electric goods, groceries and shoes, Weld comes to the following conclusions:

1. The principle of decreasing costs does apply to merchandising, but not to the same extent as to most kinds of manufacturing.
2. In the retail and wholesale trades, points are soon reached both in the intensive cultivation of the territory already covered and in the extensive cultivation of additional territory, where costs begin to rise.
3. The advantages of large-scale merchandising are greater than the figures on operating costs seem to indicate, because these figures do not measure the better purchasing power of large companies, nor the greater profit on capital that may result from more rapid stock-turn.
4. The higher operating costs of large stores are due in great part to more elaborate services performed; but this more elaborate service in many cases may be a necessary accompaniment of size.
5. The available data seem to indicate that the principle of decreasing costs does not apply as much to wholesaling as to retailing.
6. The principle of decreasing costs applies to different marketing agencies in varying degrees, depending on the number of commodities handled; whether they are handled in bulk; whether volume can be increased substantially without increasing plant or personnel; whether salesmen have to be used; and whether salesmen can sell on the premises (as in a retail store) or whether they have to travel in search of buyers (as in a wholesale house).²

These figures are the best available on this important field of business, and the conclusions can be accepted at least tentatively

¹ *Ibid.*, p. 188.

² *Ibid.*, p. 197.

as showing the operation of the principle of large-scale production in the field of merchandising.

In contrasting merchandising with manufacturing, or the mercantile operations of a manufacturing plant with the mechanical, it is more likely that the mechanical methods in different plants approach uniformity than do the commercial practices. The mechanical limits of specialization are reached when no further standardization of parts can be made without sacrificing technical efficiency. Through the various educational agencies, such as the technical schools, the technical press, etc., new methods of production soon become common knowledge. There is not the same cooperation in spreading knowledge concerning commercial methods. Hence, it would seem that the extent to which specialization can be profitably carried depends upon commercial conditions and commercial practice. When rival producers are striving to occupy the same field, success will fall to the one who has developed the more effective sales methods, to the one who exercises the best judgment as to when it will pay to force the market by advertising or by other means.

Combination and Large-scale Production.—A phase of large-scale production which has developed within the last two generations may be called large-scale management or combination. A plant may increase in size until it has attained maximum efficiency from a technical point of view and still show further economies by a combination with another plant of equal size and efficiency. This type of combination appears in two forms, which have been described as "horizontal" and "vertical" combinations.¹ If, for instance, a number of competing plants in any industry should combine into one large concern under single management, such combination would be called "horizontal," for it reaches out and unites concerns operating on the same plane and competing for the same trade. This is the kind of combination which was typical of the trust movement. It may result in additional economies of production as machinery can be standardized in the different works, and repairs and replacement facilitated. Savings may be effected also in the selling methods as orders can be filled from the most conveniently located plant, thus saving transportation charges. Other similar economies may be effected, but a still more powerful

¹ TAUSIG, F. W., *op. cit.*, Vol. I, p. 59.

motive for combination of this character is the desire to avoid the effect of competition. Many important methods have been devised by which unity of action can be accomplished and competition avoided. The most important among these are the *pool*, the *trust*, the *holding company*, and *interlocking directorates*, i.e., when one company elects members of the board of directors of another company to its own directorate.

Vertical combination is frequently called integration of industry. Whenever there is a tendency to bring under one management all of the processes of production, from the raw materials, on the one side, to the merchandising of the product, on the other, we say that the industrial processes have been integrated, or there has been a vertical or sequential combination. The United States Steel Corporation has experimented with vertical combination on a scale quite unknown in any other industry. Ore, coal, and stone owned by the company are transported over barges and railways that are similarly owned, and are then smelted in its own blast furnaces and converted into steel products of various kinds in its own mills. This principle of integration may begin at any point in the general process of production and work in the direction of an integrated industry. In some cases, it will begin with the merchandising and work backward toward production as when a firm, such as Marshall Field and Company,¹ takes over cotton or lace mills. On the other hand, the process may work from the manufacturing toward the marketing of the product as when a shoe manufacturer establishes his own retail stores, as has been done by W. L. Douglas & Co.

In all cases of integration there is an attempt to gain control of the profits which have previously accrued to the concerns producing parts of the completed commodity. Competition is also an impelling force which drives management to find more economical methods of operation, and some economies have been found in organization and large-scale administration. This fact accounts for the large amount of business acumen that is being directed toward experimentation and invention of more effective administrative methods. The possibilities of large-scale management seem to be limited therefore only by the capacity of men for organization and administration.

¹ The largest merchandising firm in Chicago.

Suggested References

- BLACK, JOHN D., "Production Economics," Chap. IX.
CLARK, FRED E., "Principles of Marketing," Chaps. XII-XXV.
FAIRCHILD, FURNISS, and BUCK, "Elementary Economics," Vol. I, Chap. VII.
HAYES, H. GORDON, "Our Economic System," Vol. I, Chap. XIV.
HOBSON, JOHN A., "Work and Wealth," Chaps. V and VI.
JOHNSON, ALVIN S., "Introduction to Economics," pp. 155-170.
MARSHALL, ALFRED, "Principles of Economics," Book IV, Chaps. IX, X, and XI.
MARSHALL, LEON C., "Industrial Society," pp. 633-668.
MARSHALL, WRIGHT, and FIELD, "Materials for the Study of Elementary Economics," pp. 189-197.
MILL, JOHN STUART, "Principles of Political Economy," Book I, Chap. IX.
TAUSSIG, FRANK W., "Principles of Economics," 3d Ed., Vol. I, Chap. IV.
VANDERBLUE, HOMER B., "Economic Principles," Book I, Chap. II.
VERLEN, THORSTEIN, "Theory of Business Enterprise," Chaps. II and III.
WELD, L. H. D., "Large-scale Production and Merchandising," *Am. Econ. Rev.*, Vol. XIII, No. 1, pp. 185-189.
Cf. also the references given at the end of the previous chapter.

CHAPTER VI

THE ORGANIZATION OF PRODUCTION

It has already been shown that nature does not provide wealth in quantities sufficient to satisfy human desires for it, and therefore, it must be produced. We have also seen that production requires the combination of certain factors known as land, labor, and capital; that plant, equipment, labor, and raw materials must be brought together and combined in such proportions as will result in a usable good. It now becomes our problem to see how this task is accomplished.

The Entrepreneur.—The responsibility for determining what shall be produced and how it shall be made rests in first instance with the *entrepreneur*, i.e., with the man who perceives a human need and undertakes the task of producing the goods that will satisfy it.¹ This man is frequently referred to as the "captain of industry" because he secures control over the factors of production, directs the operations, and assumes the risks involved in bringing the finished goods to the market. In other words, he undertakes the business of producing the wealth that men desire. In performing his tasks, he exercises authority in modern industry. While he may delegate the actual operations of the business to technical experts of one kind or another, these men are not free to proceed without authorization from him and are accountable to him. If not actually engaged in directing its operations, he retains at least the right to determine the broad, general policy of the business. It is evident, therefore, that the

¹ This is a French word that is approximately equivalent to the English word "undertaker." For obvious reasons, the English word cannot be used to describe the specialized function involved in the operation of modern industry, hence the economists have introduced this French word to convey the concept which they have in mind. Some writers use the word "enterpriser," but since the idea to be conveyed is the fact that some person undertakes the responsibility involved in the operation of an industry, the term "entrepreneur" is more commonly used in modern texts. It comes nearer to expressing the actual part played by the organizer of industry than does the term "enterpriser."

character and volume of wealth available for the gratification of human desires depend upon the manner in which entrepreneurs perform their functions. They occupy the key position to all production. Naturally, then, consumers generally are vitally concerned with the way in which entrepreneurs are selected and perform their tasks.

Selection of the Entrepreneur.—Since entrepreneurs occupy such a significant position in our modern economic life, it will be worth while to examine into the methods of selecting them. In some instances the entrepreneur is selected by the state, as in the case of the German railways and of many of the English and a few American public utilities. While men selected to direct municipal or state-owned industries exercise much of the authority associated with the entrepreneur, nevertheless, in one sense, it is correct to say that the source of this authority is the people who constitute the government and who cause the appointments to be made. The appointees are performing only the functions that have been delegated to them and are, therefore, not independent of the wishes of their constituents. But since the state is unable to operate except through officials who are either elected or appointed, it is perhaps correct for the purposes at hand to say that these officials perform the functions of the entrepreneur. Their decisions control the operations of the industries committed to them during the term of their appointment.

In the great Cooperative Societies of England, or wherever cooperation is conducted along similar lines, we find another method of selecting the entrepreneur that is somewhat akin to the one just described, but differing from it by virtue of the differences between a state or municipality and a cooperative society. The real source of authority in a cooperative society is the membership composing the cooperative group. But it is customary in all such organizations to elect representatives or officials who are charged with the responsibility of directing and carrying forward the work of the association. The precise authority which these officials exercise depends, of course, upon the rules and regulations of the particular organization. However, in so far as these officials or representatives determine the activities of the collective group, they may be said to be performing the functions of the entrepreneur. In both of these instances, it will be noted that the method of selection involves collective action

However significant selection by the state, or by a cooperative group, may become in the future, at the present time these methods are insignificant when contrasted with the self-appointment of entrepreneurs. We have already seen that one of the characteristics of our industrial order is the fact of individual initiative, the individual being free within very broad limits to choose his occupation and to enter into business for himself. If an individual thinks he can produce a good that people want and by so doing make a profit for himself, he is free to set up a business on his own initiative. This freedom of individual initiative is what is here referred to as the self-appointment of the entrepreneur. Quite apart from the relative merits of any other system, this is the chief form of selection that prevails today. As consumers, we are dependent for a continuous flow of wealth upon men who have selected themselves to perform the functions of industrial leadership. The chief rôle in determining what use shall be made of the natural resources of the country is played by entrepreneurs selected in this way.

The Functions of the Entrepreneur.—In a general way, the functions of the entrepreneur have been set forth in the preceding discussion but, owing to the importance of these functions, a more detailed examination will be made. The part played by the entrepreneur in modern industry can best be seen by an examination of some simple industry like a retail store. In such an undertaking, he must hire or erect a store building in which to conduct his operations. In either case the services of land will be required and in addition he must secure capital equipment in the form of fixtures, delivery trucks, and any further aids necessary to the conduct of the business. Then he must lay in his stock of goods and hire or furnish the labor to carry on the work of the store. All of these things must be done before he can begin his duties as a storekeeper. The point to be noted is that the decision to open a store, the choice of location, the determination of the kind of goods to be carried, the selection of help, are all questions that must be decided in advance of performing the service of a storekeeper. These are the decisions that the entrepreneur has to make in the early and simpler stages of an industry. He assumes the authority and exercises all the private control found in industry at this stage of development. But as the industry grows in size and complexity, specialization takes place and a differentiation of functions develops along two

fairly definite lines, both of which tend to narrow the strictly entrepreneurial functions. The two lines of differentiation here implied are the development of credit and lending agencies, and that of management. Since the functions of the entrepreneur in modern industry are vitally affected by these developments, it will be necessary to give them special consideration.

The Influence of Lenders.—In the illustration of the storekeeper above, it was assumed that, as entrepreneur, he possessed the funds or money which enabled him to render his decisions freely. But in actual business life, the great majority of business men do not have sufficient money to set up their business independently, so they resort to borrowing as a means of finding an adequate supply of capital. The prospective storekeeper may borrow from a friend or relative, giving his personal note or a mortgage on some property which he owns as security, or he may borrow from a bank or an investment house on such terms as he can secure, or the stock of goods may be advanced to him by a wholesaler or other distributing agency. He will supplement his own funds by any one or by a combination of these methods. Whenever he applies to any of these agencies for assistance, he must satisfy them of his ability to repay the principal with interest. In safeguarding their advances, the lenders may lay down stipulations which materially modify the activities of the entrepreneur and place a positive check upon his freedom of action. Banks and other lending agencies are coming to exercise more and more influence over business men and are requiring them to render systematic statements concerning the financial operating conditions of their business. So far as lending agencies, through their regulation and control of credit, exercise an influence over business policies, they are performing one of the functions that is usually associated with the entrepreneur himself, namely, the control of capital funds.

The Place of Management.—The second differentiation which has modified the activity of the entrepreneur is the development of what is now commonly known as "management." By management is meant the exercise of the responsibility involved in the actual operation of the business. Formerly, management was an incident to ownership and, in all small-scale operations, it is still associated with the ownership of the property. For instance, the farmer is both owner and manager of his farm. Even when he leases land, he has full rights of possession for the

term of the lease and, in addition, is owner of the farm implements necessary for the cultivation of his crops. But in the more highly developed industries it becomes impractical for the owner to attend to all the details of operation. Instead of undertaking the entire task, he appoints individuals who are capable of directing the specific work assigned to them. One man may be placed in charge of a purchasing department, one in charge of the office, another in charge of production, and so on. Generally, as the business grows in size, there is a tendency to subdivide the management more and more, the owner reserving to himself final decision on questions of general policy and general methods and delegating routine matters to the respective managers. In large concerns, the whole question of the expert operation of the business is likely to be delegated by the entrepreneur to management. Thus the function of the entrepreneur in large-scale business tends to be confined to the determination of the broad, general lines of policy, and to the appointing and holding of management accountable for the operation of that policy. The authority to exercise this influence over production flows from the legal right of private property. But as we shall see in the treatment of the corporation, there has been a further differentiation and, because of the diffusion of ownership resulting from the corporate form of organization, a different method of exercising the authority of the entrepreneur has been developed.

Forms of Entrepreneurship.—There are three principal forms of entrepreneurship, namely, the *individual proprietor*, the *partnership*, and the *corporation*. The individual proprietor is a capitalist-employer, who owns his business, furnishing a large part of the capital, and either performs the necessary work himself or hires the labor required by the undertaking. He may borrow part of his capital funds without losing his authority, except as indicated above, but he must ordinarily possess a considerable amount of property in order to gain the confidence of his creditors in his venture. The farmer is a good example of this type of an entrepreneur.

The advantage of single proprietorship, or the one-man business, lies in the fact that ownership and management are concentrated in one man who has staked his own fortune on the success or failure of the venture. This is a powerful stimulus that drives men to give to their business their full capacity. In such undertakings the owner can keep all of the details of the business in

mind and give them his personal attention. By so doing he can maintain unity in the management of the enterprise, a condition that cannot obtain in any other form of business organization.¹ It should be clear from these statements that the one-man business is likely to be small or confined to those enterprises that have relatively few details of supervision.

The limitations on this form of entrepreneurship should be evident. First, growth in size soon makes it impossible for one man to give personal attention to all of the details of a business. At this stage, and a business does not have to be very large to attain it, one of two things must be done, namely, he must hire assistants and undertake to supervise their work, or extend the ownership motive by changing the form of organization and by dividing the responsibilities among the new owners. Second, the amount of capital which one man has at his command is usually not sufficient to operate the largest of our industries. This limitation is relative, however, and is not as potent as is the growth in complexity due to an increase in size. There are some businesses of very considerable size that are privately owned or held by a family, or as an estate, and operated by an appointed management. Nevertheless, the difficulty in securing capital must be reckoned as one of the limitations on this form of entrepreneurship.

The Partnership.—As soon as we pass from agriculture and small industrial establishments, the type of organization is likely to change in the direction of collective entrepreneurship, which may take the form of a partnership or of a corporation. In the partnership, two or more persons bind themselves by an agreement to conduct a business jointly. The relations between the partners are determined by the articles of agreement, which may specify the share of profits that each partner is to enjoy and the proportion of the common losses that each must pay. However, the agreement cannot control the relations of the partnership with the general public for, from a legal standpoint, each partner

¹ The statements assume equal ability on the part of those in executive positions. It not infrequently happens that a hired manager may be called in to straighten out the affairs of a one-man business, but in such cases the hired man is likely to be superior in ability or training, or otherwise he would be unable to improve on the methods of operation already in use. Other things equal, the motive of ownership will give a superior form of management to that of a hired man who may have authority without full

is regarded as an agent of the partnership and is, therefore, liable for all of its acts. In case the partnership becomes financially involved, the creditors may proceed against any one of the partners for the full amount of their claims, notwithstanding any terms of the agreement that were designed to divide such losses among the respective partners.

The advantages of a partnership are first, the ease with which additional capital may be secured. Very frequently a partnership will be formed when a one-man business has reached a size that makes the securing of additional capital difficult. A man without business experience in a particular line, but with funds to invest, may join with one who has had experience, but who is looking for additional money as a means of expanding a profitable business. By pooling their resources, partners can conduct a larger undertaking than if each had endeavored to operate single handed. Second, the principle of specialization may be successfully applied in a partnership. One member of the firm may take charge of production, another may handle the selling, a third may keep the accounts, and so on. In this way, a partnership makes possible the enlistment of specialized ability which cannot be done when all of the functions of the entrepreneur are performed by one individual. Third, the ease with which an organization can be formed. Lastly, the policies of a partnership can be changed at any time by conference and agreement among the partners.

The weakness and perhaps the most positive limitation on the partnership as a form of organization is the fact of unlimited liability of each member of the firm. A creditor, as stated above, may levy on any one of the partners and may secure the full amount of his claim from that member. It is evident that only persons well known to each other would be willing to join in an organization which made each member fully liable for all the debts of the partnership. Another weakness is encountered in the case of the death of one of the partners, or if one wishes to withdraw from the firm. Under such circumstances, the remaining partners must either purchase the property rights involved, or find someone who is willing to buy these rights and who is at the same time acceptable to them. Unless one of these adjustments can be made, the partnership must be dissolved and its assets distributed. Such adjustments frequently entail

a serious interruption of the business and may cause an actual loss of property value.

The Corporation.—The most important form of collective entrepreneurship is the modern corporation. This organization is a creation of the state, and gets its right to exist from its charter or articles of incorporation, which is granted or approved by the state. When corporations first came into existence, their charters were special grants of the Crown, as in the case of the London and Plymouth companies, under which attempts at settlements were made in our early colonial history. Later, after the colonies had become separate states, the grants of authority were special acts of the legislature which, after the famous Dartmouth College case (1819), became binding contracts between the state and the corporation.¹ The granting of articles of incorporation was jealously guarded at this time. It was feared that because of the special powers given to corporations they would develop into dangerous monopolies, and it was not until after the middle of the nineteenth century that general incorporation laws were enacted by the states, which made the formation of corporations a comparatively simple matter. These laws specify the conditions that must be met and prescribe the methods to be followed by the incorporators. If the organizers fulfil the conditions laid down in the general laws and apply in due form to the proper state officials, a charter will be issued creating the corporation and defining its rights and duties.

In the eyes of the law, a corporation is a legal entity that is clothed with many of the attributes of a person, such as the right of contract, rights of property, etc. It is a kind of artificial, fictitious person, that is conceived to have an existence apart from the persons who were responsible for its organization. It differs in this regard from a partnership. For example, if a suit were brought by or against a partnership, it would be brought in

¹ The Dartmouth College case grew out of an attempt on the part of the state to modify the charter of Dartmouth College, which had been granted as a special act of the legislature. The decision of the United States Supreme Court held that a charter so granted was a binding contract between the state and the corporation and could not be altered or amended by the state without the consent of the corporation. As a result of this decision the common practice has developed of limiting the life of corporations, but this limit is no serious handicap to the continuity of a corporate existence because of the ease with which the rechartering of a corporation may now be effected.

the name of the partners, but a corporation may sue or be sued in the name of the corporation. But as a form of economic organization, notwithstanding the fact that it functions as an individual, a corporation is a modern method of pooling the resources of a large number of individuals for the purpose of undertaking a business venture, and in this sense it may be regarded as a form of collective entrepreneurship.

The business of a corporation is conducted through its regularly elected or appointed officials, but back of these persons there is usually a board of directors that has the authority to elect or appoint them. Hence, the officials perform managerial functions in connection with policies that have already been determined for them. So far as the board of directors determine these policies, its members may be said to be performing the functions of the entrepreneur. But a board of directors is not ordinarily free to determine any policy it sees fit to undertake because its membership is elected by the stockholders. The stockholders, who own the corporation and who have a right to vote, can determine within the provisions of the charter the fundamental policies of the business. The stockholders, therefore, are the real entrepreneurs and the corporation is the legal means by which they may pool their resources and perform the entrepreneurial functions. In actual practice, the stockholders usually delegate a large part of their authority to the board of directors, but so long as they have a right to vote and to elect the directors, the source of authority rests with them.¹

The importance of the corporate form of business organization can be seen from the following table taken from the 1920 census:

¹ In the above statement no attempt has been made to go beyond the general aspects of the corporation as a form of entrepreneurship, but the reader should be aware of the great variety of corporate forms that may vary materially from the simple structure assumed in this description. While any attempt to describe these variations would carry us beyond the scope of this text, yet the reader should realize that what has been said about the source of authority is based on the assumption that the stockholders have voting rights. In some of the modern corporations the large body of stockholders may not have these rights. They may have ownership rights in the property of the corporation with no voice in determining its policy, the voting rights being vested in a small group of officers or directors, who may or may not hold a large investment in the property of the corporation. The dangers of this form of corporation is that it gives authority without the wholesome check of ownership responsibility.

CHARACTER OF OWNERSHIP IN MANUFACTURING

Year	Number of establishments owned by			Average number of wage earners (in thousands)				Value of products (in millions)			
	Individuals	Corporations	All others	Aggregate	Individuals	Corporations	All others	Aggregate	Individuals	Corporations	All others
1919	138,112	91,517	60,476	9,096.3	623,477	875.1	597.7	\$62,418	\$3,536	\$54,744	\$4,137
1914	142,436	78,192	55,203	7,036.2	707,555	649.8	678.7	24,246	1,925	20,183	2,137

Of the 290,105 establishments reported in 1919, 31.5 per cent were corporations, which employed 86.5 per cent of the total number of wage earners, and produced 87.7 per cent of the value of the products. Similar figures for 1914 showed that 28.3 per cent of the establishments were corporations employing 80.3 per cent of the wage earners, and producing 83.2 per cent of the value of the products. During the five-year interval, the figures reveal a distinct movement toward the corporation as a form of business organization. Its position and growth in modern industry suggests that this form of organization has distinct advantages over single proprietorship and the partnership. Let us now consider what these advantages are.

First is the provision for limited liability of the stockholders. We have seen that in a partnership there is unlimited liability which means that any property owned by a partner may be attached in the settlement of the debts of the partnership. But in a corporation the holder of a share of stock is limited in his liability to the amount of his holdings. For instance, if a person owns a \$100 share of stock, all that he can lose or be held for in connection with the business of the corporation is the amount of his subscription, or \$100 in this case. In some instances, the shareholder has double liability, as in the case of the national banks, *i.e.*, he is not only liable for the amount of his holdings, but he may be assessed an additional amount equal to his holdings, in the event of bankruptcy of the corporation. In the case above, if the stock were national bank stock, the owner would not only lose his subscription, but other property could be attached for \$100 or a total liability of \$200. This distinction of limited liability as between the partnership and the corporation has been modified somewhat by legislation in recent years, giving some limitation of liability in the case of some partner-

ships, a situation that varies from state to state in accordance with the legislation on the subject.

A second advantage is the ease with which large capital sums may be accumulated. By means of limited liability the risk involved in a new undertaking or, for that matter, in a going concern, can be widely distributed and the loss in case of a failure will not be ruinous to any particular individual. This is a very positive advantage, especially in the case of new enterprises which may involve more than ordinary risks, but hold forth hope for unusually high returns. The corporation has not only furnished an effective means of spreading risk and thus fostered the accumulation of large capital sums necessary for the operation of business on a large scale, but it has stimulated savings by furnishing a convenient means of securing ownership rights in a business and of thus appealing to the acquisitive motive of individuals. The possibility of diffusing ownership among larger and larger numbers of people through the sale of shares of stock in corporations has by no means been exhausted. The social significance of this practice is very great.

Another advantage is the ease with which shares can be transferred. We saw in the case of a partnership that, in the event of the death or withdrawal of one of the partners, a great deal of inconvenience and disturbance to the business occurs. No difficulties of this character are experienced by a corporation because of the ease with which the shares can be transferred either by sale or by inheritance. Such transfers can be effected without disturbing in any way the business policy of the corporation. The long life of the charter with easy renewability insures permanence and stability that is not possible in either of the other forms of organization that have been mentioned.

A fourth advantage, which results from the ease in transferring shares, is the tendency for the control, and hence the management of the business, to fall into the hands of those most competent to operate it efficiently. While ease in transferring shares makes this desirable result possible, we hasten to say that there are other effects that are not so beneficial. The pride of ownership and the reputation of the firm mean much less to a stockholder than to a partner. A share owner is much more likely to be governed in his attitude toward a corporation by the yield of his stock than by any feeling of personal pride in the success of the concern. He is likely to sell his holdings when the business

is not showing favorable returns and to buy stock if he thinks the prospect for unusual profits is favorable. This practice is fostered by the development of the specialized markets known as the "stock exchange" where the transfer of shares is greatly facilitated.

Transferability may result in the control falling into the hands of the unscrupulous, who may use their favored position to take so-called profits out of the business at the expense of the business itself, or at the expense of the minority stockholders. While the two evils mentioned may often exist, the essential fact remains that ease in transferring shares increases the possibility of control of corporate business becoming concentrated in the hands of the industrially efficient. These various advantages account for the very large proportion of business that is now done through this form of organization.

In concluding this section it may be well to direct attention again to the fact that the authority exercised by the entrepreneur flows from the rights of private property in the business. As has already been suggested, this authority is exercised in a peculiar manner in the corporation. It rests, in the first instance, with the stockholders because they own the property. By virtue of their property rights, the stockholders elect or appoint a board of directors or other directing body which performs, during its term of office, all of the functions of the entrepreneur, such as that of determining the policy of the concern and of appointing the management which puts these policies into effect. While responsible to the majority stockholders, the board of directors acts freely, once it has been created, and continues to do so until its policies or actions are checked by the stockholders.

The simple line of authority from stockholder to management, just described, may be greatly modified in actual practice because of the variety of property interests that may be represented in a corporation. For instance, there may be both preferred and common stockholders with varying rights that are determined by the charter or by-laws of the corporation. Voting power may be exercised by both classes of stockholders or by the common stockholders only. Then corporations may raise large sums of money by the sale of bonds. The bondholders are those who have a prior lien on the property of the corporation, but who do not usually exercise authority in the operation of the business so long as the interest charges are regularly paid. In the event

a corporation is unable to meet its interest payments, however, the bondholders will force a receivership and the business will then be operated by the receiver until the bondholders' claims are adjusted. During this period the receiver, acting for the bondholders, exercises the functions of the entrepreneur. When the claims have been liquidated the property is returned to the control of the owners, unless in the meantime, there has been a financial reorganization which has changed the character of the ownership. Aside from the modifications due to the varied property interests that may be represented in a corporate structure, collective entrepreneurship is exercised through the corporation very much as it has been described in the foregoing treatment.

Diminishing Productivity.—As has already been stated, the function of the entrepreneur in modern industry is to decide what shall be produced and to determine how the productive processes shall be organized and carried on. In deciding the latter question, he must secure control over the factors of production and combine these in such a manner as will enable him to make the desired wealth with as little effort or cost as possible. In combining the factors of production, every entrepreneur soon discovers that there are some fairly definite principles governing the results that can be obtained from the use of these factors. He finds, for instance, that as he increases the amount of labor and capital applied to a definite area of land, say an acre, the product secured does not continue to increase relatively and may not increase absolutely. He may secure more product per acre but the increase will not be in the same ratio as the increase in the labor and capital used, and, in time, if more labor and capital are applied to the same area of land, the product secured may actually decline. If the rate of production per unit of labor and capital remained constant, it would be possible to secure enough wheat from a single acre to feed a nation, or the world. In like manner, if he is a manufacturer of shoes, he will discover that he cannot add more men and machines within a definite space and hope to maintain the same rate of production per man or per machine, or even to increase his total output indefinitely. It does not matter what line of industry the entrepreneur undertakes, whether it be agriculture, manufacture, merchandising, or transportation, he will be confronted by this principle of

production, which may be called, the "principle of diminishing productivity."¹

The principle may be briefly stated as follows: Whenever any one, or more, of the factors of production is held constant and the quantity of the other factor, or factors, combined with it is varied, the product secured may increase absolutely, and for a time even relatively, but after a certain point has been reached in this combination, the increase will not be in the same proportion as the varying factors. It will be noted that in this definition the statement is made that the product *may* increase absolutely. The significance of this qualification is that the amount of the varying factors applied to the fixed factor could become so great as to cause an actual decline in the product secured. It is evident that the number of machines and men employed within the walls of a given factory could be increased to such an extent that the total product would decline because the men and the machines were so crowded as to hamper the movement of the materials through the factory. Of course, no intelligent entrepreneur would be so foolish as to add the varying factors in such quantities as to cause an actual reduction in the amount of goods produced, but should the experiment be attempted there is no doubt that the results would be as indicated. In actual business operations, however, the entrepreneur is confronted with this production principle and his business decisions are affected by its existence.

The principle was discovered in connection with the use of land and was first called "diminishing returns in agriculture." Later studies have shown that, when properly stated, the principle is applicable to the other factors as well as to land, and is, in fact, a universal principle of production. While it is most commonly called the principle of diminishing returns, the better expression is that of "diminishing productivity." There are at least two reasons for adopting the latter expression. First, the principle really expresses a uniformity found in all production which affects the rate at which physical units of output can be obtained.² Second, the term, "diminishing returns," has come to imply a

¹ ELY, R. T., "Outlines of Economics," p. 384.

² Professor John Black in his "Production Economics" has emphasized this same point of view and suggests that the principle be called "diminishing physical outputs," pp. 275-310; TAYLOR, F. M. calls the principle the "law of diminishing output," "Principles of Economics," p. 137.

decline in the money returns received from the operation of a business enterprise. A decline of this character may result from a change in the attitude of the purchasers of the good without any change in the conditions of production. It would seem, therefore, that the expression which places the emphasis upon the conditions of production rather than on the money returns received from the business is preferable and will aid in avoiding confusion of thought concerning these two very different phenomena.

We may illustrate the principle by taking a hypothetical case. Suppose that we undertake to farm an acre of land with a given grade of labor, equipped with the tools ordinarily used in cultivating the crop to be grown. Let us represent the variations in the amount of labor and capital used by varying the number of days of service of each that we apply to the land. The results obtained might take some such form as the figures given in the following table.¹

Number of days' service of labor and capital used	Total yield in bushels	Bushels per day of service	Marginal increase or additions to previous product
I	II	III	IV
1	2	2	0
2	6	3	4
3	12	4	6
4	20	5	8
5	27	5.4	7
6	32	5.3	5
7	35	5	3
8	36	4.5	1
9	34	3.8	-2

¹ The reader should realize that there is no special merit in the figures given in this table. All that is claimed for them is that they portray a tendency which is encountered in all production. The results from an actual experiment would show this tendency even though the figures varied widely from those given here. Just what results would be obtained would depend upon the industry and upon the particular combination of factors that was employed. As will be shortly shown, a given product may be secured by more than one form of combination of the productive factors. For instance, there might be more labor and less capital, or more capital and less labor, or the amount of land might be varied in relation to the amount of labor and capital employed. The assumption that the amount of labor and capital

Diminishing Total Output.—The first thing to observe in this table is the fact that the total yield, as indicated in column II increases up to the application of 8 days of service, when a decline begins in the total amount of product obtained. In this illustration, with a given system of cultivation, the maximum product per acre of land is secured when 8 units, or days of service of labor and capital are employed. The yield is 36 bushels per acre. Thereafter any attempts to increase the total yield by applying more labor and capital would give an actual decline in the amount of product received. To get the varying results set forth in the above table from a given acre of land would require separate seasons, which would, of course, introduce another variable element into our calculations. We can avoid this effect by assuming that the experiment is conducted simultaneously on different plots of land of equal fertility. The fact that government experimental stations regularly conduct tests of this general character is assurance that this assumption is not out of harmony with actual experience. By assuming simultaneous experiments, the results shown can then be attributed solely to the variations in the amount of labor and capital used. All that is claimed for this hypothetical case is that it represents a tendency found in all types of production. Increasing the amounts of the varying factors used in relation to a constant factor will ordinarily give for a time an increased product, but later on further additions of the varying factors will show an actual decline in the total yield.

Diminishing Average Output.—From this fundamental tendency in production two further observations can be made. First, the average yield per unit of the varying factors (in this case per day of service) may increase for a time, but will then begin to decline. It will be observed that the maximum yield per day of service, as indicated in column III, is attained with the application of 5 days of service. Up to this point the yield per day of service has increased from 2 to 5.4 bushels. We may call this point of maximum yield per varying factors, the point of diminishing productivity, for if we continue to cultivate the

used fluctuates uniformly is oversimplified in order to stress and illustrate the operation of the principle. However, there is usually at any one time one most economical combination of the factors of production which is affected by the principle of diminishing productivity plus the cost of the respective factors.

land more intensively, the yield per unit of the varying factors will decline, even though the total yield continues to increase. It is this tendency for the rate of production to fall behind the rate at which the varying factors are applied to a constant factor that is called the "principle of diminishing productivity." The point of diminishing productivity is where the yield per unit of the varying factors is at a maximum. The location of this point will vary, of course, with the results obtained from the particular combination of the factors.

Diminishing Marginal Output.—The second observation is that as the amounts of the varying factors are increased, the additions to the previous volume of product may also increase for a time, but will later decline. These additions to previous output are usually called "marginal products," or the product at the margin of cultivation. The question may well occur to the reader, Why should a farmer cultivate his land beyond the point of diminishing productivity, *i.e.*, beyond the point of maximum yield per unit of labor and capital applied to his land? The answer to this question is clear enough, but it involves an issue quite different from the one which we are now considering. He will continue to cultivate his land intensively as long as it pays, that is, as long as the returns from the sale of the additional product are sufficient to offset the added expense involved in securing it. A balance may be reached at any point in the series of results shown in the table and is, in fact, the effect of two forces, namely, diminishing productivity on the one side, and diminishing utility of the product on the other. This latter force is in reality a problem in value, and must, therefore, be postponed for subsequent treatment. For present purposes it is sufficient to say that the extent to which the intensity of cultivation is carried depends upon the demand for the product. Put in other words, the farmer will continue to cultivate his land as long as the purchasers of his product are willing to pay for the expense involved in securing the additional volume. The recognition of this fact explains why the farmer in our hypothetical case does not stop his cultivation at the point of diminishing productivity, or where the output per unit of labor and capital applied is at a maximum. It will be noted that the maximum addition to previous product, as indicated in column IV, occurs with the application of 4 days' service and thereafter, while the total yield may increase, the additions will be at a

diminishing ratio. It will be observed that the rate of decline of the marginal increments is much more rapid than that of the average output per unit of the varying factors. This fact is of great importance to the entrepreneur for, when he is considering whether he should extend his cultivation or not, he is concerned with the additions to the product that can be obtained by the application of more labor and capital to his land. The operation of the principle of diminishing productivity is a very positive force affecting his decisions on this question.

This principle can be portrayed by means of a graph, as follows:

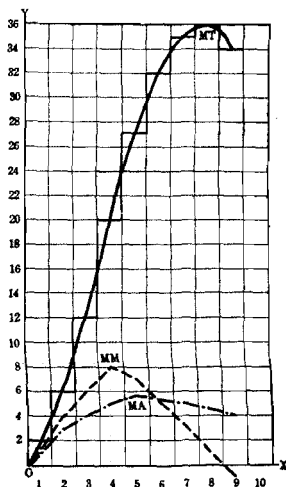


FIG. 1.

Along the line OX , equal spaces are marked off to represent equal areas of land of equal degree of fertility, and along the line OY is represented the amount of product per plot. Each rectangle represents the yield per plot obtained by varying the amounts of labor and capital applied to the land during the same

interval of time, as in a single season.¹ With 1 day of service on plot number 1 the yield is 2 bushels, with 2 days on plot number 2, 6 bushels, and so on until it is found that 8 days of service on plot number 8 give a maximum yield per plot, as indicated by *MT* in the graph. Any further attempt to increase the total product would result in a decline in the yield as is represented by the area of the ninth rectangle. Since this experiment is conducted simultaneously on plots of land of equal size and fertility we may attribute the results solely to the varying factors, and assume that identical results would be obtained from a single plot of land, if the labor and capital could be varied in accordance with this hypothetical experiment. The figure may, therefore, be taken to represent the results that can be obtained from a single plot of land (or the constant factor) when the amount of labor and capital applied to it is varied and no other changes occur to affect the rate of production. By reducing the size of the units of capital and labor sufficiently, these rectangles will approximate straight lines and instead of a broken series of steps we would have a curve *OMT*. This curve represents the course of production when one factor is held constant and the other factors are varied.

In these same rectangles, it is possible to represent both the average output per day of service and the additions to previous product, conditions which depend upon the production principle just explained. With 5 days of service the average yield per day reaches a maximum and is indicated by *MA*. With 4 days of service the addition to previous output reaches a maximum and is indicated by *MM*. By assuming that the units of labor and capital used are infinitely small, we get the curves *OMA* and *OMM*. It is the latter curve that first shows the effect of the principle of diminishing productivity, as the rapid decline of the additions to previous product is the direct result of this production principle.

The reader should understand that the shape of these curves depends upon the results that are obtained from the combination of the factors of production. It may well happen that the rate of production will decline at once instead of increasing as it has in this hypothetical case. For instance, there are many kinds of machines which can be operated most efficiently by one

¹ Equal degree of fertility is assumed to avoid the seasonal variations, which would affect the quantity produced, if the time element were not provided for.

man, and if more than one man were employed on such a machine, the output per man would doubtless begin to decline immediately. This condition would be reflected in the total yield and, hence, would change the shape of the curves shown above; however, the results would conform to the general principle of production that has been set forth in our hypothetical illustration.

This principle is of very great significance in production. Every farmer knows that it is impossible to increase at a constant ratio the output of corn or wheat or other agricultural crop on a fixed area of land. He may know that by intensive cultivation, by fertilization, or by other methods of farming his land, he can greatly increase the output, but the returns will not keep pace with the increasing effort necessary to secure the additional crop. If it were not for this principle of production, there could be no such thing as scarcity of land. Additional products or services could be secured by applying more labor and capital to very limited areas of land. But since nature does not respond with equal force to additional quantities of labor and capital, the users of land must experiment and decide when it is advantageous to use more acres and when a larger product will be obtained from cultivating the existing acres more intensively.

Principle Universal in Application.—As set forth up to this point, the principle of diminishing productivity has been illustrated as applicable to land. However, as has been suggested, it is equally applicable to any of the factors when one, or two, are held constant and the other factor, or factors, are varied. Suppose that we represent a given combination of the factors of production as follows: x labor plus y capital on z land yielding P product.¹ If we should vary the amount of land over which the given amount of labor and capital was spread, we would find that after a certain point had been reached, the product would tend to decline in terms of the varying factor or factors. In this case, x plus y plus z might yield P , but x plus y plus $2z$ would not necessarily yield $2P$, but more likely less than $2P$. Every farmer knows that, with a given equipment of machinery, he can

¹ Those who are familiar with algebra should have no difficulty with this method of expression. The literal numbers are chosen to represent any unknown amount of the factors which, when combined, will give an unknown product represented by P . The important thing to note is that the rate of production in no instance remains constant.

cultivate effectively only a limited area of land and, if he undertakes to farm more acres, he will not get proportionate returns because he is unable to cultivate the larger area effectively and may suffer an actual loss. Here the principle is shown operating where land is the varying factor and labor and capital are held constant.

But labor and capital are not a single homogeneous factor but two separate, and often competing, factors of production. The principle of diminishing productivity will apply to any combination of the factors, so long as one or more of them is held constant. If we assume that x labor plus y capital on z land will yield P product, then, if we vary the amount of labor and hold capital and land constant, we will find that after a certain point is reached, the product will tend to decline per unit of labor used. Thus, x plus y plus z yields P , but $2x$ plus y plus z will not necessarily give $2P$, but more likely less than $2P$. In other words, the amount of labor that can be effectively used with a fixed amount of capital on a fixed area of land is subject to this principle. Likewise, if capital be held constant and labor and land varied, the principle will be found to operate, for after a definite point has been reached an additional supply of labor working with a fixed supply of capital applied to an increased area of land will produce a diminishing quantity of product per unit of labor and land used.

In like manner, by holding labor and land constant and varying capital the principle will appear also. If we combine x labor with y capital on z land and get P product, it will not follow that by combining x plus $2y$ plus z that we will necessarily get $2P$, but more likely the product will be less than $2P$. This statement assumes that there has been no change in the character of the capital instruments used, but simply an increase in the number of the existing kind of tools. When stated in this way, the principle of diminishing productivity applies to all the factors and may be regarded as a general principle of production.

It has sometimes been said that this principle is applicable to agriculture but not to manufacturing; that more labor and capital in manufacturing will give constant, if not increasing, output. This statement involves two confusions. First, it fails to recognize that one or more of the factors must be held constant, while the other factors are varied. If on a fixed area of land a factory building be erected and equipped with

machinery of a particular kind, it will be found that, after a definite number of machines adequately manned have been installed, any further addition of machines and men will not give proportionate returns and, in fact, very soon the men and machines will be in each other's way and cause an absolute loss of product. In the second place, the above statement fails to distinguish between the principle of diminishing productivity and the economy of large-scale production. In the one case, the important question is, What is the best combination of the factors of production? In the other, What is the best size of plant or factory in which to carry on production? The former deals with the principle of diminishing productivity which, as has already been shown, applies to all industries; the latter with the relative productivity of an industrial unit in which the various factors, land, labor, and capital are combined in varying proportions.

Even in connection with the problem of the size of plant, the question of diminishing productivity arises in the form of the limits of administrative organization. Can managers be found who can direct an industrial plant with equal efficiency as the plant increases in size? Much will depend upon the diversity of the problems of administration as to how large a plant can be before diminishing productivity of management will make its appearance. If the nature of the industry is such as to permit of simplification and standardization of the administrative problems so that increase in size means nothing more than applying standardized tests to additional units, the size of an industrial plant that can be operated under a single management without loss of efficiency is exceedingly great. The United States Steel Corporation and the Ford Motor Company may be cited as illustrations of very large plants that are efficiently operated. However, it is probably correct to say, even here, that the principle of diminishing productivity applies but that the size of plants which can be operated under a single management without showing a tendency to diminishing productivity, due to the difficulties involved, varies from industry to industry on account of the differences in the nature of the industries themselves. It is by no means clear that productive efficiency is always promoted by large-scale management. Often the motive for the development of large-scale management is to be found in the desire to control industry for financial purposes, or for the monopoly power which may be developed by

means of centralization of control. Such motives may or may not coincide with productive efficiency. They may emanate from purely private or acquisitive desires rather than from the desire to secure the largest output for society with the least expenditure of human effort. There can be little doubt that the principle applies to management, even though the scale of production may vary widely from industry to industry.

Principle of Variable Proportions.—There are two corollaries of very great significance to the entrepreneur that flow from the principle of diminishing productivity, namely, the *principle of variable proportions* and that of the *least cost combination*. We will now consider these in order. In chemistry, we learn that elements combine in definite proportions. For instance, the combination of two atoms of hydrogen with one atom of oxygen will produce water, and this is the only way by which this compound can be obtained. What is true in this instance is also true in all other chemical combinations, or, in other words, a law of definite proportions governs the combination of the various chemical elements. But in production the same result can be obtained by a combination of the factors of production in a variety of ways. The farmer can secure a definite quantity of wheat by cultivating a large acreage, using relatively small amounts of labor and capital, or he may secure the same quantity from a smaller acreage which he has cultivated more intensively. He can use more labor, or more commercial fertilizer, or both, on the smaller area and secure the same volume of wheat. This is what is known as the "principle of variable proportions." It enters into the decisions of every entrepreneur who has the responsibility of combining the factors of production.

Professor Carver gives an account of an interesting experiment that was conducted for a number of years on the Rothamsted estate near London.¹ This experiment illustrates the principle of diminishing productivity and the way in which the problem of combining the factors of production presents itself to the entrepreneur. In this case, five plots of land of approximately equal fertility were treated alike in all respects except that different quantities of nitrogen were applied to the separate plots. To one 43 pounds were applied; 86 pounds to another;

¹ CARVER, T. N., "Principles of National Economy," pp. 479-481. The tables are quoted from an article by Eugene Davenport in Bailey's "Cyclopedia of Agriculture."

129 pounds to another; and 172 pounds to another. The following table shows the results of the experiment:

TABLE I¹

Plot	Fertilizer	Acreage yield in bushels for 8 years	Gain for 43 pounds of nitrogen
No. 5.....	Mixed minerals alone	19	
No. 6.....	Mixed minerals plus 43 pounds of nitrogen	27½	8½
No. 7.....	Mixed minerals plus 86 pounds of nitrogen	35½	7½
No. 8.....	Mixed minerals plus 129 pounds of nitrogen	36½	1½
No. 16.....	Mixed minerals plus 172 pounds of nitrogen	37½	½

¹ CARVER, T. N., *ibid.*, p. 480.

According to this table, the yields diminish for each successive dose of 43 pounds of nitrogen. The gain on Plot 16 over Plot 8 was only ½ bushel, so that this plot was discontinued after 8 years, but the other four were continued for 48 years with the following results:

TABLE II¹

Plot	Average annual yield in bushels	Gain for 43 pounds nitrogen
No. 5.....	15	
No. 6.....	24	9
No. 7.....	33	9
No. 8.....	36¾	3¾

¹ *Ibid.*, p. 480.

The results of this experiment as shown in Table II are interesting even though the number of plots is too small to be conclusive. The return per dose of 43 pounds of nitrogen on plots 6 and 7 is constant, while the third dose on plot 8 gives diminishing output per dose of nitrogen.

Least Cost Combination.—Up to this point the discussion has been in terms of physical output, and the results illustrate

the operation of the principle of diminishing productivity. While the decisions of the entrepreneur are affected by the principle of diminishing productivity and the law of variable proportions, the form in which these principles confront him is that of costs, or expenses, involved in securing control over the factors that are required in his industry. It would seldom be desirable to combine the factors in such proportions as would give a maximum yield for any one of them, unless the varying factors were free goods. Since this is a condition that never exists, it then is necessary to combine the factors in such proportions as will give the largest return with a minimum of expense or, in other words, to find the least cost combination. The consideration of this topic transfers our discussion into the field of value, a subject that is treated in detail in subsequent chapters, but since purchase and sale are so generally understood, the simpler phases of the problem may be presented here as a part of the problem of least cost combination.

Turning again to the experiment referred to above, we find the following results when the question of costs is introduced into the problem. It is assumed that the cost of 43 pounds of nitrogen is \$6.50 and the price of wheat is \$1 per bushel. Under these circumstances the results would be as follows:

TABLE III¹

Plot	average annual yield in bushels	Gain for 43 pounds of nitrogen	Value of gain	Cost of gain	Profit or loss
No. 5.....	15				
No. 6.....	24	9	\$9.00	\$6.50	\$2.50
No. 7.....	33	9	9.00	6.50	2.50
No. 8.....	36¾	3¾	3.75	6.50	2.75 loss

¹ *Ibid.*, p. 481.

It will be readily seen that had the price of wheat or the cost of nitrogen been different from that assumed in Table III the profitableness of the method of culture would have been greatly modified. If wheat had been \$2 per bushel then the third dose of nitrogen on Plot 8 would have yielded a \$1 profit instead of the \$2.75 loss shown in the table. But if wheat had been only 50 cents per bushel there would have been a loss on every dose of nitrogen used. This experiment illustrates the kind of problem

that confronts the farmer in growing his crops. It will be observed that the only variable was that of the quantity of fertilizer used. The same kind of question would be confronted in respect to tools and machinery employed and also in respect to the use of labor in place of machinery, or vice versa.

What has been said here concerning agriculture applies with equal force to all other lines of production. Enough evidence has been presented to show the significance of these principles in the operation of modern industry. Both the volume and the costs of production may be greatly affected by the intelligence exercised on the part of the entrepreneur in combining the factors of production. He should always economize in the factor that is expensive and use the cheaper ones more liberally. When land has a high value it will be used sparingly, as when it is cultivated intensively, say, for truck gardening, or when it is used as a site for the erection of high buildings in the business districts of large cities. Likewise, if labor is cheap and capital expensive, the entrepreneur will employ large quantities of labor and few tools and machinery; but if labor is expensive, then the wise business man will economize in its use and install machine methods of production.

To maintain a proper balance, the entrepreneur must watch both the output and the costs. The costs should be separated so as to enable him to measure approximately the significance of additional units of the varying factors, in whatever form the combination of these factors may present themselves, whether it be more labor and less capital and land, or more capital and less labor and land, or any other combination that may be made. It can be laid down as a general rule that it will pay to continue to add units of any one of the factors so long as this addition yields a product that is equivalent to the costs involved in securing the use of that unit. If, for instance, it costs \$4 per day to employ a laborer, the minimum product which his services must yield in order to make it worth while to hire him is the cost of securing those services, or his daily wage. It is customary to call this additional product "the marginal product," and it should be noted that the marginal product of any factor will tend to decrease as the supply of that factor increases in relation to the other factors. In other words, when the supply of labor is increasing faster than the supply of land, or of capital, we say that the marginal product, or marginal significance of labor, is

falling. Labor would, under the conditions assumed, be relatively less significant as a factor of production and, hence, any one unit of labor would add relatively less product than if the supply of labor were scarce. In like manner, and for the same reason, when the supply of land is plentiful and there is a relative scarcity of labor and capital, the marginal product of land would be small and will, therefore, command relatively a small price.

The Highest Profit Combination.—These two principles have an important bearing on the decisions of the entrepreneur. The law of variable proportions is in reality a production principle that affects the combination of the factors of production in securing any definite quantity of physical units of output. The entrepreneur deals with these combinations in terms of value, but since he can secure the same quantity of output by different combinations of the factors, he will naturally be affected in his decisions by the relative cost of the factors used. The question remains for discussion as to whether he will always stop production at the point of least cost combination. In thinking of this question one should bear in mind that the entrepreneur is concerned primarily with the net return from the operation of his business. The question may now be asked whether the point of least cost combination and that of highest net return always coincide. The relation of these two points may be illustrated by the following graph:

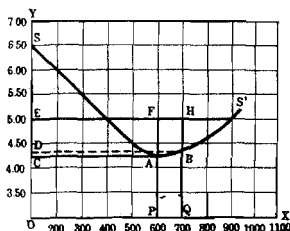


FIG. 2.

Along the line OX is indicated the number of units of product, and along the line OY is measured both cost and selling price in dollars. Let SS' equal the cost per unit as the number of units increases. Let OE equal the selling price. The least cost com-

bination would be at P with 600 units of product at a cost of \$4.25 per unit, and is represented by the rectangle $OPAC$. The profit in this case would be $CAFE$. If production should be carried to Q with 700 units, the cost per unit would rise from PA to QB , let us say from \$4.25 to \$4.30 per unit. The total cost is now represented by $OQBD$ and the profit by $DBHE$. The profit obtained under the circumstances assumed would be greater when production was carried to 700 units, notwithstanding the fact that the least cost combination is found with the production of 600 units. The answer to the question, then, is that the point of highest profit combination and that of the least cost combination do not necessarily coincide when one factor is held constant. This fact is accounted for by the attitude of the buyers. If demand is fairly elastic and the market can be expanded, as indicated in this illustration, production may be carried profitably beyond the point of least cost combination. If all the factors are regarded as variables, the evidence seems to point to the conclusion that these two points will coincide. Since the entrepreneur is engaged in business for the net returns which he can secure, the location of the point of highest profit combination may affect the combination of the factors that he employs.¹

The Marginal Choices.—In deciding what combinations to make, the entrepreneur will usually add more units of any one of the factors as long as the return from its use will be sufficient to offset the costs involved in securing it. If an additional machine does not increase the product sufficiently to pay for itself during its lifetime, it will not be advantageous for the entrepreneur to purchase it. In like manner, if an additional laborer does not increase the product enough to cover his wages, there will be no inducement to employ him. When any factor has been utilized to the point where the return just offsets the increased cost due to its use, we may say that the margin of use for that factor has been reached. It is at the point of marginal uses that the entrepreneur normally stops the combination of the factors of production.

Finally, it may be observed that the principle of variable proportions, or of proportionality, appears in maintaining a balance

¹ Here again it must be urged that the significant thing is not the figures used, but the recognition of the operation of the principle. For more complete treatment of this whole subject, consult Black, J. D., "Production Economics," pp. 275-344.

in the utilization of the resources of a nation. The economic prosperity of a nation is highest when its resources are utilized in a manner that will yield the maximum satisfaction of human wants. If too much land is devoted to the growing of wheat and too little to grazing purposes, wheat prices will be unduly low and meat products unduly high. If too much land is devoted to agriculture and too little to timber, the prices of farm products will be unduly low and those of lumber and wood products unduly high. If there should be an over-specialization of capital equipment in a particular industry, say in the railways, as was experienced in the United States following the Civil War, or an over-development of the automobile industry and a shortage of railway equipment, as seemed probable at the close of the World War, the productive power of the nation will be impaired by virtue of this over investment in the lines mentioned. Or, in the field of labor, if too many men are factory hands and too few are building laborers, the prices of factory products will be low while those of houses will be high. The welfare of a nation demands a balancing of all of its resources. To some extent the normal market forces distribute the use of the various factors of production in the direction of their most intensive use, but the individual does not ordinarily have the information on which to determine what is the best use of these factors. Price serves as a rough index, but at times it may deceive the managers of industry. The unconscious forces that work through the market to distribute the factors in the direction of their most significant use are giving way to some extent to a conscious policy based upon an intelligent analysis of the facts surrounding the conditions of production. An extension of scientific method to the problems of management, both public and private, will increase conscious control over the productive processes and will not only increase the volume of wealth produced but will also result in a more economical and better-adapted utilization of the natural resources of the country.

Suggested References

- BLACK, JOHN D., "Production Economics," Chaps. XI-XIII, and XVIII-XXI.
BYE, RAYMOND T., "Principles of Economics," Chap. IX.
CARTER, GEORGE R., "The Tendency toward Industrial Combination."
CARVER, THOMAS N., "Principles of National Economy," Chaps. XIV and XVII.
———, "The Distribution of Wealth," Chap. II.

- CHAPMAN, S. J., "Outlines of Political Economy," Chaps. XI and XIII.
ELY, RICHARD T., "Outlines of Economics," 4th Ed., Chap. XIII.
FAIRCHILD, FURNISS, and BUCK, "Elementary Economics," Vol. I, Chap. VI.
HAYES, H. GORDON, "Our Economic System," Vol. I, Chaps. XII, XIII, and XV.
LOUGH, WILLIAM H., "Business Finance," Chaps. I-IV.
MARSHALL, ALFRED, "Industry and Trade."
MILL, JOHN STUART, "Principles of Political Economy," Book I, Chap. XII.
MITCHELL, WESLEY C., "Business Cycles," Chap. VI.
TAUSSIG, FRANK W., "Principles of Economics," 3rd Ed., Vol. I, Chap. VI.
TAYLOR, FREDERICK M., "Principles of Economics," Chaps. VIII-IX.

CHAPTER VII

RISK AND PRODUCTION

Nature of Risk in Production.—In the organization of production we have seen that through the division of labor one group of persons produces the goods which another group consumes. At first sight it might appear that the consumer is in a better position to estimate his needs and, hence, to control production than the producer, but in practice the responsibility for making these estimates devolves upon the producer. It is the entrepreneur, or where this function has been delegated it is the management, who decides what shall be produced and in what quantities. Production is carried on for a general market ordinarily in anticipation of demand, and frequently without contracts in advance for the goods. The estimates of demand may be erroneous. They may be too large or too small, and when we remember that the amount produced is the result of the estimates of many entrepreneurs who do not cooperate in determining the quantity to be produced, it becomes evident that more or less uncertainty is inevitable.

The nature of this uncertainty may be illustrated by a few concrete examples. When the farmer plants his crops in the spring or fall, as the case may be, there is a large element of uncertainty as to the size of the crop he will be able to harvest. Much will depend upon the character of the weather and freedom from pests of many kinds that may attack his crop and materially affect the output. There may be too much or too little rain to mature the crops, or the weather may be too hot or too cold for the plants to develop.

In addition to the hazards affecting the production of his crops, there are also those that may be just as serious that affect the demand for them. While the farmer is engaged in producing many of the necessities of life, the demand for his products is by no means assured, as conditions quite beyond his control may affect the sale of his crop. Heavy harvests in competing regions, as wheat grown in Russia or Argentina, may depress the price

at which the American farmer can dispose of his products. While changes affecting the conditions of demand are more serious in the case of many other forms of production, yet the farmer is by no means free from hazards arising from this cause.

The farmer is not the only producer who is confronted with uncertainty in the conduct of his business. The manufacturer ordinarily must make up his goods in advance of the demand for them, and it is extremely difficult for him to estimate how large a volume of goods the consumers will use. Take, for instance, the case of a manufacturer of clothing. The volume of clothing purchased in any season is affected both by weather conditions and by the behavior of the purchasers in respect to styles. If a summer season happens to be late and cool, summer goods will not be taken from the market as rapidly or in such quantities as when it is hot. The manufacturer must either have laid in a sufficient stock of materials to take care of his estimated summer trade, or he must have contracted for them in advance of their delivery. In either case, he has assumed obligations which he hopes to meet from the sale of goods. Expected profits may be turned into losses as the result of weather conditions that are beyond the control of the manufacturer.

In addition to these seasonal hazards, the manufacturer is confronted with style changes. Cotton goods may have been in vogue and suddenly give way to woollens, or woollens to silk. Fashion may dictate one or the other of these goods, and unless the manufacturer has foreseen the change he will be caught with unsold goods on his hands and thus suffer loss. His plant and equipment and his labor force have been adjusted to, and trained in, the handling of one type of material, and the change in fashion will cause a readjustment both in the factory equipment and in the direction and supervision of the labor force. It is evident that if either weather conditions or style changes are unfavorable to him, the manufacturer will be caught with a stock of goods which he cannot dispose of at a profit.

In like manner, the merchant must stock up his shelves in advance of his sales, and he, too, is affected by the seasonal demands and the individual tastes of his customers. He has to face the uncertainty of disposing of the goods which he has purchased from the manufacturer. The greater the length of

time intervening between the beginning of the processes of production and the completion of the finished goods, the greater the degree of uncertainty that is likely to be present. Hence division of labor and the "round-about" methods of production, which are so characteristic of our modern economic organization, contribute very greatly to the uncertainty in business. It is this uncertainty that gives to all business its speculative character and introduces what is generally known as the risks involved in business. By risk, as here used, is meant the uncertainty existing in the outcome of any part of the productive process.¹

Causes of Uncertainty.—As has just been indicated, uncertainty in business arises from a number of causes, which may be classified as follows:

1. *Uncertainty Due to Physical or Natural Forces.*—Here would be classed the uncertainty arising out of general weather conditions, storms, floods, lightning, etc. The volume of loss from these causes is very large. Society is endeavoring to deal with some of these by forecasting weather conditions, but in the present state of knowledge, our ability to prevent loss and damage is quite limited; therefore we may expect a large amount of risk to continue in industry from this cause.

2. *Uncertainty Arising from Ignorance of Scientific Principles of Production.*—Among the great mass of people engaged in production, there is a large amount of empirical knowledge that has been handed down from generation to generation, which has been sufficient to carry on the simpler forms of industry. With the development of large-scale production, this empirical knowledge has had to give way to scientific and experimental knowledge because of the complicated processes involved. Lack of exact information concerning the materials and forces that enter into production introduces an element of uncertainty. This uncertainty may be due either to the actual limits of human knowledge concerning these forces or it may be due to the fact that the verified knowledge concerning the productive processes has not been disseminated among those who are engaged in industry. In either case, lack of knowledge introduces a degree of uncertainty in the outcome of production.

¹ The fact that uncertainty can be measured and a method devised by which the burdens can be distributed in a definite cost does not change the fact that risk is present.

3. *Uncertainty Arising from the Unpredictable Character of Individual Conduct.*—Some men may steal, or forge, or fail to pay their debts. Then, also, dissatisfaction and unrest may result in strikes, riots, or even wars. Since it is not possible to predict these events accurately, they result in a greater or less degree of uncertainty in the conduct of business. For this reason, it is impossible to foretell precisely what, and how much, the consumers will buy. This fact gives rise to market fluctuations, so that it is not often possible to match with precision the beginning and the end of production. There will always be a considerable degree of uncertainty in production due to this cause.

4. *Uncertainty Due to the Time Element.*—Uncertainty may also arise because of fluctuations in prices and conditions of production occurring between the beginning and the completion of the productive process. This cause may be regarded as the influence of time on our economic life and is, in fact, a summation of most all other causes of uncertainty. But since its manifestations through price fluctuations are so significant, it is put down as a separate cause. The larger speculative gains today are not secured through differences in the price of commodities in different markets. Improved methods of transportation and communication have reduced these differences to a minimum. The large gains are secured from differences in price from month to month or season to season or, in other words, from differences due to the element of time. While the uncertainty arising from the attempts to match the beginning and end of production in advance may involve all forms of risk, the time element is so significant in most business transactions that it justifies special mention.

5. *Uncertainty Due to Governmental Action.*—Lastly, production is materially affected by the action of governments through legislation, tariffs, taxes, etc. No one can know how the next session of the legislature will affect his business, which is only another way of emphasizing the uncertainty which he faces from this cause.

Methods of Dealing with Risks.—The presence of risk in production should be apparent to every observer.¹ The next problem is how to deal with these risks. There are three principal ways of treating the risks arising in the conduct of modern industry, namely, the attempts at reduction, or elimination, the

¹ HARDY, C. O., "Risk and Risk-Bearing," Chaps. II, III, and IV.

assumption, and the distribution of risks. These methods will require extended discussion.

1. *Reduction or Elimination.*—The elimination of risk means the substitution of certainty for uncertainty in the conduct of a business. Certainty may be introduced to a limited degree in a variety of ways, some of which are considered here:

(1) Among the most important methods of eliminating or preventing risk is the application of the results of invention and discovery. The introduction of safety devices, the installation of a sprinkler system to prevent fires, the building of dykes to prevent floods, and innumerable other examples which might be cited as methods whereby science and invention have contributed to the elimination of risk. In engineering and other scientific fields as well as in most professions, this influence has been at work to eliminate uncertainty. But scientific methods have not been applied to the same extent in business administration as in these other fields of endeavor. As our knowledge of business becomes more exact and the results of research and discovery are applied to the planning and operation of industry, there is reason to believe that uncertainty in business will be reduced.

(2) In the second place, much risk may be eliminated by market analysis and scientific forecasting. The modern merchant analyzes his market by studying the possible customers in a total population and, on the basis of past experience, establishes a definite amount or quota to be sold. Exact information concerning the buying power and buying habits of a people is a necessary condition for the estimation of the potential market for any good. In like manner, the manufacturer may be able to build up guides by which he may control his production by the method of market analysis. In the factory, also, the introduction of scientific management tends to give a greater degree of certainty in the quantity of goods produced. The limits of this method of eliminating risk are the costs involved in finding the facts necessary to establish and maintain the method. Some economy in spreading these costs may be obtained through cooperation in supporting the research required. Firms engaged in like lines of business may unite to support a research agency, or they may apply to a private agency formed for the purpose of furnishing this information. In recent years, there has developed a large number of private, investigative agencies that sell their services to manufacturers or other persons engaged in

production. These agencies maintain the statistical and research facilities for the gathering and interpreting of information and for scientific forecasting. As examples of such agencies the Harvard Economic Society, the Roger Babson Service, the Brook-mire Service, etc. may be cited.¹

(3) A third method is that obtained by the consolidation of individual risks into large groups. It often happens that we have very accurate information concerning large groups of data, while we have very limited knowledge concerning the individual items that make up the groups. Thus, death rates, marriages, and many other similar contingencies may be reduced to averages that give a high degree of certainty for a large population group, although they are of little value when applied to an individual instance. The death rate in large cities is known with reasonable certainty for most diseases, but no one can tell what individuals may be stricken by any of them. In the case of fire, the risk to an individual house owner is large and the uncertainty great, but by combining a sufficient number of houses into one group the liability of loss by fire can be reduced to a high degree of certainty, and the risk when distributed over the whole group becomes relatively small for each house owner. In so far as business judgments can be based upon the certainty of an average instead of the uncertainty of a single instance, this method assists in eliminating risk.

(4) Lastly, risk may be reduced, and to that extent eliminated, by the method of compensation, *i.e.*, by combining business in such a way that the profits in one line will offset or compensate the losses in another. Hedging contracts, to be described later, are an illustration of this method.

2. The Assumption of Risk.—After the most scientific methods known to man have been applied to their fullest extent as a means of removing uncertainty, there will remain a large amount of risk in production that will have to be assumed in one way or another. We will now examine the methods of distributing or assuming risks.

As industry is now organized the individual entrepreneur, in first instance, assumes the risk. As owner and manager of the business, he undertakes to bring the factors of production

¹ An example of the way a firm may use scientific methods to reduce risks is described in H. B. VANDERBLUE's, "Problems in Business Economics," Chap. VII.

together and to produce the goods for the market in advance of the demand for them. He agrees to pay wages to his employees, and to pay interest on any capital funds loaned if he has to borrow for the conduct of his undertaking. Should his venture be unsuccessful, his creditors have a prior lien on all of his property, so that he assumes the risk of the business up to the full value of his investment. When the entrepreneurship takes the form of a partnership or a corporation then the risk follows the line of liability which obtains in these two forms of organization. Unless the partnership be limited as to liability, each partner is responsible up to his total property holdings (subject to any bankruptcy laws that may provide for a limited exemption). In the corporation, the stockholders run the risk of loss of their investment or, as has been explained, there may be double liability in some cases. In other words, in the event of a failure of a business, where double liability has been imposed by law, the stockholders may have any other property that they own attached for an amount equal to their holdings in the corporation. In these various ways risk is first assumed by those who engage directly in any kind of industry and losses from such business ventures are first assessed against the property of these persons.

But there are still others who assume risks besides the entrepreneurs. Next in importance are the creditors who advance money to those engaged in business. If an individual proprietor borrows on a personal note, the lender runs the risk of his inability to pay back the principal and interest. Should a partnership become involved, the lenders lose only if the total loss exceeds the value of the total property owned by the partners. The bondholders of a corporation in like manner assume the risk that its assets may be so depleted as to be less valuable than its bond indebtedness. This sometimes occurs when the management of a corporation exploits the property and leaves the depleted assets to the bondholders.

Finally, the laborer assumes some very positive risks in the operation of business. While his wages are usually paid at regular intervals in advance of the realization of the returns from business, and while he is by law a preferred claimant on the assets in case of failure, nevertheless, he assumes some very important risks. Among these unemployment is perhaps the most significant. His contract with his employer is ordinarily indefinite and

can be terminated at the will of the employer. Since the wage bill is a large item in the expenses of most business concerns, it is customary in case of seasonal fluctuations or any unusual conditions of demand for the employer to let out his labor force as a means of reducing his money outlay. In only a few instances in this country, as in the case of the Cleveland Ladies' Garment Manufacturers, some of the men's clothing manufacturers of Chicago, and a few others, have the employers experimented with a definite guarantee of a certain number of weeks' work and made provision for spreading the risk of unemployment among the firms that are party to the agreement. Unemployment insurance has been applied in England, and advocated to some extent in this country as a means of relieving the employees from some of the risk from inability to find employment.

In addition to the risk of unemployment, the laborer assumes the risk of the loss of skill which is caused by the invention and use of machinery and the discovery of new methods of production. Whatever may be the ultimate effect of machinery upon the relative economic position of the workingman, the first effect is most likely to be a decline in the importance of his skill and, for a short period at least, a loss of employment for some of those affected by the new methods of operation. Besides these risks from unemployment, the laborer bears a very large part of the risk of physical injury arising from industry, although in recent years by means of workmen's compensation laws the economic losses due to industrial accidents have been shared to some extent by employers and consumers.

Laborers assume the risks of industry herein described but they do not assume any risk for the success of the undertaking, unless they are investors, or the business is run as a cooperative venture. In either of these exceptions the workmen assume the risk not as laborers, but as investors, or entrepreneurs. Thus, while the investors and laborers assume certain risks in the course of production, the success of business rests primarily with the entrepreneurs.

3. Methods of Shifting Risk.—The entrepreneur does not ordinarily assume risk as a matter of choice, and he will usually shift it as far as it is possible to do so. The tendency is to shift risks to specialists or experts in risk taking. There are a number of methods of transferring risk from one group to another, among which the following may be considered, namely, the transference

to middlemen, to experts in the terminal markets, and to insurance companies. We will now consider these methods in the order mentioned.

(1) *Middlemen as Risk Bearers.*—The most common method of shifting risk from the producer to the dealer, or middleman, is to produce only on orders booked in advance. By working only on orders a manufacturer can avoid the risk of unsold products, and can even protect himself against price changes during the time necessary to produce the commodity in question. For instance, a textile manufacturer may produce cotton cloth only on orders from merchants and maintain no stock in his warehouse. In this way he avoids the uncertainty involved in the disposal of a stock of manufactured cloth and relies upon the merchant for the continuity of his business. If a stock of such material is maintained, the merchant carries it and assumes the risk that may arise from any shift in demand, from the payment of the carrying charges, such as interest on the investment in the stock, insurance, and any other charges incident to the maintenance of the supply of goods.

The manufacturer may go a step further and protect himself against price fluctuations. He may agree with the merchant to furnish the cloth at a definite price fixed in advance of its production. By purchasing the cotton fiber at the time he receives the order from the merchant, he can include in this price all of his manufacturing costs, including those of the raw materials used in the production of the goods, and thus protect himself against any price changes in these materials. For a certainty in his costs he shuts himself off from any gains that might arise from possible lower prices of these materials at the time the cloth is woven, but avoids the possibility of higher prices. By such methods the manufacturer shifts the risks of maintaining stocks of raw materials and finished goods to others and contents himself solely with a manufacturing profit.

While large numbers of entrepreneurs are enabled by means of contracts or orders to transfer the risks involved in production in anticipation of demand, the industry as a whole cannot avoid risk in this manner. What happens is that dealers or middlemen are likely to become the risk bearers. They will give orders in advance of sales and hold the stocks until the consumers' needs have reached the market. Wholesalers and jobbers frequently assume the responsibility of estimating demand, and on

the basis of these estimates issue orders for goods to the manufacturers.

As consumers, we expect to find the goods available at the time that our desires become active, but in order that the goods may be obtainable at this time, the manufacturer must produce them in advance of demand, and either he or the jobber or middleman must hold the stocks. Thus, by working on contract or orders, the manufacturer foregoes the chance of gain from advance in prices in return for security against loss in case of a price decline. This does not mean that the risks have been eliminated, but that the chance of losses and the opportunity for gains are shifted to specialists or middlemen. One group, anticipating demand, holds the stock of finished goods, another, the stock of raw materials. Thus the uncertainty which develops as the result of specialization leads to a further extension of this principle by developing a group who specialize in anticipating demand and in bearing the risks of production.

(2) *The Terminal Market and Risk.*—A further concentration of risk bearing has developed in the central or terminal markets, such as the stock exchanges which deal in corporate securities, or boards of trade which deal in grains and produce. Here a specialized group of traders come together and perform two important functions. They constitute a market in which price is established, a topic to receive further consideration under the subject of value, and serve as an agency for carrying and distributing risk. It is in the latter capacity that terminal markets will be considered at this point. In order to understand the effect of terminal markets upon the distribution of the risks from production, it will be necessary to examine the nature of speculation and speculative trading.

Speculation consists in forecasting price movements and in buying or selling on the basis of these forecasts. The buying and selling of the speculative trader differs from the buying of the middlemen by virtue of the fact that the middlemen usually have in their possession the stock of goods which they purchase, while in speculative trading the speculator is making contracts for a supply of a commodity, the stock of which may never be in his actual possession. He is engaged in buying and selling rights to a supply of the commodity, either on his own initiative or on orders from other buyers. The speculator hopes to make a profit out of his sales on account of changes in prices from one time to

another. The ordinary middleman may enjoy some profit from speculative sources, but receives most of his profit from differences in prices in different markets. A commission merchant, or broker, may buy wheat from a dealer who has a supply of wheat to sell, and then sell it to a miller who expects to grind it into flour. He secures as profit the difference in price at the elevator and that offered by the miller. The speculative trader is engaged primarily with price movements at some future date. This leads to a consideration of future contracts, or the sale of *futures*.

a. Nature of Future Contracts.—The operation of the sale of futures may be illustrated in the wheat market. Wheat is a world commodity in the sense that it is grown in the northern and southern, and in the eastern and western hemispheres. The market price is therefore affected by the conditions of supply wherever wheat is grown. Hence, any terminal market dealing with wheat must have accurate information concerning the world supply and the per capita consumption. Since wheat is cultivated in both northern and southern hemispheres a crop is growing almost continuously, so that dealers must keep themselves informed concerning the condition of the growing crop or prospective supply, as well as with the existing supply. To obtain this information a board of trade must be in telegraphic communication with all of the leading wheat-growing countries and with all wheat markets, and at the same time it must keep accurate statistics of available supplies as well as estimates of the prospective supply. With this information at hand the speculative traders forecast price movements.

There are two groups of dealers operating on a board of trade, those who think future prices will be higher, hence, they buy at present prices contracts for future delivery and expect to make a profit by selling these contracts at a higher price at a later date. These speculators are commonly known as "the bulls." Another group of dealers proceed on the theory that future prices will be lower than present prices, so they sell contracts for future delivery, hoping to make a profit from their ability to buy at a lower price in the future a supply sufficient to fulfil their contracts. This group of dealers is known as "the bears." The practice of "the bears" in selling contracts for wheat which they do not have is known as "selling short." Should they make a mistake in their estimates, they will lose because, when they

come to fill their contracts, they will find that the price of wheat has not fallen, hence, they will be unable to purchase a supply at a lower price. If their estimates are correct, however, they will make a profit. In like manner "the bulls" may, or may not, make a profit, depending upon the correctness of their estimates of supply and future price movements. As profit can arise only from the accuracy of their estimates, these dealers become experts in forecasting supply and demand. When it is stated that there are two groups of traders, it must not be inferred that any particular trader is permanently a "bull" or a "bear," for, in fact, a trader may be a "bull" today and a "bear" tomorrow, depending upon which theory of price movements he is operating.

With this knowledge of a terminal market in mind, the nature of a "futures" contract may be studied. Let us suppose that A is a "bull" and is proceeding upon the theory that the price of wheat 3 months hence will be higher than it is today, and that B is a "bear" and thinks the price will be lower. A is willing to buy, we will say, 10,000 bushels for delivery 3 months hence, and B is willing to sell the same amount. Let us assume that a contract has been made between A and B involving 10,000 bushels of wheat. Theoretically, the supply of wheat for the present market has been decreased by 10,000 bushels and the supply available for the future market has been increased by the same amount. The effect upon present prices and the estimate of prices 3 months hence is the same whether B actually possesses the wheat, or depends upon his ability to purchase that amount at the future date. Contracts will continue to be made between the "bulls" and the "bears" until there has been an equalization of the estimates of the supply and demand through the 3 months' period of time. The net result of trading in *futures* in accordance with this assumption is to regularize the present and future price of wheat during the periods affected by the contracts, and thus to minimize the risks from price fluctuations as between sales at different periods of time. The variation in the price of wheat as between September and the following May, for instance, will not be as great as it otherwise would have been were it not for this practice, and the reduction of this variation in price tends to reduce the risk from this cause.

The effect of the sale of futures on the productive processes may be illustrated by the following example: Suppose a miller

secures an order for flour from a merchant, the flour to be delivered 3 months later. The miller is able to quote the merchant a definite price on the flour because he can get from a wheat broker a definite price on wheat for future delivery and to this cost he can add his own manufacturing expenses and quote a specific price at which he can furnish the flour. If the order for flour is given, the miller can then give the broker an order for an amount of wheat necessary to make the required quantity of flour, the wheat to be delivered in due time to be milled. The broker will go to a speculative trader and buy for future delivery the amount of wheat ordered by the miller, say 10,000 bushels. In this way, the miller has shifted the risk of price change in wheat to the dealers in the terminal market who make a business of dealing in speculative contracts. If it were not for this possibility, the miller would either have to provide himself with a supply of wheat sufficient for his yearly volume of business or do his own estimating as to the probable changes in prices at the time an order was under consideration. In either case, he would have to assume a large risk, and in most instances he would not have the information necessary to render an intelligent judgment upon its amount. The probability of error in estimates made in this way would be much greater than through the present practice of buying and selling *futures*.

The miller in the above example contents himself with a manufacturing profit. If it were not for the terminal market and the sale of *futures*, he would have to include in his costs the speculative elements due to the possible price changes. But by means of the present practice he is able to shift the risk of these changes in his raw materials on to those who make a specialty of buying and selling future contracts. Day by day, these speculative dealers are engaged in buying and selling contracts for wheat and, as the season proceeds and contracts from millers take the available wheat out of the speculative market and it is transformed into flour, these future contracts have to be settled. There results from these settlements an equating of the actual supply with the estimates of that supply. Thus, in this way the speculative traders in the terminal markets assume a large part of the risks which result from changes in price that occur during a period of time.

b. *Hedging*.—As a means of shifting risk there has developed a special form of speculative contract known as “hedging.”

Hedging consists in the making of two contracts at about the same time, but of an opposite nature. Let us take the case of the broker cited above, and assume that the miller placed an order for 10,000 bushels of wheat on September 10 for April 15 delivery. When the broker accepts this order, he has sold short 10,000 bushels, which he has obligated himself to deliver on the specified date. There are three courses of action open to the broker in the process of fulfilling his obligation to the miller. He may purchase the wheat in the cash market and make provision for storing and holding it until the delivery date; or he can wait until the delivery date and then buy the wheat in the cash market, thereby assuming the risk of price changes; or he can immediately purchase a futures contract for 10,000 bushels of wheat to be delivered at a date that will enable him to fulfil his contract with the miller. Since the futures contract is as effective in protecting him against unfavorable price changes as if he bought the wheat outright and arranged for carrying it until the date of delivery, he is likely to choose this method of shifting the risk, because of the inconvenience and cost in storing the wheat.

Now let us suppose that the quotations of wheat on September 10 were as follows:

Cash wheat.....	\$1.00
December wheat.....	1.02
March wheat.....	1.04
May wheat.....	1.05

In order to cover his risk it will be necessary for the broker to purchase a futures contract for May because the regular quotations are for March, May, July, September, and December. The purchase of this contract will protect him against any price changes that may occur between September 10 and the following May. When the time comes to secure control of the wheat to be delivered, the broker can go into the cash market and make his purchase at the prevailing price. The actual price at this date may be identical with the estimate made in September, or it may be either above, or below that estimate. If the cash price in April is identical with the September estimate, then the broker can purchase wheat for delivery at \$1.05 per bushel. He can also sell his futures contract at that same price, so that he neither gains nor loses on the two transactions. In this case, there is no special advantage in negotiating a hedging contract but no one

can foresee accurately the course of commodity prices. Because of this uncertainty in the movement of prices, the practice of selling futures has developed.

But if the wheat price in April should rise to \$1.10 per bushel, the advantage of the hedging contract will be apparent, for when the broker now comes to purchase the wheat for delivery, he will find the cash price higher than it was estimated in September. The difference of 5 cents per bushel represents his loss in case he failed to protect himself against a price change of this character. However, he can dispose of his futures contract at the higher price, so that what he loses on the one contract is offset by his gain on the other. In this instance, the hedge has protected the broker against the loss that he would have suffered from the increase in the cash price of wheat between September and April.

Should wheat prices fall to 90 cents per bushel, the two contracts will again neutralize each other. In this case, the broker could buy the wheat in April for 15 cents per bushel less than it was estimated at the date of his contract with the miller, which would have given him a profit of this amount, if he had not purchased the futures contract. In other words, if he had carried the risk himself, he would have gained the amount specified as a speculative profit. But when he comes to sell his futures contract for wheat, it will command but 90 cents per bushel, so that he will lose 15 cents per bushel on this contract. Hence, the broker loses on one contract the same amount that he gains on the other. He gives up the chance of receiving a speculative profit for the assurance that he will be protected against loss in the event that prices move upward. In this way, the hedging contract protects those who use it against the loss arising from unfavorable price changes and passes this risk on to the speculative traders.

The following tables will illustrate concretely how the hedging contracts mentioned above neutralize each other. Two things should be remembered in connection with these tables. First, they are oversimplified in order to bring out boldly the way in which hedging operates. Carrying charges are eliminated from the calculations and it is assumed that the price of futures moves in the same direction with cash prices. In practice, the hedging contract is never as perfect nor is the process as simple as these tables imply; hence, the risk is not ordinarily as perfectly offset as the tables show, but the tables do indicate how the hedging con-

tract enables those who make use of it to shift a large part of the risk from price changes. The second thing to be remembered is that cash prices may be identical with, above, or below their estimates at an earlier date. The following tables are designed to show the effect of a hedge in each of these three possible price conditions.¹

TABLE I.—FUTURE PRICE AND PRESENT ESTIMATE IDENTICAL

First contract	Second contract
Sold (short) Sept. 10, 10,000 bushels wheat for April delivery @ \$1.05..... \$10,500	Purchased Sept. 10, 10,000 bushels wheat in futures market for May delivery @ \$1.05..... \$10,500
Purchased April 6, 10,000 bushels in cash market @ \$1.05..... 10,500	Sold contract for 10,000 bushels April 6 at cash price \$1.05..... 10,500
Gain or loss..... \$00,000	Gain or loss..... \$00,000

In Table I, it is assumed that the April price is just equal to the preceding September estimate of that price. In the first contract the broker sold short to the miller the 10,000 bushels of wheat to be delivered in April at the price at which wheat futures was quoted on the date of the sale. In order to cover his obligation to the miller he immediately purchased in the futures market the same amount of wheat for May delivery. The broker has thus made two contracts of an opposite nature; he has a credit due him from the miller for \$10,500 and he owes the speculative trader the same sum. When it is necessary to secure control of the wheat for delivery to the miller, the broker goes into the cash market and purchases the 10,000 bushels and immediately sells his right to the 10,000 bushels which he obtained on the futures contract. Since the cash price coincides with the September estimate, there is neither a gain nor a loss on either contract.

In Table II, it is assumed that the April price advances above the September estimate. The process is the same as that just described, but the effects are somewhat different. In this case the broker loses \$500 when he purchases cash wheat to fulfil

¹ TAYLOR, F. M., "Principles," p. 235; Cf. also CLARK, F. E., "Principles of Marketing," pp. 364-365, for a discussion of this subject.

the order of the miller, but since he can sell his rights obtained through the futures contract at the higher price, he gains a similar sum, so that the loss is neutralized by the gain on the speculative contract.

TABLE II.—FUTURE PRICES HIGHER THAN PRESENT ESTIMATES

First contract	Second contract
Sold (short) Sept. 10, 10,000 bushels wheat for April delivery @ \$1.05..... \$10,500	Purchased Sept. 10, 10,000 bushels wheat in futures market for May delivery @ \$1.05..... \$10,500
Purchased April 6, 10,000 bushels in cash market @ \$1.10..... 11,000	Sold contract for 10,000 bushels April 6 at the cash price of \$1.10..... 11,000
Loss..... \$ 500	Gain..... \$ 500

TABLE III.—FUTURE PRICE LOWER THAN PRESENT ESTIMATE

First contract	Second contract
Sold (short) Sept. 10, 10,000 bushels wheat for April delivery @ \$1.05..... \$10,500	Purchased Sept. 10, 10,000 bushels wheat in futures market for May delivery @ \$1.05..... \$10,500
Purchased April 6, 10,000 bushels in cash market @ 90¢..... 9,000	Sold contract for 10,000 bushels April 6 at the cash price of 90¢..... 9,000
Gain..... \$ 1,500	Loss..... \$ 1,500

In Table III, it is assumed that the April price is below the September estimate. Under these circumstances the broker can buy the wheat in the cash market at 90 cents per bushel, thereby making a speculative gain of \$1,500, but since he will have to sell his rights at the same price, he will lose an identical sum on his futures contract. Here again the profit and loss just offset each other, but the broker has protected himself in this way against the risk of loss that might arise from changes in prices during the interval of time assumed. He foregoes the chance for a speculative gain for the assurance that he is protected against loss from an advance in the price of wheat.

These three illustrations show a perfect balance in the hedging contracts. However, an actual hedge is not quite so simple as these hypothetical instances would lead one to believe. The spread between the cash and future prices does not always remain perfectly uniform as is here assumed, so that the making of such a contract is not as simple as appears from these examples. Nevertheless, when negotiated with skill and judgment, a hedging contract does serve as a means of shifting the risk of price changes that may occur during a period of time from the maker of the contract to the operators in the speculative market.

It may be asked at this point, why anyone should wish to make such a contract, and the answer is that he is satisfied with a stipulated commission, or profit for his services, and desires to be free from the possibility of loss due to price changes, even though these changes might yield him a profit if they were favorable to him. By means of a hedging contract of the kind described, the dealer, whether he be broker or miller, or other purchaser of wheat, can protect himself against most of the loss and receive a stipulated commission or profit for his services.

The advantages of hedging are not confined to brokers, for manufacturers may take advantage of the practice to protect themselves against price changes during the period that the raw materials are being worked up. The miller mentioned above may buy wheat in September to manufacture into flour to be delivered in May. If, when he purchases his wheat in September, he can sell flour for May delivery, he protects his manufacturing profit against the possible loss from the fall of flour prices. Or, if he should contract in September to deliver flour at a stipulated price in May, he can purchase wheat for future delivery in time for milling and thus protect his manufacturing profit against a rise in the price of wheat. Similar instances of the operation of the hedging contract might be cited from other industries, but enough has been said here to show how, by means of such contracts, a very large part of the risk due to price changes during a period of time can be shifted to the dealers in the terminal and speculative markets.

The social effect of the practice of selling futures is worthy of further consideration. In addition to serving as a means of absorbing the risk due to price changes during the period required for the manufacture of flour, the speculative dealers have an important influence on the price of wheat during the period in

which a season's supply is marketed. Those who buy now for future delivery are taking supply out of the present market and, therefore, tend to increase the present price of wheat. But when they sell their holdings at a later date, it has the effect of increasing the supply at that date and thereby lowering the price that otherwise would have obtained. Thus the buying in September for May delivery will tend to raise September prices and lower May prices, or in other words, the sale of futures will tend to stabilize prices during the period in which a season's crop is marketed. This conclusion is based upon the assumption that in selling futures there is no attempt to manipulate the market. In so far as the sale of futures results in narrowing price fluctuations, it is a social benefit. With the greater stability of prices, uncertainty from this cause is reduced, and all producers, whether they be the farmers, or the manufacturers of goods from farm products, are benefited thereby.

c. *Manipulation.*—From the foregoing discussion it will appear that a considerable amount of the forecasting of price movements is concentrated in the terminal market, and is done by the specialized traders in futures. By means of the operation of these traders, the manufacturers, jobbers, and other middlemen can shift the risks of price changes to this specialized group. The social advantage of this method of handling risks is obvious, but like all good things, this form of speculation is subject to abuse. So long as the buying and selling by these traders is based upon accurate information as to supplies and per capita consumption of the commodity dealt in, these men are in a better position to forecast price movements than are ordinary dealers. But since the speculative phase of production is separated from the technical operation, these traders are not concerned so much with the social effects of their trading as they are with the private profit they are able to make from the buying and selling in the market. As a matter of fact, only a fraction of 1 per cent of the contracts made in the wheat market are completed by actual delivery of the grain, the balance being settled by transactions in the "pit."¹ Unfortunately, a trader may make a profit by methods which do not coincide with the best interests of society. There are three types of evils that have developed in connection with speculative dealings, namely, trading by inexperienced per-

¹ U. S. Dept. of Agriculture, *Technical Bull.* 79, p. 3.

sons or outsiders, as they are sometimes called; trading with insufficient capital; and manipulation of prices.¹ Each of these will receive separate treatment.

- (1) *Speculation by the Outsider.*—The terminal market lends itself to speculation by the uninformed outsider, who wishes to get rich quick, and takes a "flyer" in the speculative market. The operation by the outsider is likely to be without accurate knowledge of market conditions, hence the forecasts of price movements by such a trader are likely to be incorrect. He may not only lose personally but his operations may force prices from their normal trend and thus upset the estimates of the most skilful and well-informed traders.
- (2) *Trading with Insufficient Capital.*—The practice of buying on "margins" stimulates operations by uninformed traders, although the practice is not confined to them. The method is an encouragement to one form of gambling. The practice consists of advancing a sum of money on a contract to purchase, the broker making the sale, holding the stock or commodity purchased as security. In the event the prices fall, the broker calls for additional sums of money, or an increase of the margin, and should the purchasers be unable to make this advance, the broker promptly sells the holdings outright. Besides being a disturbing factor in the market, persons who take a "flyer" usually operate with insufficient capital. If their estimates prove to be correct they will make large profits, but if they be incorrect, they not only lose what they have advanced but they are likely to be forced into bankruptcy with insufficient assets to meet their obligations. Innocent traders who have made future contracts with a person operating in this fashion will suffer by such methods. This danger may be guarded against in the organized exchanges by requiring the traders to keep on deposit with the exchange a sum sufficient to cover any differences between the price at which they have bargained to deliver or to accept futures, and the current price of the futures in that market.

- (3) *Manipulation of Prices.*—Many traders operate on a very large scale and maintain their own statistical organizations.

They may cause false reports to be given out as to the

¹ CLARK, F. E., "Principles of Marketing," pp. 375-377.

conditions of a growing crop, or of the available supplies of, or demand for, a given commodity, thus causing artificial price fluctuations. Another method that may be used to manipulate prices is for the traders to operate in the opposite direction from the normal trend of prices based upon the available information. By sudden and extensive sales of futures, they will depress future prices below what the actual supply and demand would fix. Secretely, such traders will begin to buy after the price has been depressed, and in this way they may get control over large quantities of the good, and thus increase their profits. While such manipulations may result in a profit to one group of traders at the expense of another with little social consequence, yet, in other cases, because production is carried on in anticipation of demand, and price is taken as the indicator of that demand, manipulations that artificially affect price will result in a disadvantageous use of the factors of production. Land, labor, and capital may be diverted to less effective uses, and the supply of wealth available for the gratification of our desires will not be well balanced. Since business relies upon price changes as the indicator of supply and demand, any falsification of that indicator violates the interest of society and is, therefore, economically wrong.

(3) *Insurance*.—The last method of distributing risk to be considered is that of "insurance." The nature of insurance is a pooling or consolidation of a large number of instances in which risk appears. As was pointed out above, we frequently know more about groups of facts than we know about individual instances making up these groups. This is particularly true of many kinds of uncertainties found in business. Every business man runs the risk that his mill, store, or other type of business may be destroyed by fire. If he makes no provision in advance for this contingency, the loss, if it occurs, may be so serious as to ruin his whole business. Actuaries, however, can with great accuracy figure out from past experience what the probability of such a loss is and from this estimate they can calculate what sum must be set aside yearly in order to make provision for this contingency.

An individual might follow such an experience table and accumulate a reserve by setting aside a sum sufficient to cover the

loss when it occurs. In fact, many large corporations do provide self-insurance by setting aside yearly a definite sum as a reserve to provide protection against the contingencies which are thought to be hazards to the business. But ordinarily, the carrying of risks of this character is turned over to corporations that specialize in the business of insuring individuals or business firms against various kinds of uncertainty that occur in industry.

The economic service performed by means of insurance can be illustrated by showing how the risk from fire can be distributed in the case of residences in a village or a city. Suppose there are 1,000 houses involved, and that each house is worth \$10,000. It is much more likely that some one house will burn than that all of them will be destroyed by fire during any given period of time. Hence, there is greater certainty that there will be a fire within the group than there is as to which house will be burned. By pooling the risk, the loss can be distributed and each house owner will lose \$10 in case of each fire, while he receives the assurance that he will be reimbursed in full in case his own house should burn. The individual house owner can well afford to pay \$10 per fire so long as the number of fires remains relatively small. By this means, each house owner changes the uncertainty of a large loss into a series of small losses. The advantage to the individual of this procedure is apparent unless the number of fires causes the number of payments to approximate the value of his own house—a situation that is quite unlikely except in the case of a general conflagration. The probability of a fire that will destroy the entire village is remote, although great areas in cities have been burned, as in the case of the Chicago fire, the Baltimore fire, or other similar instances of wholesale destruction. It is the practice of insurance companies to reinsure with other companies in case they have assumed a highly concentrated risk. In this way the risk is spread over a wider area, and the likelihood of a general destruction covering this wider area is extremely remote.

By means of consolidating the risks, the individual is freed from uncertainty. By the payment of a definite and certain amount, known as the premium, he shifts the uncertainty of a larger loss to the company that accepts the risk. It should be noted that, from the point of view of the company that carries the risk, a very large part of the uncertainty disappears by means of consolidation. In so far as the actuaries can build up experience

tables, they can figure the exact payment or premium that will be required to cover the risk. This premium, plus the interest earned on the investment of any reserves held by the company, must be adequate to meet all obligations assumed. Unless the premium is figured on an actuarial basis and is sufficient to meet all of the contingencies during the whole period that the risk is to be covered, then those who insure are not adequately protected against the uncertainty involved. The premium is made up of two parts, the real cost of the insurance, and a loading to carry the cost of operating the business. When the premium has been built up in this way, the insurance company assumes no risk, but is performing a service for the individual by carrying what would be an uncertainty and hence a risk for him. From the certain payments made to the company, there is paid to those who suffer loss the amount of insurance which they have purchased. The company has collected these small amounts, invested any reserves that it has accumulated, and pays back to the individual a sum to cover the loss that he has sustained.

The kind of risks that are insurable are ordinarily those where the uncertainty can be calculated, or reduced to an actuarial basis. Whenever an experience table can be built up and the losses estimated with reasonable accuracy, then it is possible to carry the risk in the manner described. Many insurance companies will assume risks, even when the uncertainty is not calculable. In these instances the company makes a guess as to the probable loss and places the figure high, so that the hope of unusual returns stimulates the managers to assume the unusual uncertainty involved. When workmen's compensation laws were first enacted in this country, the insurance companies had little basis of experience on which to fix the premium rates. In Illinois the rates at which the private insurance companies would write compensation insurance in the coal industry were so high that a large number of the mine operators found relief in a mutual insurance company which enabled the members to pool their risks and to carry their liability at a much lower rate than the private companies were willing to give them. Whenever a company assumes a risk in the absence of information on which to build up the rate or cost involved, it assumes the uncertainty of loss. In this case, the insurance company takes a gambler's chance and risk taking under these circumstances is

akin to gambling. In practice, the risks assumed by insurance companies vary all the way from those which can be figured with great accuracy, to those in which there is no experience on which to make an estimate of the uncertainty involved.

Mutual versus Stock Companies.—It should be noted that all insurance is essentially mutual, that is, it represents a group of individuals making regular contributions for the purpose of restoring to any member of the group who suffers loss, the amount of that loss. While this is the essential nature of all insurance, there is a difference in the form of organization of the companies that specialize in insurance. When a group of persons incorporate primarily for the purpose of carrying their individual risks, such a company is called a mutual. The policy holders receive their insurance at the net cost of carrying the risk and there is no attempt to make any profit from the business of insurance. When the premiums exceed the losses involved, or the earnings on invested reserve are greater than the estimates, the policy holders will receive the benefit by a dividend or by a readjustment of the premium.

A stock company, on the other hand, is one which is organized and run as any private business. The risks are accepted at a figure that will cover the losses and in addition provide a sufficient margin for a dividend upon the capital stock of the company. A stock company assumes risks for the purpose of making money for its stockholders, while a mutual company is primarily interested in distributing the risk for the policy holders. Without attempting to go into detail to show the distinction between the different forms of organization which insurance companies take, the above description will enable the student to distinguish the two most important types. Other forms are only a variation of one of these.

From the foregoing discussion it should be apparent that insurance is one method by which the business man, whether he be an individual proprietor or an officer of a corporation, may shift some of the risks. With the development of large-scale production and specialized industry, the insurance business has grown to enormous proportions. Most of the important contingencies that confront the business executive can be converted by means of insurance from an uncertain to a certain amount. The enormous sums of insurance written yearly are evidence of the extent to which business men use this method of

transferring their risks. It is not possible, however, to eliminate or transfer all risk; some of it must be borne by those engaged in business. The methods by which industrial risks are distributed among the various classes of business men have been described and some of the most important ways of eliminating or reducing them have been set forth in the previous discussion. This treatment should be regarded as incidental to the general discussion of production and the student interested in this topic of risk and risk-bearing methods should turn to the many treatises on the subject.

Suggested References

- BLACK, JOHN D., "Production Economics," Chap. XXIII.
BYE, RAYMOND T., "Principles of Economics," Chap. X.
CLARK, FRED E., "Principles of Marketing," Chap. XVII.
CLAY, HENRY, "Economics for the General Reader," Chap. IV.
FAIRCHILD, FURNESS, and BUCK, "Elementary Economics," Vol. II, Chap. XXXII.
HARDY, CHARLES O., "Risk and Risk-Bearing," Chaps. I, VI, and XII.
HAYES, H. GORDON, "Our Economic System," Vol. I, Chap. XVII.
HENDERSON, HUBERT D., "Supply and Demand," Chap. VII.
HUEBNER, S. S., "The Stock Market," Chaps. I-VII.
KNIGHT, FRANK H., "Risk, Uncertainty and Profit," Chaps. I, VII, and VIII.
MARSHALL, LEON C., "Industrial Society," pp. 470-513; 516-518.
RIEDEL and LOMAN, "Insurance Principles and Practices," 2nd Ed., Chaps. I-III.
SMITH, J. G., "Organized Produce Markets."
TAUSSIG, FRANK W., "Principles of Economics," 3d Ed., Vol. I, Chap. II.
TAYLOR, FREDERICK M., "Principles of Economics," Chap. XVIII.

CHAPTER VIII

VALUE AND UTILITY

Up to this point in our discussion, we have been concerned mainly with a description of the general structure of our economic organization and with a study of the relations existing among the various phases of our economic life. The examination of this structure begins at the point of ungratified desires—a fact of human experience which constitutes the main stimulus to industry. While some industry may arise in response to the desire for activity, and while the “play-spirit,” or “love of the game,” may be present to some extent in all industry, we must look beyond this motive to explain why men will continue to work often to the point of great fatigue. The explanation of this fact is found in the desire for goods, which is the active stimulus that accounts for the existence of the whole industrial mechanism of our time. Production in all its varied forms arises in response to this stimulus. In the study of production we were led to a consideration of the division of labor or specialization in its three principal forms, occupational division of labor, division of processes, and territorial division of labor. In addition to these external phases, further study of the subject led to a consideration of the factors of production, namely, land, labor, and capital, and of the problems associated with the most effective combination of these factors for productive purposes. The responsibility for organizing and directing industry rests in our economic system with the entrepreneur, or captain of industry. It is his responsibility to determine the policy of the business establishment with which he is identified. His decisions have a vital bearing upon the economic welfare of mankind, because the realization of the ultimate objective of all industry, *viz.*, the production of a large and continuous flow of wealth available for the gratification of human desires, depends upon the decisions which he makes as to the use of the factors of production.

The Market.—Our next step is to inquire into the forces that integrate the separate parts of this structure and bring system

out of the apparent diversity. When viewed externally and from the standpoint of the individual, modern industry appears to be divided into a multitude of separate and independent units, but closer examination reveals a large degree of unity in the productive processes, especially when they are considered from the standpoint of society. The integration of all the separate producing units into an industrial mechanism is accomplished through the market.

The market is the coordinating agency in modern industry and occupies a strategic position among the social institutions of the present age. This position has been attained primarily because the vast majority of goods today are made to sell. As we have seen, this situation has not always obtained. Before division of labor and specialization had been developed, production was mainly for personal use or, at most, for a very restricted market. But improved methods of transportation and communication, the development of stable governments, the extension of specialization, and the use of machinery have changed the simple conditions of early times into the complex economic relations that characterize the present day, and we may repeat, that it is the market that brings system out of the apparent diversity of economic activity. Because men now direct their energies to the production of one good, or at most a few goods, it becomes necessary for them to find a way to exchange their own products for the products of others. Markets have evolved as the organized means by which such exchanges take place. Today we hear constantly of commodity markets, produce markets, money markets, labor markets, real-estate markets, in fact, a market for every commodity or service that is regularly exchanged. These facts are conclusive evidence of the significance of trading and exchange in the conduct of modern industry.

Middlemen and Market Operations.—It is at this point that middlemen, such as wholesalers, retailers, jobbers, brokers, commission men, etc., come into the operation of our economic life. The function of these men is to bridge the gap between the producer and the user of a good. The method of trading varies from industry to industry and, as a result, there has been built up a highly organized structure which is calling forth a large and rapidly growing volume of literature describing its operations. The economic importance of the service rendered by the middleman is obvious, but the many specialized questions of marketing

and marketing organization must be left for the specialists in this field. To undertake here an elaborate treatment of this subject would divert attention from the main purpose of this text.

A Market Defined.—A market may be defined as the area within which the influence of buyers and sellers operates to effect exchanges. This definition implies that there are two or more persons engaged in exchanging two or more commodities. Usually the buyers and sellers are in close proximity, although physical proximity is not a necessary condition for the existence of a market. The telephone, the telegraph, and the mails are effective means of bringing about exchanges of goods, even though the buyers and sellers are separated by thousands of miles. From what has already been said, it should be apparent that markets are a necessary consequence of specialization in production and have developed because of the division of labor that characterizes modern industry.

As exchange is now carried on it usually involves the trading of goods for money, or money for goods. However, trading began in the form of barter, *i.e.*, the exchanging of goods for goods. The inconvenience experienced in this form of trading greatly limited its use, and very early in the history of the human family some commodity came to be used as a means of facilitating exchange. This commodity became known as money, a topic to receive treatment later, and by means of its introduction the trading of goods was greatly expanded. Markets would still be very limited in scope if it were not for the use of money and especially money in the form of the precious metals.

It is sometimes thought that a market is a place or a location for exchanging goods, as a board of trade, or a produce exchange. Actual trading does require space and a location in which to conduct business, but place is not the central idea of a market, for a market is really the whole area within which the influence of buyers and sellers operates. This may be a very restricted area, as when an individual bargains with a farmer for some commodity such as tomatoes, or it may be national or even international in scope, as in the case of a commodity such as wheat. The trading in wheat on the Chicago Board of Trade is affected by conditions all over the world. There are telegraphic communications with similar trading centers both in this country and abroad, and the conditions prevailing in the whole wheat-growing

area are immediately reflected in the operations in Chicago. Reports of unusual demand in England, or a crop failure in South America are taken into consideration by both the buyers and the sellers in Chicago and have an influence in determining the basis of the exchanges made there. It is essential, therefore, to remember that a market is the whole area through which the forces that affect the exchanges operate.

There are a number of forces that limit the area of a market, among which the following may be mentioned. High transportation charges place a very definite limit on the area through which buyers and sellers can operate. Before the coming of the railroads the cost of transportation was so great as to confine traffic to very limited areas along the Atlantic coast. Perishable commodities have a more limited market area than durable goods, but improved transportation and modern methods of refrigeration have greatly increased the market area of many commodities like fruits and vegetables. When goods must be inspected before an exchange can take place, the area of the market is more limited than when the goods are exchanged on the basis of well-recognized standard grades.¹ These and other forces tend to limit the market area, although in our day the tendency is in the direction of ever-expanding markets, many of which are national and some even international in extent.

There are two aspects of market transactions that should be kept clearly in mind, *viz.*, the transfer of title from seller to buyer, and the transfer of the goods themselves. In the early markets, the physical goods were present at the time of the exchange and, hence, tended to concentrate in the market and then to be dispersed to the final consumer, a condition that is very frequently found today. In many of the highly specialized and organized markets of today, the concrete goods are not present but may be stored in warehouses many miles away from the point of exchange. These goods may be sold by samples or, where grades or specifications have been standardized, by description alone. When a market has developed to this extent, there are but two functions performed by the trading, namely, the transfer of title and the determination of the price or rate at which the exchange is to be effected. However, the change of title still controls the direction and use of the goods purchased, and the storing in warehouses, elevators, or other convenient

¹ CLARK, F. E., "Principles of Marketing," pp. 442-443.

points is only an economical method of handling the goods and of eliminating unnecessary waste in the passage from the stage of raw material to the finished product. While there are many other important market functions when the trading processes are considered from the point of view of the private business establishment, yet the transfer of title and the determination of the rate of exchange are the two which are of major significance in the development of the fundamental principles of economics. We will now direct attention to the determination of the rate of exchange, a topic that will require major interest during the remainder of the present and through the next three chapters.

A Competitive Market.—For the purpose of the development of economic principles, markets may be classified as either *competitive* or *monopolistic*. A competitive market is one in which there is rivalry between both the buyers and sellers. In a freely competitive market, or what has sometimes been called a perfect market, both the buyers and the sellers have full knowledge of all the conditions that affect exchanges. They know how large a volume of supply will be offered for sale, how many buyers will be in the market, and how much each will purchase. Each seller will strive to sell each buyer while each buyer stands ready to approach each seller with an offer to purchase. A stock exchange or a terminal market, such as a board of trade, is the closest approximation to a perfect market that can be found. Such a market presupposes that individuals are actuated solely by economic motives, that they are striving to secure the maximum benefits from the exchanges, that they always buy in the cheapest and sell in the dearest market, and that they are free from caprice and prejudice and unswayed by sympathy or charity. Under these conditions, competition may be said to work perfectly.

Perfect competition does not imply unlimited competition. It does not mean that the seller will lower price below cost in order to run a rival out of business, but it does mean that he will continue to lower price under the stress of competition until he is able to get only the cost of producing the good and at this point he will drop out of the market. Likewise, the buyer will continue to buy until the price he is willing to pay is an accurate measure of the satisfaction that he expects to get from an additional unit of the good. At this point he also will drop out of the market. Put in another way, perfect competition

presupposes that bidding for goods will continue until a perfect balance is established between the amounts that the sellers are willing to offer and the amounts that the buyers are willing to purchase at specific prices.

A Monopolistic Market.—The second form of market may be called monopolistic. While competitive forces are the dominating influences in most markets, there is a greater or less degree of conscious control exercised over the trading of many goods. Wherever conscious control is thus exercised, whether for private or public benefit, the market may be said to be monopolistic. We may, therefore, define a monopolistic market as one in which there is some degree of conscious control exercised over the conditions of exchange. Usually this control is exercised over the supply of goods offered for trade. For instance, if a merchant in a village of 500 people held all the available sugar, and it was impossible for the villagers to secure a supply from any other source for a period of 3 months, we could say that during this period there was a monopoly on the side of the merchant, even though there was competition among his customers. In this case the exclusive control over the supply of sugar gives to this market its monopolistic character. The competition among the villagers to buy sugar would operate precisely as in a competitive market, but the control over the supply would give to the merchant the power to manipulate this supply in such a fashion as to force the purchasers to pay the highest price they were willing to offer in order to enjoy the benefits from the use of the good.

This simple example illustrates the presence of conscious control over supply, which, as has been indicated, may be exercised by private individuals or by public authority. In most civilized countries the postal system is a government monopoly and may be cited as an illustration of conscious control exercised by public or governmental agencies. The example cited will also illustrate the fact that even where conscious control is an important market influence, there is likely to be present competitive influences, as in the case of the competition among the villagers for the use of the available supply of sugar. Complete control over supply is infrequent, but a greater or less degree of conscious control is present in many markets and, so far as it is present, the market may be said to be monopolistic. Although much less common, similar control over demand would also make a market monopolistic.

Competition *versus* Monopoly in Economic Analysis.—The main body of economic principles found in modern treatises has been developed upon the assumption of a competitive market. In practice, however, there is no market which conforms in every detail with the competitive ideal assumed above. Men are never guided solely by economic motives. Their knowledge and judgments are never perfect, nor are their actions always rational and consistent. In addition, the factors of production are never perfectly mobile, but tend to become specialized and adapted to specific purposes. These are some of the obstacles to the operation of competition and their influence has sometimes been called "economic friction."

Recognizing, therefore, that actual conditions are extremely complicated and that no simple rule will fit perfectly every situation that may arise in our day-to-day experiences, two alternative methods of dealing with economic phenomena present themselves to every thoughtful student who wishes to go beyond the mere description of economic facts and processes and to attempt to find some explanation of the relations existing among these phenomena. In the first place, he may assume, as has been done in much of the literature of the past, perfect competition and perfect mobility of the factors of production and then work out standards of economic behavior which he formulates into principles. In applying these principles, it will be necessary to make qualifications which will take account of the other market forces that constitute the obstacles to the operation of perfect competition.

Or, in the second place, he may proceed upon the assumption that economic forces are mainly controlled forces and formulate his explanations and principles accordingly. Because of the significance of conscious control in the marketing of so many commodities, some writers are inclined to abandon the older method of explaining price phenomena through the device of a competitive market and are favoring the development of principles upon the assumption of controlled or monopolistic forces. This method of treatment is possible, and economic principles could be formulated upon the assumption of a purely monopolistic basis. However, if the principles developed in this way were applied to the actual conditions of economic life, qualifications would have to be made to take account of the presence of competition, for every one admits that some competi-

tion is present even where monopoly is most actively at work. It is maintained here that if the analysis is carried forward with equal skill from each of these two points of approach, the net result in the form of principles will not be very far apart, except perhaps for minor differences in the phrasing that take account of the differences in the assumptions underlying the two methods of treatment.

Two advantages lie with the older method of explanation. First, there will be no break with the historical development of the subject, for the literature of economics has been formulated almost wholly upon the assumption of a competitive market. Second, the adjustments to take account of monopolistic influences will be much less difficult to make than those that will be necessary if the analysis proceeds upon the assumption of a controlled market, because monopoly relies very largely upon the presence of competition to determine market price. Conscious control, as indicated above, is usually exercised over supply, but when this supply has been placed upon the market, reliance is put upon competition among the buyers as a means of maintaining the price. The monopolist is not independent of the buyers in the determination of the price at which his goods will be sold.¹ Because of the considerations here presented, it has seemed advisable to the author to adhere to the traditional method of explaining price upon the assumption of a competitive market and to take due account of control in the determination of prices whenever attempts are made to apply general principles to actual problems.

The assumption of a perfect and freely competitive market as a means of explaining exchange relations is merely a logical device to deal with an extremely complicated group of social data. It is a perfectly proper method of analysis to assume that competition is working perfectly and that all other forces that may affect market conditions remain constant. By so doing, we are in a better position to evaluate the significance of competition as a market influence. Subsequently, we may subject each of the other market forces to the same kind of treatment, and in this way be able to formulate a judgment as to the significance of each of them, or, expressed in another way, we will thus gain

¹ Further treatment of this subject is postponed until the chapter on monopoly.

knowledge of the unimpeded operation of market forces as a basis for the development of fundamental principles.

Students may at first find difficulty with this method of analysis because the assumptions may appear unreal to them. They are much more likely to be led into confusion, however, if they undertake to deal with the complexities of actual life before they have developed some standards of evaluation. Without a knowledge of principles, the human mind soon becomes bewildered by the complexities of actual conditions—a situation which frequently occurs in popular thinking concerning current economic problems. It is of far greater importance that the student should master a few fundamental principles than that he should become lost in his knowledge of actual conditions, or overwhelmed by the obstacles and exceptions to the perfect operation of these principles. Once the principles have been grasped, they become the constants which help to explain economic phenomena. Principles, when derived in this way, will serve as the basis from which adjustments may be made in order to apply them to the actual problems of life.

No apology needs to be offered, therefore, for the development of economic principles from the assumption of a perfect market, for this method does not differ essentially from that followed in those sciences where experimentation is conducted under known or controlled conditions. The assumption of a perfect market is an attempt to follow the same method in the study of economic phenomena by isolating, even hypothetically, the various forces that affect them and by observing their operations unaffected by any disturbing influences.

Value Defined.—With these general statements in mind concerning the method of analysis to be followed, we may now proceed to a consideration of the specific problems growing out of exchange. Before goods can be exchanged some basis must be found for comparing them. How can a grocer and a shoe dealer trade their wares without each having some knowledge concerning the goods to be traded and some basis for measuring a unit of one in terms of a unit of the other? Of course it is possible to measure goods by some concrete unit as the pound, or yard, or bushel, etc., but such units do not express the social significance of the goods measured. A yard of silk may have much greater social significance than a bushel of oats, but, before trading can take place, a method of comparing the relative

significance of the goods to be exchanged must be found. The consideration of this problem brings us to the subject of value which is recognized in all modern treatises as the basis of measuring the social significance of goods and of establishing a method of exchanging them. As used in economics, value means the power which one good has to command other goods in exchange, or, briefly stated, value is power in exchange.

Value and Evaluation Distinguished.—From this definition it will appear that value is a market phenomenon which arises only when goods are being exchanged. In this sense, then, value must be regarded as an objective fact. However, value should be distinguished from the process of evaluation which is a subjective act. Prior to an actual exchange, the prospective purchaser is confronted with a number of courses of conduct. He may be presented with an opportunity to buy real estate and the chance to establish a home; he may have a security, such as a stock or a bond of some industrial enterprise offered him as a means of making provision for the future needs of himself and his family; he may be sought by auto salesmen, each expounding the relative merits of the car that he is selling, and with a gentle suggestion that man lives but once and why not enjoy the comforts of life as he goes along? Or, it may be, he is contemplating a trip abroad, or the purchase of books, fine pictures, statues, or some other method of contributing to his intellectual satisfaction and growth. These are a few of the great variety of alternatives that are constantly besieging the individual and calling for a decision on his part. But before a trade can be effected there must be some weighing of these various alternatives.

The comparison of these alternatives is a subjective process and, so far as conscious choice is exercised, the individual relates these opportunities to the urgency of his desires and selects those which he believes will yield him at that time the largest degree of satisfaction. This process of weighing the significance of goods as want-gratifiers may be called "evaluation," and consists of the estimate of the satisfaction which the individual expects to get from the consumption or use of a good.

Evaluation, as here explained, would take place in the mind of an isolated individual. He would have occasion to determine the relative significance of goods in relation to the intensity of his desires, not for the purpose of exchange, as there could be no trading under the conditions assumed, but for the purpose

of disposing of his productive energies. He would tend to produce goods that would gratify his desires in the order of their intensity. For instance, if his desire for vegetables were more intense than for meat, he would devote more time and energy to the growing of vegetables and less to the procuring of meat. In this way evaluation would take place in the mind of an isolated individual and would serve the functional purpose of guiding the distribution of his productive efforts.

In modern economic society, where production is primarily for exchange rather than for direct use, there is need for a basis of comparing the relative significance of goods to be traded. Evaluation furnishes this basis. The mental process for purposes of exchange does not differ essentially from that which goes on in the mind of an isolated individual, even though the objective results may differ somewhat. Market evaluation serves two purposes. First, it establishes a rating of goods in accordance with their capacity to gratify human desires, thus creating a basis for exchanging them. It is apparent that this rating process takes place antecedent to exchange and is the result of the estimates of the whole group of buyers and sellers who come into contact with each other for purposes of exchange. Just as the estimate of an individual for any particular good may fluctuate from time to time, depending upon the intensity of his desire for it, so the market estimate may vary with the changes in the urgency of the desires of the whole group of buyers and sellers.

For instance, if the supply of wheat grown in any year should exceed the normal yield by 25 per cent, it is safe to assume that the desire for a definite quantity of wheat would be less keen than if the yearly production had just been normal. Or, if we suppose that the yearly production is normal and then assume some unusual influence affecting its use, as the outbreak of a war, the desire for wheat and its products would likely be intensified. In either case, these fluctuations in the significance of the good as a want-gratifier will be registered in the market. The rating, placed upon the goods to be exchanged by the group of buyers and sellers involved, is the result of the subjective process of evaluation. By means of this rating there is established a basis of exchange, out of which there emerges power in exchange, or value. We may say, then, that value results from the evaluation process that takes place in the minds of those wishing to make

exchanges. If traders attach a higher capacity to gratify human desires to a unit of good *A* than they do to a unit of good *B*, then a unit of good *A* will have greater value than a unit of good *B*. Therefore, value, or power in exchange, may be regarded as an objective evidence of the subjective process of evaluation. The only objective test that we have of the significance which an individual attaches to a good is the amount that he is willing to offer of goods or services in exchange for the good in question.¹

In the second place, market evaluation from the standpoint of society serves the same purpose that evaluations do for an isolated individual, *viz.*, as a guide for the disposition of the productive energies of the whole social group. If, as the result of market evaluation, a high estimate is placed upon any good, that fact will be registered in the market by a high relative exchange power, which indicates that the good in question is gratifying keen desires. Entrepreneurs will recognize this fact and will tend to devote more time and energy to the production of this good than to the production of others. In this way, market evaluation serves a functional purpose in guiding entrepreneurs in the most effective use of the factors of production. The effect of evaluation is registered in the market through price. A high market price for any commodity indicates a relative scarcity of that good and puts a premium upon the production of additional quantities of it. Thus, price in a competitive market becomes an index by which entrepreneurs are guided in producing those goods which are highly esteemed by the consumers.²

Value and Utility.—We have already seen that utility means the capacity of a good to gratify a desire. Before a good can have value it must have the capacity to gratify some known human desire, even if this desire be only a passing whim. It is not necessary that commodities be "good" for people in the sense of improving their moral or physical natures in order that they may have value. It is only necessary that people desire them and are willing to give something valuable to obtain them. No one would be willing to trade a unit of one good for some other good unless he expected the second one would yield him some satis-

¹ MARSHALL, ALFRED, "Principles," 4th Ed., p. 76.

² This statement must be qualified by the recognition of changes in price due to changes in the value of money, a topic to be treated later.

faction or benefit. It does not follow from this that because a good has utility it, therefore, has value. A free good may possess utility in a high degree, as in the case of air. We could not live without an adequate supply of air, but under normal circumstances no one would be willing to trade anything for definite units of it. It has no value, therefore, in an economic sense. From this we conclude that a second quality must be present before a good has value, namely, *scarcity*, i.e., the quantity of the good must be insufficient to gratify all known human desires for it. If, for instance, air becomes scarce, men will pay for it as in the case of the cost involved in ventilating a building. The air itself is still free but, in order to get an ample supply to all parts of the building, machinery must be installed to pump it where it is needed and to this extent we may say that fresh air is scarce and involves a cost to secure it. The two conditions necessary for the existence of value are, therefore, utility and scarcity.

Value and Price Distinguished.—Value can be expressed only as a rate of exchange. We can compare two goods, as wheat and corn, and find that one bushel of wheat will exchange for two bushels of corn. It is incorrect to say that the value of the one bushel of wheat is the two bushels of corn, for this is a comparison of unlike qualities, viz., power in exchange with either the weight or volume, of the bushels of corn.¹ The power in exchange of wheat is not a quantity of corn, even though a quantity of corn may be used as a method of expressing the value of the wheat. The correct expression is, "The value of one bushel of wheat is equal to the value of two bushels of corn." Here we are comparing like qualities of wheat and corn and, by such a statement, we have expressed a rate of exchange. So long as the conditions affecting the evaluation of the two goods remain unchanged, wheat and corn can be exchanged in the ratio given. However, the rate of exchange should not be confused with the power of exchange. In all civilized countries it has become customary to express the rate of exchange in terms of some one or more commonly accepted goods which are known as money. The expression of the value of a commodity in terms of money is *price*. If we expressed the value of wheat and

¹ We may think of a bushel of corn as a definite number of pounds, or as a volume containing a definite number of cubic inches.

corn as given above in terms of money, we might say that the value of wheat is \$1 per bushel, and that of corn is \$0.50 per bushel. The rate of exchange which gives the basis of comparing the value of the two commodities is maintained in this expression and is still two bushels of corn to one bushel of wheat. The distinction between value and price, then, is that value refers to the power in exchange which a good possesses, while price is the quotation of that power in terms of money.

Because most exchanges are effected by means of money, *i.e.*, we sell goods and services for money and buy goods and services with money, should not obscure the fact that in reality goods are being compared with goods and are being exchanged for goods through the instrumentality of money. In fact, all trade is fundamentally of the nature of barter, although it is customary to use this term only when goods are actually traded for goods, as wheat for corn. Whenever the farmer sells his wheat for money and pays money for clothing, he is, in fact, comparing the value of wheat with that of clothing through the instrumentality of money. This type of transaction is representative of all of our economic activities, and in all economic trading value and price are both present.

In order to emphasize the strict sense in which value is used in economics, consideration is here given to two erroneous uses of the term. Frequently, we hear persons refer to the *intrinsic value* of some good. This expression usually implies that the good in question possesses some inherent property that gives to it an absolute, or at least a stable, value. This expression is inaccurate and will lead to confusion. Every good possesses some inherent properties which enable it to gratify a desire, and the possession of these properties will have an influence on its value. However, there is no difference in goods in this respect. The inherent properties of a good remain unchanged, but their value may fluctuate from time to time and from market to market, depending upon the evaluation placed upon them in a given market at a given time. Every one knows that the value of a bushel of wheat changes from time to time and to some extent from market to market, without any change in the inherent properties of the wheat itself. This change of value simply registers the varying intensity of the desire for wheat. The student, therefore, should avoid the use of the term *intrinsic value* in the development of his concept and remember that

value always expresses power in exchange, and is always a relative thing.

A second consideration that flows from the nature of value is that there can be no general rise nor general fall of values. It is unthinkable that the value of all commodities could rise or fall at the same time. By way of illustration, we may take wheat and corn again. Can the exchange power of both wheat and corn in relation to each other rise at the same time? Clearly not, for, if the value of wheat rises, one bushel of wheat will command more than two bushels of corn, and this will mean a fall in the value of corn in terms of wheat. What is true in this simple illustration is likewise true if we should multiply the number of commodities to be exchanged in keeping with actual market conditions. Therefore, a general rise of values is an impossibility. However, we may have the value of all commodities moving in a common direction in relation to the value of one commodity. The exchange power of all commodities might increase in relation to corn, which would mean that the value of corn had fallen in relation to all other commodities. This situation occurs quite frequently in comparing the value of all other goods with the value of money. In other words, we may have a general rise or a general fall of prices, which, we will find later, is due to a change in the value of money.

Before proceeding further, emphasis must be placed once more on the fact that the value of commodities is determined primarily as the result of market influences. Although the price of goods might be fixed by public authority, as was the case of some commodities during the World War, or custom might have such binding influence on prices as to permit little or no change, as is the case today in some parts of the Orient, yet, in the vast majority of cases the value of goods is determined by conditions prevailing in markets in which the goods are exchanged. Whether these conditions be monopolistic or competitive, value is always a market phenomenon and serves as a means of measuring the significance of wealth. We have seen that it is possible to measure wealth by some concrete standards such as yards, bushels, tons, etc., when we wish to know volume or quantity; but the only way of measuring its social significance is to express its importance in terms of value. By means of value, therefore, we are able to express the social importance of wealth and

through this expression to compare the relative significance of different kinds of wealth.

Demand and Supply as Value Determinants.—With these general considerations concerning the nature of value in mind, we are now in a position to inquire into the cause or causes of value, or, in other words, why commodities have power in exchange and what force or forces determine the amount of that power. If we were to ask the first business man whom we happened to meet, what determines the value of the commodities in his business, he would most likely reply demand and supply. This is the most commonly accepted explanation of value among business men and by the general public. In the discussion of current economic issues, newspaper and magazine writers frequently make reference to the natural law of demand and supply and clearly imply that these are self-operating, natural forces like gravitation. Since this explanation of value is so generally accepted, it will serve as a satisfactory departure for our analysis.

The Nature of Demand.—In proceeding with this analysis, we will first inquire into the nature of demand as a market influence. At the outset of our inquiry we must distinguish *demand* from *desire*. An individual may desire an automobile but that desire can have no economic consequence unless the person possesses the ability to offer something valuable in exchange for it. A hungry beggar standing before a bakery window may have a very urgent desire for the food displayed therein, but his desire does not constitute a demand in an economic sense and cannot do so until he finds some valuable thing which he can offer in exchange for these goods. We may conclude, then, that the essential nature of demand is the desire for a good coupled with the ability to pay for it.

It should be noted that demand always relates to some particular good at a definite time and place. We may speak of the demand for wheat, or for any other commodity, in a particular market at a definite time. In this sense *demand* means the quantity of wheat which the buyers are willing to take from a market at a price. When the price is high, buyers are likely to take fewer units of the good than when the price is low. Buyers will continue to purchase units of a good as long as they expect to get some benefit from them, or until the benefit expected from another good becomes greater. It would, therefore, appear that while demand in its ordinary usage in the market implies

the quantity or volume of goods that is being purchased, it nevertheless registers the buyers' estimate of the significance of the goods through the prices at which they are willing to make their purchases.

Individual Demand Schedule.—Demand, as a market influence determining value, consists of the combined effect of the estimates of all the individual and prospective buyers of any good that is offered for sale. In order to understand the effect of this influence, it will be well to examine how an individual purchaser reacts in a market, but, before proceeding with this analysis, it should be recalled that we are assuming a perfect market, as described above, as the basis for this discussion. A little reflection on one's own mental processes will convince him that his willingness to buy varies with the amount that he has to offer in exchange and with the intensity of his desire for the good in question. At any given moment of time, because the general intensity of his desires and the amount that he has to offer in exchange may be assumed to be constant, the amount of any good which he will purchase will depend upon the range of prices at which the good can be secured. This fact can be illustrated as follows:

Take the case of wheat and assume that a buyer will purchase wheat in the amounts indicated and at the prices quoted:

50 bushels at	\$0.50 per bushel
40 bushels at	0.75 per bushel
30 bushels at	1.00 per bushel
20 bushels at	1.50 per bushel
10 bushels at	2.00 per bushel
5 bushels at	3.00 per bushel
0 bushel at	4.00 per bushel

Such a schedule is called an "individual demand schedule," and represents the amounts of a good which an individual will buy at the respective prices. It also expresses his estimate of the significance of the good in terms of money. The total demand of an individual for any good at a given time is the sum of the quantities of that good which he is willing to take at all possible prices. The student should recognize that this schedule is a hypothetical conception intended to represent the buyer's willingness to purchase definite quantities of a good, provided the price at which the good can be bought varies as indicated in the schedule. What actually occurs, of course, is that pur-

chasers buy definite amounts at definite prices, or fail to buy, if the price is higher than their estimate of the significance of the good to them. But every purchaser is affected by the price at which he can secure the good, and is likely to buy more units as the price declines, so that the demand schedule does fairly represent the buyers' attitude at any definite instant of time and is, therefore, a useful, logical device to illustrate the influence of demand in a market.

The amount which a purchaser is willing to buy ordinarily varies inversely with the price. An exception that we need to make to this general rule is that there are commodities that are so highly esteemed that an individual will continue to buy them even though the prices rise. In winter, coal is such a necessary fuel for the heating of homes that consumers will continue to buy it even though the price rises sharply. Some economy may be effected in its use by shutting off the heat from some parts of the house, if this is possible, or by maintaining a lower temperature. Any further economy will be in the use of other goods.

It should be noted that this demand schedule represents the estimates of an individual at a definite instant of time. It has assumed that his purchasing power has not changed and that the general intensity of his desires has remained constant. At a different instant of time, his schedule of purchases might be quite different from that indicated above on account of a change either in the intensity of his desire for wheat or in his ability to buy it, or in both. But even though these changes have occurred, he would be willing to buy definite quantities at varying prices, and these quantities would constitute his demand schedule under the changed conditions.

A Market Demand Schedule.—In what has been said thus far, it has been assumed that the reaction of a single buyer was being analyzed. However, each buyer in a market has a demand schedule for the commodity in question, even though these schedules may vary greatly because of differences in the intensity of the desires of the different individuals and in their ability to pay for the goods. Thus the demand for wheat in any given market at a definite period of time would be the resultant, or effect, of the influence of all the buyers present in this market. By consolidating, as it were, all of the individual demand schedules into one, we can arrive at what may be thought of as a

market demand schedule which would represent the amounts of wheat which all purchasers were willing to buy at all possible ranges of price. Such a schedule would have many of the same characteristics of an individual demand schedule, and especially would the amounts sold ordinarily vary inversely with price. More wheat could be sold at a low than at a high price.

From this statement, it is evident that demand, as a market influence, means the quantity of a good which all buyers stand ready to take from the market at a price. Demand always registers the buyers' influence in the market, but it is not a self-operating force because the quantity which the purchasers are willing to take from the market is always affected by the price. It is evident, then, that the price of the good is a factor in determining demand. Hence, demand cannot be regarded as an independent force in the market, as popular discussion so frequently implies, but must be recognized as subject to the influences which actuate the buyer, chief of which are his evaluation of the goods and his ability to pay for them. As a determinant of value, demand is used in the sense of the amount of any commodity which the buyers are willing to take from the market at stipulated prices.

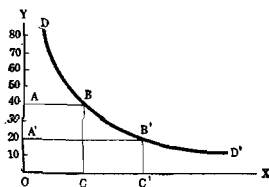


FIG. 3.

It is customary in the presentation of this topic to illustrate demand by what is known as a demand curve. This device should be understood simply as an aid in visualizing the effect of an abstract force. The demand curve is a graph which aims to represent a market demand schedule for any commodity. Along the line OX in Fig. 3 is represented the amounts which buyers are willing to take from the market. Along the line OY is represented the variations in price. At a price of 40-cents the buyers are willing to take OC quantity, while at 20 cents the

buyers are willing to take from the market the amount represented by the line *OC'*. By this method of representation all possible variations in demand can be presented and visualized. The shape of demand curves will vary greatly, depending upon the character of the commodity and upon all of the conditions which affect the estimates of the buyers, such as personal preferences, racial differences, etc. The inclination of this curve, however, will generally be negative, *i.e.*, declining toward the right, because ordinarily purchasers will buy more goods when prices decline. The truth of this conclusion is generally accepted. If we undertake to set down the actual amounts that buyers will purchase at various prices, as was done in the case of the individual demand schedule, the figures must be interpreted as hypothetical and arranged for the purpose of showing the general truth that the volume of goods purchased will increase as the price declines.

Diminishing Utility.—As has been stated above, demand expresses the buyers' influence in a market. Our next question is, How does this influence become effective in determining value? In attempting an answer to this question, if we assume that the buyers are always consumers of the goods purchased, we discover at once an important principle that has a vital bearing upon the explanation of the value of a good.¹ As the individual consumes successive units of a good within a given period of time, he finds that his desire for that good diminishes. This tendency of the desire for a good to diminish as successive units are consumed is a common and universal experience of man which rests upon his physiological organization.

To illustrate, let us take the desire for apples. We will assume that an individual gets a large amount of satisfaction from eating the first apple. Perhaps his appetite for the second may be almost as keen as for the first, but should he continue to eat a third, a fourth, and so on, he will soon find that his desire for apples is declining, and if he still continues to eat them

¹ This assumption is, of course, contrary to fact, as many buyers expect to sell the goods again or to use them in the manufacture of some other good which will be offered for sale, but for purposes of our analysis, it is not so far from the truth as at first appears. All buying is directly or indirectly connected with the satisfaction of the consumer's desires, so that the significance of all goods is derived from the contribution which each makes to the gratification of human desires.

his satisfaction may be turned into positive pain. This tendency of desires to become less urgent as their gratification is continued is reflected to the goods themselves. Since the utility of a good is its capacity to gratify a desire, it follows that, as an individual consumes additional units of a good in a given period of time, the capacity of the good to gratify his desire, hence, its utility, tends to decline. This tendency is known as the "principle of diminishing utility," which may be formally defined as the tendency for the utility of a good to diminish as successive units of it are consumed in a given period of time.

Because the change in the urgency of the desire is the fundamental influence, this principle might more accurately be called "diminishing desirability." However, usage, together with the fact that the question to be solved concerns a good, justifies the shift of attention from the change in the subjective conditions of the individual to the objective circumstance of the quality of the good, or its capacity to gratify a desire. For this reason, the principle is expressed as diminishing utility, but for a thorough understanding of the problem one should not overlook the causal fact of the change in the intensity of the desire of the consumer.

This principle holds true even though after a lapse of time the desire for the good may be recurrent and the appetite may be just as keen at the later as at the earlier date. In fact, in some instances the recurring desire may become stronger. Many persons have to acquire a taste for olives, which means that in coming back to this good after a lapse of time they find that their desire for it has become more intense and, therefore, its utility is greater. But this fact does not affect the general tendency of utility to decline as successive units of a good are consumed in a given period of time.

The principle has been illustrated by an example taken from the field of direct consumption, an experience common to all consumers. The question may be asked, Does the principle apply to producers' goods, such as machines, tools, etc.? In answering this question we must not lose sight of the fact that producers' goods are never desired solely for their own utility. They are agents that assist in the production of some good that contributes directly to the gratification of human desires. From this point of view, we conclude that producers' goods must conform, even though indirectly, to the principle governing consumers' goods, as the significance of the former is derived

from the importance of the latter. But even apart from this fundamental relation, the principle of diminishing utility applies to producers' goods, especially in the sense of diminishing desirability. In the accomplishment of a particular purpose, the manufacturer would regard one machine of greater significance than two, unless he expected to duplicate all of the conditions of production in his factory. This observation is in accord with all experience, namely, that additional units of a good tend to yield diminishing satisfaction.

Marginal Utility.—The recognition of the principle of diminishing utility is the starting point for the determination of the value of a good. Why do men buy goods? The expected satisfaction or benefit to be derived from their use is the force which brings every consumer into the market as a buyer, and in making his choices he is affected by the principle of diminishing utility. If the goods were free, he would carry his use of them to the point of complete satiation of his desire at that period of time. For instance, we use all the fresh air that we desire and if there were twice as much of it in the world, we would not consume any more per hour or day than we do now. But in the case of economic goods we pick and choose between alternatives. We continue to consume any particular good, say apples, until the satisfaction expected from some other good is larger than that expected from another apple. This observation brings us to an important step in the determination of the value of goods, so far as value is affected by the influence of consumers or buyers.

Let us now assume that a buyer expects a definite amount of gratification from the consumption of one apple, less from a second, still less from the third and fourth, and has no interest at all in the fifth. He will continue purchasing apples until his desire is completely satiated, or more likely until some other good will yield him greater satisfaction. Let us assume that that point is reached in this case after he has purchased three apples. Assuming that the apples are identical in size and quality, what will determine the significance for him of his stock of three? It is a matter of indifference which one of the stock he places in the third position, since all are alike in size and quality. The significance of each unit in the stock, then, will tend to coincide with the significance attached to the last one added. The satisfaction, or utility, derived from this added unit is known as "*marginal utility.*"

What this expression means may be further elucidated by the following statement. When an individual has consumed units of any good to the point that he expects to get a greater degree of gratification from a unit of some other good, that point constitutes for the time being his margin of use for the particular good in question. The satisfaction which he receives from that last unit, or the utility of the unit at the margin of his consumption of that good, is generally known as the "marginal utility" of that good to him at that moment of time. Because of the tendency for utility to diminish, it should be observed that marginal utility varies inversely with changes in supply of a good and directly with changes in the intensity of the desire for it. The significance of this concept in the explanation of value is the fact that it is an aid in explaining the buyer's influence upon value. It is the positive influence that lies back of demand. A buyer's estimate of a stock of goods is determined by the utility of the last added unit of the good (marginal utility) times the number of units of the stock. This is a statement of the subjective significance of the good to the buyer, and without that significance the good could have no value or power in exchange.

This idea can be represented by the following diagram. If we measure along the line OX in Fig. 4 the number of units con-

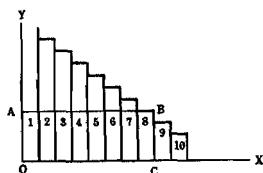


FIG. 4.

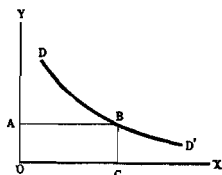


FIG. 5.

sumed, and along the line OY the satisfaction derived per unit, it will appear that each successive unit will yield lessened satisfaction. This conclusion is in accord with the principle of diminishing utility already described. The small, rectangular figures are intended to represent the amount of satisfaction which each unit will afford. It will be noted that unit 1 is not closed. The reason for this is that it is assumed that the satisfaction expected from the first unit is not capable of measure-

ment, as for instance, the utility of enough food to sustain life. After the most urgent pangs of hunger have been satisfied, the utility of additional units will yield a measurable degree of satisfaction, as shown in Fig. 4 by the rectangles 2 to 10 inclusive, and is measured by what he is willing to give for a unit of the good. This satisfaction will decline as the number of units available increases. At some point, let us say, at the eighth unit, he stops his purchases. The utility of the eighth unit in this case is known as the marginal utility of the good under the given circumstances. Since the utility of the last unit determines the utility of each of the other units taken, then, the area *OCBA* represents the significance of this stock to him, *i.e.*, the utility of the last unit times the number of units in the stock. If we assume that the units are small enough, the rectangles will become lines and the broken line becomes a curve as in Fig. 5. In this manner we can represent the way in which marginal utility determines the significance of a good to the buyer.

We have here represented the principle of marginal utility as determining the significance of a single good to an individual buyer. The reader should understand that this principle applies not only to all of the evaluations of an individual but it applies in the same manner to all buyers. From an economic point of view the choices that an individual makes are always in the order of their relative significance to him at that time. In making his purchases, he chooses those goods which he thinks will yield him the largest satisfaction. If he buys an automobile rather than a home, the only conclusion that can be drawn from this act is that, economically speaking, he regards the automobile of greater utility to him than a home. In other words, the individual tends to distribute his buying power among the great variety of goods which he purchases in accordance with the relative marginal utility of these goods to him. He tends to maintain a balance among the satisfactions derived from the last units purchased and to keep the intensity of his unsatisfied desires approximately equal. The extent to which he carries his purchases in all directions may be regarded as his margin of consumption. The location of this margin will differ among individuals, due to differences in income, to differences in personal tastes, and to different estimates placed upon the significance of present desires as compared with estimates of future needs.

In the determination of the value of any single good, all of these influences are at work. Since the evaluations of one individual have been taken simply as representative of the evaluations of all the buyers, it should be clear that in every market, from the side of demand, there is an estimate of the significance of the good in question, which is marginal. It may be that one or more of the buyers are on the point of dropping out of the market for that good. They constitute marginal buyers. It does not matter whether these marginal purchasers have bought one or many units of the good, the significant thing is that their estimates of the importance of the last portions of the supply are marginal. We may conclude that, from the side of demand, the value of this good is determined by its marginal utility to the marginal buyers.

We know by general observation that there are some purchasers in every market who do not economize in their use of any particular good; they satisfy completely their desires for it. Take a commodity like salt. How many persons are there who do not use as much salt as they would if it were a free good? What is true of salt is true in some degree of other commodities. Clearly, it is not the estimate of those persons who consume a good to the full satisfaction of their desire for it that determines its value. But the estimates of some persons for every good are marginal. These are the persons who economize in its use. It is these marginal estimates of the utility of the good that tend to determine the significance of that good in that market as a want-gratifier. Hence, we may say that, from the side of demand, the marginal utility of a good to marginal buyers tends to determine its value, and gives to the good its power in exchange.

Inequality of Income and Value.—There are two difficulties in the application of these principles to the concrete conditions of actual life, namely, how to measure objectively subjective influences, and how inequality of income affects the use of marginal utility as a determinant of value. We will now proceed to give consideration to each of these problems. We have said that marginal utility expresses the buyers' influence on value, but how can we measure the satisfaction which any one buyer hopes to get from a good when the gratification of desire is a subjective phenomenon? There is clearly no absolute measure of the satisfaction which two individuals get from the con-

sumption of identical units of a good. No one can state precisely whether A gets more satisfaction from an apple than B. The only objective evidence we have of the subjective effects which an individual gets from the consumption of a good is the willingness of the buyer to pay something to secure it. The order of preference in choosing goods and the amount that he is willing to forego in order to secure them furnish a basis for approximating the satisfaction that an individual hopes to get from their use. Before a consumer can obtain an economic good, he must pay for it either in effort, goods, or money. If he produces the good himself, the amount of time and energy which he is willing to devote to its production is an indication of the satisfaction which he expects to get from it. The utility must be of equal or greater significance to him than the effort expended in its production, together with any additional sacrifices that must be endured in order to secure it, otherwise there would be no rational explanation of his conduct. On the basis of this assumption we can at least infer that an individual prefers one good to another because he is willing to expend more time and energy or endure greater sacrifices in securing the first than the second. We are unable to express the subjective reactions of an individual quantitatively more accurately than this, but this basis of measurement is ordinarily sufficient for practical purposes.

But most goods are purchased with money so that the problem of measuring satisfactions objectively is usually concerned with money transactions. The question then is, Can we infer the degree of satisfaction which an individual gets from a good by the amount of money he is willing to pay for it? Assuming no change in the value of money itself, there are two inferences that can be drawn from the amount of money which an individual is willing to offer for goods. First, we can infer that he gets more satisfaction from the first unit of a good than from a second, if he is willing to pay more for the first than for the second unit. Then we can infer that he prefers one good to another, if he is willing to pay more for the first than for the second. This relation between the willingness to pay money for goods and the gratifications derived from their use is a fairly satisfactory basis for measuring the relative significance which an individual attaches not only to different units of the same good, but to units of different goods. It may be asserted also that this relation

obtains for every individual, whatever may be the amount of money at his disposal.

A serious difficulty is encountered, however, when an attempt is made to use this method of comparing the satisfaction which different individuals get from the use of the same goods. It would be incorrect to infer that A gets the same degree of benefit from a good as B, because the two are willing to pay the same amount of money for it. If A and B had equal amounts of money to spend, then the inference that the satisfaction was about equal, would be substantially correct if they were willing to pay the same price for a unit of a good. But inequality in the amount of money at the disposal of buyers complicates the use of this method of testing objectively the degree of satisfaction which different individuals get from the use of any good.

If A has twice as much money as B, and each is willing to spend 25 cents for a unit of some good, it does not follow that the benefit derived is equal. Under these conditions A can carry his margin of consumption lower, *i.e.*, to less important or urgent desires, than is possible for B. Therefore, assuming that A and B have the same rate of diminishing utility, the expenditure of 25 cents by A would yield less satisfaction than the expenditure of the same sum by B. This evidence is conclusive that the expenditure of the same sums of money cannot be made the basis of inferring equality of benefit between different persons.

But what bearing does the recognition of this fact have on the use of the marginal utility principle in the determination of value? In considering this problem, we are not concerned with the degree of satisfaction which a definite amount of wealth might yield if it were differently distributed. If wealth were differently distributed or, in the case above, if A gave B a portion of his money income, it would be capable of yielding a larger maximum of satisfaction, because the less important units which A gave up would presumably yield a greater satisfaction when consumed by B. But the real problem in the determination of value is to discover how the influence of buyers is brought to bear upon the value of a good.

In attempting to solve this problem, one must remember that the principle of diminishing utility affects every user of a good; hence, each user will have a marginal estimate of the significance of that good as among the various goods for which he has a

demand. As was stated above, in the market for every good, there are those whose estimates have little bearing upon the determination of value; their estimates are distinctly above the margin. But in every market, there are those whose estimates are marginal, and who are on the point of dropping out of the market. The estimates of these persons, whether it is the fifth purchase by those who have much money, or the first by those who have little, is immaterial. The fact that there appears a marginal offer is the significant thing in the determination of the power in exchange of any good. As a determinant of value, marginal utility is used in the sense of the utility at the margin of use, or the utility to the marginal users in that market. In this sense, the willingness to pay for a unit of any good is a fair measure of the benefit marginal users expect to get, and the amounts which they are willing to offer for the good tend to determine the exchange power of that good. In this way marginal utility to marginal users is the influence that determines value from the side of demand.

Before leaving this point the following caution should be expressed. The reader should guard himself against inferring from the above statement that the marginal users, in case there is more than one, necessarily get like degrees of satisfactions from their marginal purchases. The marginal unit may be the fifth one purchased in the one case while it may be the first in the other. To conclude that the utility of the fifth unit to one person is equal to the utility of the first unit to another person, both of whom may be at the margin of use of this good, is unwarranted. But the problem is not that of finding an expression to compare the utility of marginal units to different persons, but rather one which will express the influence of marginal buyers on the determination of value.

Whenever an exchange is being made and, hence, the phenomenon of value emerges, it is important to remember that two conditions that have a bearing on the point under discussion are always constant, namely, the distribution of wealth among the buyers and the value of money or other medium of exchange. While these conditions may change from one point of time to another, *i.e.*, wealth may be differently distributed today from what it will be tomorrow, or the value of money may be higher now than it will be a year from now, yet they may be said to be constant when an actual exchange is being made. This means

that whenever an exchange takes place, marginal purchasers are likely to have unequal abilities to offer something valuable in exchange for the goods they wish to buy. Furthermore, the money medium in which the exchanges take place may, for all practical purposes, be said to be constant in value during the process of exchange.

Since these two conditions remain constant whenever exchanges are being made, the significant influence in determining the value of any good from the side of demand is the marginal price offer by marginal users. If the marginal buyers are willing to offer a definite sum, say 25 cents, for the marginal units of a good, that offer tends to determine the exchange power of that good in that market. The utility of this unit of money, or the marginal price offer, may vary between the marginal buyers because of the differences in the amount of wealth each has at his disposal. Nevertheless, each is subject to the principle of diminishing utility in the disposal of his wealth and each will offer at the margin of purchase an amount that will just equal his estimate of the satisfaction he expects to receive from an added unit of the good to be purchased. Since the principle of diminishing utility tends to govern the choices of each of these persons and since each tends to carry his use of the good up to his own margin, it seems correct to assert that marginal utility, or the utility to marginal buyers, is the force that determines the value of a good from the side of demand.¹

¹ Taussig has introduced the expression "marginal vendibility" in place of marginal utility to take account of the inequality of income of individuals who are competing for the control of goods. ("Principles," 3rd Ed., Vol. I, pp. 123, 134, 142, 148, 171, and 179.) While there is a significant point here that escaped the attention of earlier writers, it does not seem to me that the expression "marginal vendibility" is distinctly superior to that of "marginal utility," because I see no satisfactory reason for stressing inequality in the distribution of wealth in the explanation of value. Everyone would admit, I think, that a given stock of wealth might yield greater utility if it were differently distributed, but that is not the problem before the student of value. The value problem appears wherever exchange takes place, and exchange occurs even though the distribution of wealth is very unequal. Since every buyer is subject to the principle of diminishing utility and since each one will carry his purchases up to the point where the satisfaction expected from an added unit is marginal, it would seem correct to say that marginal utility to marginal users determines the value, or exchange power, of the good in question. To consider whether these marginal units yield different degrees of satisfaction to different persons is approaching the realm

PRINCIPLES OF ECONOMICS

Total Utility—Consumers' Surplus.—From the consideration of the principle of diminishing utility and marginal utility we get the concepts of "total utility" and "consumers' surplus." By total utility is meant the total satisfaction which consumers actually get from the consumption of successive units of a stock of goods. It represents a summation of the satisfactions derived from a good as successive units of it are consumed. As has been shown, the satisfaction derived from the first unit is greater than that derived from the second, and so on. The total satisfaction yielded by the whole number of units consumed is the total utility. Total utility is not the element that determines the value of a good from the side of the buyer but, as has been said, it is the utility of the last added unit of that good that determines for him its exchange power. It is evident from this discussion that a buyer may, and often does, get a surplus of satisfactions when he purchases a good. In other words, he may find a situation in the market which will enable him to buy a commodity for less than he would be willing to pay for it. His estimate of its significance is greater than the market estimate. Most buyers find themselves in this situation in respect to many of their purchases. Who would not pay more for bread or other food products than the market price in order to get a daily supply of these commodities? If any buyer of any commodity is willing to pay more for a unit of a good than the market conditions require, that buyer is enjoying a surplus of satisfactions. In Fig. 5 the area above the line *AB* represents the satisfaction in excess of that at the margin and hence may be regarded as a surplus. This surplus is usually called "consumers' surplus." It has no influence in determining the value of the good in a competitive market. We will find later in the discussion of monopoly value, that consumers' surplus becomes an important consideration.

of social philosophy and is unnecessary in the explanation of value, as we are not concerned in this explanation with the problem as to how wealth should be distributed in order to get the maximum satisfaction out of a given stock, but only in explaining why units of any good have power to command units of other goods in exchange. To me a better expression than "marginal vendibility" is "marginal price offer" which retains the influence on the side of the buyers where the expression "marginal utility" originated. The former expression shifts the point of consideration from that of the demand to that of supply, which, I think, is objectionable.

Nature of Supply.—Up to this point in the consideration of the determinants of value, the discussion has been confined to the nature and influence of demand. It will now be necessary to inquire into the nature of supply and its effect upon the determination of value. Demand and supply are always associated in the market and react upon each other. By "supply" is meant the combined amounts that sellers stand ready to offer on the market at a price. If the sellers hold some goods which they will not part with at any price, these goods should not be included in the concept of supply for they can have no bearing in the determination of value. This definition of supply should not be confused with the total amount in the hands of the producers or dealers, which should be thought of as composed of two parts, namely, "actual supply," or the amount that the sellers are willing to dispose of at the prevailing price, and "potential supply" or the amount they would dispose of if the price offers were more favorable.¹

Actual supply is the influence that is most potent in determining value in the current market. Potential supply, or the amount which remains in the hands of owners for purposes of sale, will affect both buyers and sellers if the volume of this stock is generally known. If the buyers are informed as to the volume that may come into the market, this knowledge will affect their price offers. For instance, the supply of wheat actually on the market may be 1,000,000 bushels, but from crop reports the season's yield may be known to be 900,000,000 bushels. This latter sum constitutes the limit of supply for that year, but only that portion of the total amount that finds its way to the market for sale can have any influence in determining the value of wheat. How much the farmers will hold over from one season to another, how large the wastage from loss of weight due to poor methods of storing, how much the farmers may decide to feed to stock cannot be known definitely, so that the market is concerned with the evaluation of the amounts that the sellers stand ready to offer for sale, affected, of course, by the knowledge of the potential supply. This dependence of supply on price gives rise to the concept of "supply price," which means the price that is necessary to call forth a given amount of a good and bring it to the market.

¹ Professor Taylor calls the total amount in the hands of producers and dealers "the stock." Cf. "Principles," *op. cit.*, p. 268.

With this definition of supply in mind we may again revert to the action of an individual and inquire how, as a seller, he will respond to market influences. A farmer probably would be willing to dispose of his entire crop of wheat if he could get as much as \$1.50 per bushel, but he certainly would hesitate to sell his crop at 70 cents per bushel.¹ Between these figures, we may assume that his willingness to sell would vary. As the price declines, he would be less willing to sell and offer a smaller amount. There would doubtless be some price at which he would be unwilling to part with any of his crop but would hold it with the hope of a later advance in price, or he might devote it to some alternative use, as having it ground into feed for his cattle or hogs.² The amounts of wheat which this farmer is willing to sell at the respective prices would constitute his individual supply schedule. Such a schedule may be illustrated as follows:

Range of Prices	Amounts Offered Bushels
\$1.50	5,000
1.40	4,500
1.20	3,000
1.10	2,000
1.00	1,000
0.80	500
0.70	0

This supply schedule may be represented as a supply curve. Along the line *OX* in Fig. 6 measure the amounts offered, and along the line *OY*, measure the range of prices. From this graph it will be seen that at 80 cents only 500 bushels would be offered; at \$1, 1,000; at \$1.10, 2,000; at \$1.20, 3,000; at \$1.40, 4,500, and at \$1.50, 5,000 bushels.

By connecting the respective points we get an individual supply curve *SS'*. But every farmer owning wheat has a supply schedule that represents the amounts he would be willing to

¹ The student should understand that the argument presented does not depend upon the particular range of prices mentioned. If these prices did not call forth the responses stated, some others would. The important point to get clearly in mind is the fact that sellers do have a supply price, which will bring their supply to the market for sale.

² In 1893 the price of wheat was so low—40 to 50 cents per bushel—that the farmers of Southern Indiana fed a considerable amount of their crop to hogs. It was worth more to them as feed than as wheat at the existing prices.

offer for sale at the respective market prices. The combination of these supply schedules of individual sellers would constitute the supply offered in the market at all possible prices. By assuming the units of change to be infinitely small, we get a supply curve which represents the amounts offered by all sellers at all possible prices. Such a curve may be illustrated as follows:

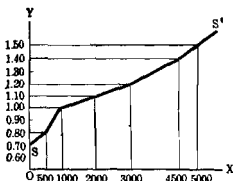


FIG. 6.

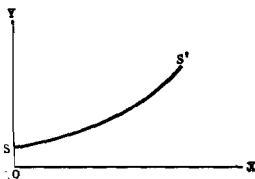


FIG. 7.

The amounts offered for sale are indicated along the line OX and the prices along the line OY in Fig. 7. The curve SS' represents the amounts offered by all sellers at the various prices. This curve constitutes the supply curve for this particular market. The shape of this curve will depend upon the way in which the sellers respond to changes in price. From this discussion, it should be apparent that supply, like demand, is not a self-operating, independent force, but is affected by the price that prevails in the market. If that price is high, the supply will tend to be large, and *vice versa*. We may conclude, then, that supply tends to fluctuate directly with price.

From the previous discussion, we have seen that demand tends to fluctuate inversely, while supply tends to vary directly with price. By bringing these two influences together in a competitive market, we find that, because they work in opposite directions, there is one price at which all of the supply offered will be taken from the market. Supply and demand will be equal at some price point. The effect of these two forces under the circumstances assumed can be shown graphically by combining Figs. 5 and 7, as is done in Fig. 8.

Indicating supply along the line OX and price along OY , it is assumed that larger quantities will be offered for sale as the price increases, as shown by the curve SS' , and that the utility of the good becomes less significant as the supply increases,

as shown by the curve DD' . A few units might be taken at the price of OF , but this price could not prevail, because the competing sellers would bid against each other and offer units of the good at lower prices. This competitive bidding would continue until the point P was reached where the amounts offered

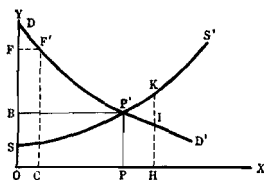


Fig. 8.

for sale at the price PP' were just taken from the market by the marginal purchasers. The sellers would be unwilling to sell more unless the prices were higher than PP' and, if more were offered, the marginal utility would be less, so that marginal buyers would be unwilling to pay more than HI , while no seller could be induced to part with the increased amount for less than HK . Thus, the amount that could be sold upon this market would be measured by the line OP , and the price would be PP' . At this price the supply offered for sale would just equal the amount purchasers were willing to buy at that price.

From this analysis we get a market principle, *viz.*, that in every perfectly competitive market there can be but one price for the same good at the same time. Competition among buyers prevents one buyer from getting the good at a lower price than that at which every other buyer may secure it and, in like manner, competition among sellers prevents every seller from getting a price higher than that offered every other seller. As a result of these competitive forces, demand and supply tend to be balanced at the price that will clear the market of the entire stock of the good. This principle is stated boldly on the assumption of a perfect market, but competition, while present in practically all markets, does not work perfectly. To the extent that there are obstacles to competition, or that there are conscious influences controlling the market, this principle needs to be modified. Yet, since competition is so actively present in the marketing of most commodities, we can conclude that there is a tendency for competition to establish a single market price in a given market at a given time.

From the preceding discussion, we can state the essential elements of the principle of demand and supply, which, we have

seen, is so generally accepted. In a competitive market, price will rise as long as demand is in excess of supply and will fall as long as supply is in excess of demand. This movement of price up or down will continue until some price is found that will clear the market. This is the principle of demand and supply. Its formulation depends upon the existence of a perfect market, as described above, but even where competition does not work perfectly, there is a tendency for the market conditions to conform to the principle as stated. Wherever competition is present there will be a tendency to establish a single market price that will balance demand and supply, and changes in these two elements will react on price as stated.

Demand in the Market and the Schedule Sense.—There are two further considerations concerning demand that should now receive attention. We have defined demand as the quantity that purchasers are willing to take from the market at a price, and have stated that the higher the price, the fewer the units of a good that will be taken, and the lower the price, the larger the number that will be purchased. Changes in demand in this sense simply indicate the quantity of goods that can be disposed of at different price points. At lower price points, in addition to the larger quantity which previous purchasers are likely to buy, new purchasers will enter the market and, as a result, a larger volume of supply will be absorbed. Increase in demand may, therefore, mean merely that more goods will be sold as the price per unit falls. But there is another sense in which we may say that demand has changed. Suppose the intensity of our desires for a good has increased, as for instance, our desire for ice during an exceptionally hot summer, or for coal during a severe winter. Under these circumstances marginal purchasers will be willing to offer higher prices for marginal units than they would otherwise be willing to give. Such a change is common enough and would be recognized by merchants as an increase in demand.¹

We may illustrate this point by assuming that the same amounts are sold as before, but that the price has been materially

¹ This argument may be reversed. As is well known to merchants, an exceptionally cool summer may retard the buying of summer goods, and in order to clear the market sharp reductions in price may have to be made. What consumer has not benefited by such market changes and purchased a straw hat or a palm beach suit at greatly reduced prices?

increased. To put the case concretely, suppose that during a certain week ice is selling at 45 cents per hundred pounds and that the next week, on account of the extreme heat, the purchasers are willing to pay 65 cents for the same amount. Under these circumstances, the intensity of the desire has increased so that buyers are more eager to buy than before. This situation is rightly regarded as a change in demand.

These two uses of demand are represented in Fig. 9. In this figure the curve DD' represents the total demand schedule. It indicates the amounts of the good that would be purchased at all possible prices. It shows that more ice would be purchased at the price BP' than at AP . Changes of demand of this character

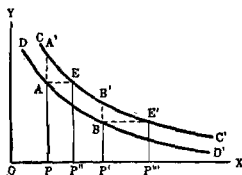


FIG. 9.

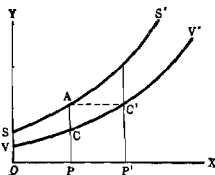


FIG. 10.

may be regarded as a change in the market sense.¹ The purchasers will take no more pounds of ice at the various price points but they will buy more as the price becomes lower. But when the intensity of the desire has changed, the whole demand schedule becomes modified as indicated by the new curve CC' . The purchasers are willing under these circumstances not only to take more units at the old prices as indicated by the lines OP'' and OP''' , but are willing to pay higher prices for the same quantities, as indicated by the lines PA' and $P'B'$. This situation may be called a change of the demand in the schedule sense.

Supply in the Market and the Schedule Sense.—The same situation exists in respect to supply. We may speak of an increase of supply in a market sense, which means an increased number of units offered for sale as the price is increased per unit, or in the schedule sense, which means an increase in the amounts offered for sale at each price point. These two condi-

¹ This subject is well treated in FISHER, I., "Elementary Principles of Economics," pp. 258-277. Cf. also FAIRCHILD, F., "Essentials of Economics," pp. 133-163.

tions of supply are represented graphically in Fig. 10. The curve SS' represents the amounts offered for sale at the different price points. The increase in amount from OP to OP' is increased supply in the market sense. But should the conditions of supply so change that the sellers are willing to offer a larger amount, as OP' , for the old price AP , or if they are willing to sell the same amount OP at the lower price PC , then we can say that the supply has changed in the schedule sense.

It is interesting to note that a change in demand in the schedule sense is always associated with a change in supply in the market sense, while a change in demand in the market sense is associated with a change in supply in the schedule sense. The following figures will show this fact graphically. In Fig. 11, the change in

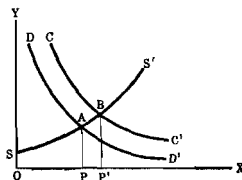


FIG. 11.

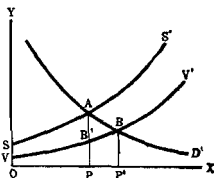


FIG. 12.

demand in the schedule sense is shown by the two demand curves DD' and CC' . It will be apparent that the two price points A and B , which represent increased intensity of the desire of the consumers, both lie in the supply curve SS' . At these two points, supply and demand are equal and market forces are in equilibrium. Marginal utility of the good has increased as is shown by the willingness of buyers to pay a higher price. The increased supply here indicated is in accord with the market supply schedule and represents the willingness of the sellers to offer additional units at the higher price point.

Now, if we draw our curves so as to show an increase in demand in the market sense, we will see that it must be accompanied by an increase of supply in the schedule sense, as shown by Fig. 12. Here the increase in demand shown by OP' has taken place because the new amounts can be purchased at lower prices. The marginal utility of the good has fallen. The market estimate has not changed for the same quantity OP , but it would not be

possible to secure the larger volume at the lower price unless the sellers had changed their attitude toward the whole volume to be sold. They are now more willing to sell and will, therefore, dispose of larger amounts at the former price or accept a lower price, PB' , for the former amounts. Such a situation could arise only when there was a complete readjustment of the seller's estimates or willingness to sell. This means that the supply in the schedule sense has changed.

How do these changes in demand and supply affect market price? There can be no change in market price unless there has been a change in supply, or in demand, or in both. Other things remaining constant, an increase in demand in the schedule sense will cause an increase in price. Under these circumstances, sellers will find an increased willingness on the part of purchasers to buy, as shown in the demand for ice in the hot period referred to above. In a competitive market, the increase in demand will be followed by an increase in price. So far then as competition is present in any market, there will be a tendency for price to rise when the increased demand results from greater intensity of desire for the good. Likewise, a falling off in the intensity of the desire will be followed by a fall in price. If the changes should be on the side of the supply and there be an increase in supply in the schedule sense, there will be a tendency for price to fall off and *vice versa*. Thus, we can conclude that, other things equal, a change in demand will cause price to move in the same direction as that of demand, while a change in supply will cause price to move in the opposite direction.

Elasticity of Demand.—A second consideration of demand that requires further elaboration is based upon the foregoing discussion. It is well-known that the demand for some commodities changes much more rapidly than for others. The more vital and urgent the desire for a good, the more constant will be the amounts sold, irrespective of price changes. When demand changes slowly with changes in price, we say that the demand is *inelastic*; and when it changes rapidly with price changes we say that demand is *elastic*. Figures 13 and 14 illustrate these two different conditions of demand. In Fig. 13, the quantity sold changes slowly with changes in price. At a price OA the amount sold is OB , but if the price should fall to OA' the amount sold will increase to OB' , but the increase is not rapid. In Fig. 14, the fall in price from OA to OA' is followed

by an increase in sales at a much faster rate. In the first case the demand is inelastic and in the second it is elastic.

The difference between elastic and inelastic demand is one of degree only. There is no sharp and fast line dividing the demand for one type of goods from that for another. But in general, we can say that the demand for the *necessaries* of life is inelastic, as it changes slowly with changes in price, while the demand for *luxuries* is elastic. Roughly speaking, this classification holds true. The amount of bread purchased will vary comparatively little with changes in price. People may economize slightly when the price is high, but the amount of increase in the consumption when the price declines is relatively small. On the

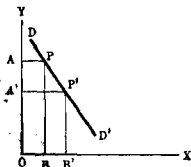


FIG. 13.

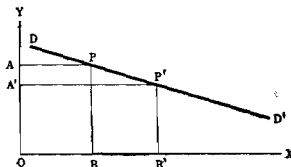


FIG. 14.

other hand, the number of Ford cars that can be sold at \$495 is very much greater than if the price of the same cars were \$1,000. In other words, the demand for Ford cars is elastic. As the price of the car has been lowered, new purchasers have been found in great numbers, hence, we may say, that the demand is elastic. What is true in this illustration would be true in all similar cases.

There are a number of influences that affect the elasticity of demand. As stated above, the demand for the necessities of life is less elastic than the demand for luxuries. If the use of a good constitutes a habit, the demand is less elastic than is the case for commodities where the choice of purchases is not so restricted. Then the existence of substitutes renders demand more elastic. If purchasers can easily find some other good that will answer their purposes, they will respond quickly to changes in price, hence, the volume that can be sold will be greatly affected by these price changes. Inequality of incomes also has an important bearing on the elasticity of demand, for if incomes were equally distributed, then diminishing utility would be the only cause affecting the elasticity of demand

and would accurately portray the satisfactions derived from successive units of a good. But inequality of incomes enables the rich to pay a high price for the first increments while the poor, or those with small incomes, cannot purchase the goods until the price has fallen materially. Thus, we may have a rapid change in the demand for a good at one end of the price schedule and a slow change at the other. Generally speaking, the demand for a particular commodity by persons with large incomes is less elastic than that of persons with small incomes. This probably does not hold to the same extent for the total purchases of the rich as for the purchase of a single commodity, because of the diversity of their wants. But even taking this fact into account, it probably is correct to say that the demand of the rich tends to be less elastic than that of those having small incomes. We frequently hear it said that the extremely rich save in spite of themselves, since they cannot find ways of spending their entire income. In so far as this statement portrays an accurate situation, we can conclude that the total demand of the extremely rich is less elastic than that of the extremely poor.

The elasticity of demand does not affect the fundamental principle of value, but its recognition has an important bearing upon the application of this principle. From the side of demand value tends to be determined by the marginal utility of the good to marginal purchasers. What they are willing to pay tends to determine the price of the good at that time. This is the simple principle. The elasticity of demand affects this principle in the following manner: In the previous discussion it has been shown that the elasticity of demand varies as between different goods and it will follow from this that the demand for one commodity will be affected by changes in price of other commodities. Suppose a shortage of bread stuffs. The price will tend to rise. The demand for bread being inelastic, purchasers will buy about as much bread as when the price is lower. But the higher price of bread will tend to lessen the amount of other goods which these persons can buy. In this way, then, the demand for one commodity may affect the demand for all other commodities. The competitive forces that tend to fix the price of any one commodity are affected by the demand for all other commodities. The differences in elasticity will modify the value of the different commodities as indicated.

Suggested References

- BYE, RAYMOND T., "Principles of Economics," Chap. XV.
CARVER, THOMAS N., "Principles of National Economy," Chaps. XXIV-XXV.
CHAPMAN, S. J., "Outlines of Political Economy," Chaps. XIV, XV, and XVI.
CLAY, HENRY, "Economics for the General Reader," Chap. XV.
EDIE, LIONEL D., "Economics, Principles and Problems," Chaps. VIII and IX.
ELY, RICHARD T., "Outlines of Economics," 4th Ed., Chap. X.
FAIRCHILD, FURNESS, and BUCK, "Elementary Economics," Vol. I, Chaps. XI and XVI.
FETTER, FRANK A., "Economic Principles," Chaps. I-VIII.
FISHER, IRVING, "Elementary Principles of Economics," Chaps. XV-XVI.
GIDE, CHARLES, "Principles of Political Economy," Rev. Ed., pp. 187-197.
HAYES, H. GORDON, "Our Economic System," Chaps. XVIII-XX.
HENDERSON, HUBERT D., "Supply and Demand," Chaps. II-III.
RUFENER, L. A., "Principles of Economics," Chaps. III, V, and VI.
TAUSSIG, FRANK W., "Principles of Economics," 3d Ed., Vol. I, Chaps. VIII-X.
TAYLOR, FREDERICK M., "Principles of Economics," Chaps. XIX-XXIV.
VANDERBLUE, HOMER B., "Economic Principles," Chap. V.

CHAPTER IX

COST OF PRODUCTION AND SUPPLY

In the preceding chapter, after explaining the nature of value, its relation to demand and supply was analyzed. It was shown that demand and supply were not self-operating, independent forces like gravitation, but were themselves dependent upon more fundamental forces. In the case of demand, the fundamental force was found to be the evaluation of the good as a want gratifier by marginal consumers, and the term used to signify the satisfaction derived by these consumers was "marginal utility." The significance of the last added unit of the good to marginal purchasers is the influence which determines the importance of a given stock of goods from the side of demand. In considering supply as a market influence, no attention was given to the forces that determine the amounts that sellers are willing to offer for sale. It was tacitly assumed that the sellers intended to offer their entire holdings for sale. But just as in the case of demand, there are fundamental forces which lie back of supply. It is our task in this chapter to examine into the character of these forces and to analyze their bearing upon the determination of value.

Scarcity and Value.—One of the conditions necessary for a good to have value is scarcity. If a good is useful and at the same time scarce it will have power in exchange. Let us now consider the influences that cause goods to be scarce. Four reasons may be assigned to account for this scarcity: (1) "The niggardliness of nature,"² (2) the expansion of desires, (3) the cost of production, and (4) monopoly.¹ The first of these causes means simply that nature does not provide free and ready for use all of the goods that mankind desires. Nature has, however, bounteously supplied man with raw materials from which he can, by his own efforts, create a vast flow of wealth.

¹ CARVER, T. N., "National Economy," pp. 354-363. Professor Carver has presented an excellent chapter on the subject of scarcity in relation to value, from which I have drawn heavily. Cf. p. 355 for a classification of the causes of scarcity.

But it takes effort on man's part to secure these goods and this effort is an important influence in limiting their supply. In the second place, scarcity is caused to some extent by the fact that human desires expand. This expansion may result either from the fact that population increases faster than wealth is produced, or from a change in the character of the desires which will show itself in a higher standard of living. The development of civilization has been accompanied by a rapid expansion of desires in both of these directions and, as *has* already been shown, they seem capable of indefinite expansion.¹ The net result of this increasing capacity to consume is pressure upon the supply of wealth available for use. The expanding desires of the human family, therefore, tend in the ways indicated to create a scarcity of goods.

The third cause of scarcity is cost of production. Since nature does not furnish free all the goods that man wishes but *does* provide raw materials in large quantities, it becomes necessary for him to direct his mental and physical energies toward the production of goods in a form in which he can use them, or to bring them into the time or place in which they can be used. An examination of produced goods will reveal that some things from which mankind gets satisfaction are absolutely fixed in quantity, as first editions of books, or paintings and musical compositions of old masters. But the vast majority of the goods which man consumes is capable of reproduction. Why, then, should there be scarcity when goods can be made by human effort? The only answer to this question is that the effort required to produce goods is irksome.

If men got pleasure from the making of goods and if they never experienced fatigue, we could imagine the supply of goods becoming so plentiful that they would have no value. No one would be willing to exchange anything for them. Some goods are produced for the love of the exercise required in their making. For instance, men like to hunt and fish, and the pleasure derived from the sport itself will induce them to put forth the effort necessary to secure the game, quite independent of the satisfactions obtained from its use. But if the demand for game and fish is not met by the effort that creates its own satisfaction, then some compensation will have to be offered in order to induce

¹ In Chap. III, this subject was discussed under the topic, "insatiability of desire."

the men to exert themselves further in hunting and fishing. The price paid is an attempt to overcome the disinclination of producers to carry their efforts beyond the point at which the effort is self-compensating. The price paid to get the work done is a cost that affects supply and is, therefore, a cause of scarcity.

Finally, monopoly is in modern production an important influence in creating scarcity. Monopoly has already been defined as the conscious control usually exercised over supply for the private benefit of the monopolist. The monopolist may decide that he will produce a definite quantity and then sell that quantity for what it will bring, or he may decide that it is to his advantage to sell his goods at a definite price and then produce such quantities as the market will absorb at this price. In both cases, he relies upon the market forces to determine the price of the goods offered. At this point we are concerned only with the effect of monopoly as a cause of scarcity, and it may be stated that wherever monopoly exists there is always present the possibility of a limitation of supply. A fuller treatment of its relation to value will be found in Chapter X.

Nature of Cost.—Returning to a consideration of the nature of costs, it should be apparent from what has been said that this term may be used in two senses. It may refer to the fatigue, sacrifices, or other disutility which the producer assumes in making the good, or it may refer to the payments that are offered him as compensation for the assumption of these various forms of disutility. Costs, in the sense of effort or sacrifice, may be called *psychic costs*, or disutility, while the payments made to induce men to endure these psychic costs may be called *expenses of production*. In ordinary business transactions, the term "cost of production" is used to refer to the money outlays necessary to get a good produced and therefore may be more properly referred to as expenses of production. Cost of production may be used in both senses as we proceed, but it will be possible to determine from the context whether emphasis is placed upon one or the other of these two uses. At this point, it is important to realize that costs of production in the psychic sense are the real cause in determining scarcity. These costs are usually translated into expenses of production when some form of compensation is offered in order to induce men to assume

Kinds of Costs.—Three different types of psychic costs appear in industry which have to be compensated in one way or another. Some forms of effort are disagreeable from the start. Work may be dirty, or dangerous, or otherwise disagreeable. Costs of this character may be called *pain costs*. The progress of civilization and the application of mechanical means of producing things tend to reduce pain costs as an influence in production. By the exercise of his ingenuity, man keeps turning over to machinery the disagreeable forms of work. One needs only to observe the operations in the erection of a modern building to be convinced of the truth of this statement. Excavations are now made by steam shovels and the debris is hauled away in motor trucks. Foundations are quite generally made of concrete which is mixed in power-driven mixers. Innumerable instances similar in character could be cited from this and other industries in support of this statement, but enough has been said to suggest additional cases to the thoughtful reader.

In other cases, the effort at the start is pleasurable but, in order to get as much of the good as society desires, effort has to be carried to the point of fatigue. Cost in this form may be called *fatigue cost*. Most forms of production involve this type of cost. Men generally enjoy a certain amount of activity for its own sake and without doubt many men get some pleasure out of their daily tasks. If all work were of this character, fatigue costs would have a very small influence in limiting the supply of goods, but since this is not the case, these costs are a very positive and significant force causing scarcity.

Then, men may be reluctant to undertake some form of effort, not so much because the effort is fatiguing as because its use in one direction prevents them from using it in another. Such a cost is known as an *opportunity cost*. The giving up of opportunities to direct effort in other lines may become irksome and require a payment in order to induce the individual to choose a given occupation or to perform a certain task or job. The general improvement of the economic welfare of a people increases the importance of opportunity costs. High wages make long hours of work irksome, not solely because they are fatiguing but because the larger income creates opportunity for recreation and amusement that would not otherwise exist. Laboring men generally have come to demand higher pay for overtime and holidays. Work on a holiday requires no more physical energy

than on other days but, because of the opportunity for recreation and amusement which this time affords, it has become well-nigh universal to pay a higher wage for work on these than on other days. This is clearly an attempt to offset an opportunity cost, and the significance of this fact in the production of goods has become everywhere increasingly important. It is a factor limiting the supply of economic goods.

Diminishing Productivity and Increasing Costs.—In dealing with costs in relation to scarcity, we find that as population has increased it has been necessary to utilize the natural resources more intensively, or to bring the goods from greater distances. To transport goods from greater distances requires more effort and hence greater costs. To attempt to produce a greater supply within the more limited area brings producers into contact with the principle of diminishing productivity.¹ While it is possible to increase absolutely the amount of production by increasing the labor and capital devoted to land, the rate of production does not remain constant and after a certain combination of the factors has been reached, the increase in the volume of wealth produced will be at a diminishing rate. Because of this fact, more effort will be required to secure an added volume of product, which will increase the psychic costs involved, and will, therefore, add to the expenses of production necessary to induce men to incur these increased psychic costs. Increasing cost of production, whether viewed as a psychic cost or as an expense, is the most important factor in limiting supply and, therefore, of creating scarcity.

Items Entering into Expenses of Production.—Costs appear to the business man in the form of money outlays, or expenditures. In order to bring the foregoing discussion somewhat closer to the experience of the business man, the items which enter into costs in the sense of expenses of production may now be examined. Viewed from the standpoint of an average factory, or other industrial enterprise there are at least four groups of costs: (1) labor cost; (2) material cost; (3) burden, or overhead cost; and (4) selling costs.² All of these call for a direct outlay of money and, therefore, have an effect upon the supply of goods

¹ This principle was explained in Chap. IV under the subject of production.

² In retaining the term "costs" here the student should remember the two senses in which this term is used. It is here used in the sense of expenses, or money outlays.

which a factory can produce and sell. These items will be considered separately.

Labor Costs.—Labor cost does not mean the wage rate per day or other unit of time, nor the total payroll of wages, but the cost of labor per unit of output. While the manufacturer is concerned with the total outlay in the form of wages, the determining factor in his ability to meet competition is the unit cost of his product. There are many factors affecting the labor costs per unit of product, among which may be mentioned the organization and layout of the plant, the control of materials so that there will be the least possible lost motion both of man and machinery, the methods of pay, and lastly the good will of the employees. The psychological attitude of the workmen toward the management has a tremendous influence upon labor costs and one of the significant problems of management is to keep these unit costs low in order to meet market competition.

Material Costs.—Next are the material costs which include not only the outlay for the materials from which the product is made, but the expenditures for freight, cartage, handling, and storing. Any items of expense that may be assigned to the material from which a good is made should be included in this group. Material costs enter into all production, except in the rendering of personal services, as the services of a lawyer, preacher, teacher, etc., from which no tangible product appears.

Overhead Cost.—In addition to labor and material costs there is in all production an overhead cost or burden. Classifications may vary for accounting purposes, but there are always costs for general administrative expense, including superintendence, clerical, and accounting services. All labor which is necessary for the successful operation of a business, but which does not contribute directly to the output, is included in the idea of burden. For instance, a shoe factory has need for a carpenter, a machinist, a storekeeper, and the like—services that are required to keep the plant and equipment in proper condition—but the costs of these services cannot be assigned directly to the making of the shoe. All expenses of this character are a burden upon production that must be covered by the sale price of the good, if the plant continues its output.

Selling Costs.—Up to this point costs have had to do with the processes of manufacturing, but when these processes have been completed, there are additional expenses in selling or marketing

the goods produced. In most lines the merchandising costs are relatively large. The selling expense in the retailing of men's clothing averages more than 20 per cent of the sale value.¹ Selling expenses vary between industries as well as between establishments in the same industry, but from the standpoint of scarcity and in the determination of value, these costs must be included.

Interest as a Cost.—All of the above-mentioned costs would clearly involve money outlays. There remains another cost concerning which there is frequently some confusion of thought, namely, the earnings on investment. The question is, Shall interest be regarded as a cost of production? There is some conflict of opinion between economists and accountants on this subject. The difference of opinion is to a large extent due to differences in the definition of terms, and in the purposes for which the classification is made. To many accountants, interest on investment that is owned by the proprietor of the business should not be reckoned as an item in the cost of production, but interest payments on borrowed capital should be counted as part of the cost. The accountants who dispose of interest in this way are likely to reckon interest on capital owned as a financial expense but not as a cost of production. The economist is viewing the problem from the standpoint of society rather than from that of an individual business concern, hence, he will include as a cost a return on all of the capital equipment necessary for the production of a good. To him it is immaterial whether the capital is borrowed or owned, for the investment must, in the long run, earn an amount sufficient to retain it in the industry. Interest, in this sense, is a compensation for the services of waiting, which we will later see is a necessary condition for the existence of capital. In this sense, all interest, whether on borrowed or owned capital, is a part of the cost that must be covered by the market price of the good.

If the market price is not high enough to pay all of the costs above mentioned, including a return sufficient to retain the capital invested, there will tend to be a readjustment of the

¹ SACRIST, H., "Costs, Merchandising Practices, etc.," Vol. III, p. 187. In this volume it is shown that the total selling expense per \$100 net sales averaged for the three years 1914, 1918, 1919, \$22.69. These costs varied from a low of \$21.49 in 1919 to a high of \$24.56 in 1914. Professor Sacrist has given this subject a most complete and thorough analysis in the above publication of the Bureau of Research of Northwestern University.

productive equipment to the extent that the market price will cover all of these items. This difference in the treatment of interest as a cost is due mainly to a difference in the approach to the problem. The accountant is ordinarily concerned with this question as a matter of internal administration of a private business. If he regards interest on capital owned as a cost, he will have to charge his income account with an identical sum as earnings. Therefore, this item will appear on both sides of the ledger. He argues: Why set up this item on both sides of the ledger, when the effect will be shown in the profit and loss statement? But the economist is dealing with a slightly different problem. He figures against an economic good all of the charges that are required to bring the good into existence and unless the capital invested in a business yields an income it will not be available for use. To him, the cost of a good is all of the charges that are assignable to the good and react upon its scarcity. He will, therefore, include as a cost whatever payment is necessary to call forth the waiting requisite for the existence of the capital used in industry. In this sense, interest in this text will be regarded as a cost, and the subject will receive further treatment in the chapter on interest.

Rent as a Cost.—The relation of rent to cost presents a special problem that will require further consideration under the subject of rent, but at this point it may be said that entrepreneurs compete for land as they compete for labor or capital. To the individual, the payment for the use of land is a cost. Whether this outlay is a cost from the point of society is a more difficult problem which will be treated later in the chapter on Rent. However, it may be said here that land can be used for many different purposes, and in order to secure its use in a particular way the entrepreneur must either buy it at the capitalized value of the rental, or lease it for a sum that approximately measures the rental in its present use. In either case, there is a rental charge which, from the individual point of view, may be regarded as a cost in the production of the good. Later consideration will show that, from the social point of view, rent is not generally a price-determining item in the cost of producing a good.

Managerial Salary as a Cost.—Another item in cost is the return to the entrepreneur for the distinctive services which he renders in the processes of production. This return might very well be regarded as a form of labor costs, or wages of manage-

ment. In practice, the entrepreneur's compensation is more likely to be associated with what is called "profit." For the present it will be better to regard the compensation of the entrepreneur for services rendered as a labor cost that should be reckoned as a part of the cost to be covered by the price of the good. The amount chargeable for this item can be measured either by what the entrepreneur could get if he hired out to operate some similar plant, or by what it would cost to hire some one to operate his own plant as effectively as he is now running it. In case he were an exceptionally able manager and could get a larger product from the business than any one else, the additional product, over what the average manager could get, should be regarded as a surplus. It is of the nature of rent and has frequently been called "rent of ability." Such sums do not enter cost as a limiting force on supply.

Fixed Supply and Value.—In the preceding discussion, the nature of costs has been analyzed and the essential elements have been pointed out. It remains now to relate these costs to the determination of value. Cost can affect value only through supply. As has already been stated, supply may be either definitely fixed or it may be reproducible. If the supply of a good is fixed, we may ask, How will its value be determined? Under these circumstances neither psychic costs nor expenses of production can have any influence upon the value of the good. Costs in both forms have been incurred in the past and the goods are being evaluated in the present. The effort which a Michael Angelo put upon one of his paintings, or the expense he incurred for brushes, paints, and canvas can have no bearing upon the value of that painting in a modern market. We must conclude, then, that when a supply of a good is fixed, its value is determined by marginal utility, that is, by the estimate placed upon it by marginal users. What such a good will bring in the market will depend entirely upon its evaluation by the purchasers. The buyers' influence will be the controlling factor in the determination of its value.

This argument may be accepted as the cause of value whenever the supply of a good is absolutely fixed, as in the illustration given. But how significant is the condition of fixed supply in the marketing of goods? Is it confined to rare paintings, first editions, etc.? If so, this explanation of value is of minor significance. An examination of the conditions under which

goods are regularly sold, however, will show that fixed supply as an influence on value is much more significant than at first appears. Indeed, there are those who are inclined to say that whenever goods are actually being sold on the market, the supply is fixed; that the costs have been incurred in the past and hence they can have no influence in the determination of present value. Persons who argue thus are inclined to explain value solely by marginal utility. While we may admit that goods actually in the process of sale are fixed and, hence, are subject to the evaluation by the prospective purchasers, the possibility of reproducing the goods is also an important influence both upon the buyer and upon the seller. Hence, while recognizing the influence of fixed supply upon the determination of value, it is not possible to disregard costs, especially if the good has considerable durability.

There is still another possibility of fixed supply having an important bearing upon the explanation of value. In the production of all goods, a time element is involved which is greater in some cases than in others. If the processes are so continuous that active selling can be reflected back along these processes from the finished product to the raw materials, increased activity is likely to result in speeding up the flow of goods to the market. Under such conditions of production, the fixed-supply influence is not sufficient for an explanation of value. The consumer's influence under such conditions is no more important in the determination of value than that of the producer. But where the productive processes cannot be speeded up, as in the case of all agricultural products, then the condition of fixed supply becomes important. Between crops, the volume of wheat that can come upon the market cannot exceed the amount of the last crop, and this supply cannot be increased until the next crop is harvested.¹ The present market price can have no influence on the amount of the last crop and can influence the next crop only by stimulating the farmers to increase or decrease the acreage devoted to wheat. Whenever the supply of any good is secured under like conditions, the fixed-supply influence will be significant in determining its value. The value of wheat during the time

¹ This general statement needs to be qualified by the extent to which previous crops have been stored. This qualification, however, does not materially affect the application of the fixed-supply notion in the determination of value.

that a crop is being marketed, for instance, tends to fluctuate around a price point which measures the buyers' estimate of the importance of that supply, irrespective of the costs involved in its production. If this price does not cover the costs in growing the wheat, the farmers will tend to limit the acreage devoted to that crop the following year. In this way, demand and supply tend to establish an equilibrium that will cover the long-run cost involved in the production of a good.

In drawing this conclusion, the student must bear in mind the two conditions assumed. First, that the market is competitive and second, that all of the good is coming on to the market for sale. This conclusion is modified to some extent by the control that may be exercised in the selling of wheat and also through the practice of selling futures. For the present, we can say that in marketing agricultural products the price between crops tends to conform to the marginal utility principle. The fixed supply influence tends to be controlling and value to be determined by marginal utility to marginal users. Only future additions to the supply will be affected by costs of production.

Reproducible Supply.—The second condition under which goods are brought to the market is that of a reproducible supply. Three different conditions of a reproducible supply will require our attention. First, when the supply may be regarded as perfectly flexible, or when the costs of production remain constant. If the costs of producing a unit of a good remain constant, regardless of the supply offered, then the value in the long run will be determined by the costs of production. While the market price of such a good may depart for a short time from its costs and be determined by marginal utility, this condition cannot prevail for any long period, because, according to the assumption, if the price goes above costs, additional units can be offered at the same figures and the operation of the principle of diminishing utility will bring the price of the additional units back to costs. If price, for any reason, falls temporarily below costs, it will be possible to check production and price will come back again to costs. Thus, demand and supply will tend to be in equilibrium at the costs of production when the costs are constant for all units produced. Under these circumstances, or so far as constant costs represent an actual condition of production, we may conclude that value in the long run is determined by costs of production. Marginal utility, or demand, determines

the amount offered on the market. All buyers willing to pay the costs of producing the good can find a supply to satisfy their desires.

The determination of value under constant costs can be illustrated by Fig. 15. The cost is here represented along the line OY which in this case is OS . Since the cost is constant the supply curve is represented by a straight line parallel with the line OX , or SS' . The demand curve DD' intersects the supply curve at F . At this point the supply OB just equals the amount demanded at the price BF which by assumption is equal to OS . If the supply offered were OA , competition among buyers would tend to fix the price at AE . Should this price be established, producers would tend to expand their supply until equilibrium was reached at the price point BF , which just equals the cost of producing the

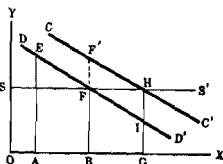


FIG. 15.

good. If for any reason supply were increased to OG , then price would fall to GI , which is less than costs. Producers would reduce the supply and equilibrium would again be restored to BF . If demand in the schedule sense increased, as for instance a change in the intensity of desire for the good, then the effect would be as shown in the demand curve CC' . The equilibrium would be established in the long run at GH , which, it will be observed, is just equal to BF . From this discussion, it will be apparent that the market influences tend to fix value at the costs of producing the good when the costs are constant.

No industry conforms in all of its processes to the assumption of constant costs, but there are conditions that approximate it. In the handicrafts or domestic industries the differences in unit cost are so small that an additional supply of goods can be furnished at approximately the same cost. Industries of this character approximate a constant cost condition but they do not bulk large in total production. In most industries costs are unequal, even if in some the effect of competition is to work in the direction of uniform costs by weeding out the high-cost producers. The second and third conditions of costs in relation to the determination of value will be considered under the topics of increasing and decreasing costs, which are generally

recognized as more nearly representative of the conditions under which goods are normally produced for the market.

Increasing Costs.—Second, we will now take up for discussion the case of supply that can be increased only as unit costs are increased. In these circumstances, both demand and supply influences are at work in determining value or, in other words, the influence of both the buyers and the sellers is important in the determination of value. This situation can be illustrated by Fig. 16. Let it be assumed that as supply is increased along OX , the utility to marginal users falls, and that additional amounts of the good can be disposed of only at falling prices. This situation is represented by the demand curve DD' . Let

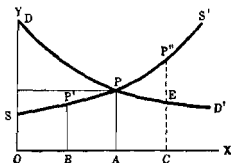


FIG. 16.

it be further assumed that the increased supply can be secured only under conditions of increasing unit costs, and that this condition is represented by the inclination of the supply curve SS' . It will be observed that supply and demand will come into equilibrium at the price AP , which means that at that price marginal purchasers are willing to pay a sum for the good that will just cover the costs of marginal producers. In a competitive market, this is the only price that would balance production and sales. Purchasers would be willing to take additional units of the good only at a lower price, say CE , but this additional amount would force marginal costs up to CP'' . As there would be no buyers at this figure, the supply would fall off to OA , at which point the price offered by marginal purchasers would just balance the costs to marginal producers. Under the circumstances assumed, AP would tend to be the competitive price. Both utility and costs are mutually operating to determine the value of a good so produced. Value, in the long run, tends to coincide with the costs of the marginal portions of the supply, for at this point marginal utility just balances marginal costs. Market price may depart temporarily from this sum, but will tend to revert to it because at no other figure will the amount produced just balance the amount demanded.

It will be observed from Fig. 16, that costs differ among the different producers. Some producers are more favorably situated

in respect to costs than others, as is shown by the distance that the curve SS' is above the base line OX . Those producers who furnish that portion of the supply that is represented by OB have lower costs on a portion of their supply than those who furnish the marginal supply at A . Between those with lowest costs at O and those with highest cost at A , the expenses of other producers are fairly evenly distributed.¹

These differences in costs may be temporary or permanent in character and their effect on value is quite different. Temporary differences, due to better plant, equipment, or organization, tend to disappear. If one man introduces an improved method of producing a good which gives him an advantage in costs, the competitive forces will bring pressure on his rivals to install this or some other method that is equally as effective. If they do not introduce more economical methods, they will find their customers deserting them for the producer who can furnish the good at a lower price. The determination of value, where there are non-permanent differences in costs of this character, tends to follow the principle of constant costs. The equilibrium price point will be costs, but in this case it tends to be the costs of the better equipped plants, *i.e.*, those which in the long run have the lowest costs.²

If the differences in costs are permanent, then the equilibrium point will coincide with the expenses of the marginal increment of the supply. This marginal increment may come from the least efficient or marginal producer, or it may come from the least effective application of the factors of production by the more efficient producers. In Fig. 16 the differences in costs were noted. It might very well be asked, Why do not those who have low costs, *i.e.*, those represented near O in Fig. 16, furnish the whole of the supply, thus eliminating the marginal producers at A ? This is what would take place if the differences were non-permanent in character; but if the differences in expenses are permanent, then any attempt to expand production with the view to absorbing a larger portion of the supply is checked by the operation of the principle of diminishing productivity. A producer at O may have a permanent advantage over one at A for a portion of

¹ Secrist has shown for one branch of merchandising that the selling expenses for different establishments are distributed fairly regularly from low to high cost firms. Cf. *Bull.* No. 8, Sec. II, pp. 8-10, of the Bureau of Business Research, Northwestern University.

² TUGGLES, F.W., "Principles," Vol. I, pp. 180-181.

his supply, but if he tries to expand his production, he will find that his unit costs will rise. In fact, the cost of the marginal increments of his supply will tend to be equal to the costs of the producer at A. Whether the marginal increment of the supply comes from either the extensive or intensive margin of production, the effect on value is the same. In either case, value will tend to coincide with the costs of the most expensive portion of the supply.

The Representative Firm.—In the treatment of this subject, it has been customary to identify marginal expense with the expenses of the marginal producer, and to say that value tends to coincide with the expenses of that producer who is producing at the greatest disadvantage. It has become evident, however, that in every market there are some producers whose costs are so high that they cannot continue to operate on the existing basis and, unless they can reduce their costs or unless prices rise, they will eventually be eliminated from the market. When it became generally recognized that some part of the supply may come from firms whose costs are higher than those that can permanently prevail, the idea of a *representative firm* was introduced as a means of explaining the determination of value. It was argued that instead of the costs of the firm producing at the greatest disadvantage, it was those of the representative firm, *i.e.*, the firm that could persist and continue to produce for the market under the given circumstances, that really determined value from the side of supply. The highest cost firm was likely to be eliminated in the long-run competitive struggle, so that its costs were not controlling. Recent investigations of actual cost data show that the idea of a representative firm has to be modified.¹ An examination of the actual costs in the manufacture of certain products, as lumber, copper, iron, sugar, coal, etc., and also of the selling expense in men's retail clothing stores, has shown that the same firm does not necessarily occupy the same position in respect to costs from year to year. A firm may have high costs one year and the following year its costs may be lower. The most intensive study of this subject has been on the costs of selling men's clothing. Professor Secrist has shown that firms which have high costs one year tend to have lower costs the

¹ ROBBINS, LIONEL, in the *Economic Journal*, pp. 387-404, Sept., 1928 argues that the notion of the representative firm is not only superfluous but is also misleading.

following year, while those that have exceptionally low costs are usually forced out of their favorable situations.¹ In other words, competition tends to operate most vigorously upon the firms occupying the extremes in respect to cost, and to force them to move in the direction of average costs.

The bearing of the results of Professor Secrist's study of selling expenses in the men's clothing industry on the relation of costs to the determination of value is presented in Fig. 17. In this Figure,

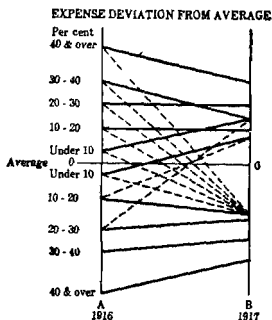


FIG. 17.

the experience of identical firms is shown for two separate years, 1916 and 1917. The expenses of 170 firms have been averaged for each year and this average constitutes the basis for comparing the relative positions in respect to the selling expenses of the firms studied. This average is represented in Fig. 17 by the horizontal line *OO*. The position of the individual firms, which has been determined by the percentage of their deviation from the average expense of the whole number of firms, is arrayed on the two vertical lines *A* and *B*. The position occupied by a firm in 1916 is located on the vertical line *A* and the position of the same firm in 1917, on line *B*. Thus, if Jones & Co. were able to do business at an expense of 10 per cent below the average, that fact would be shown by a position on the vertical lines indicating a position 10 per cent below the average expense of all firms.

¹ Secrist has no figures on firms that were eliminated, so that his information shows what happens to firms that remain in business.

Likewise, if the expenses of Brown & Co. were 30 per cent above the average, that fact would be shown by a position indicating a 30 per cent deviation above the average expense. The lines between *A* and *B* connect the positions of identical firms in the two years presented. The unbroken lines connect the positions of firms that remain either above or below the average for both years. The broken lines connect similar positions of identical firms whose yearly expense was below the average in one year but above the average the next year, or *vice versa*. In other words, these firms had changed positions in respect to the average expense of operation during the two years under consideration.

From this Figure, it will be seen that the position of a firm does not remain constant but tends in the direction of average costs. It is also apparent that the pressure of competition is greater on firms whose costs depart farthest from the average as these firms show a greater change in position than those whose expenses deviate less from the average. Secrist's study covered a period of five years (1916-1920), and the tendency shown in Fig. 17 for the two years here portrayed was perceptible for each of the years covered by his investigation. An analysis of the costs of identical firms revealed a constant pressure toward average costs, and also the fact that the identity of the firms having the highest costs did not remain the same from year to year.¹ The firm having the highest cost in 1916 might have a lower cost in 1917.

From this and previous studies in the manufacturing field, it now seems conclusively established that the marginal portion of the supply does not necessarily come from a marginal firm, in the sense of a firm that has from year to year the highest costs. The recognition of this fact has an important bearing on the statement of how value is determined, where there are permanent differences of costs of production. Instead of saying that value under these conditions tends to conform to the costs of the marginal firm, which may be true during a short period of time, we should say that value in the long run tends to coincide with the costs of producing the marginal increment of the supply. This statement of the principle removes the emphasis from the firm to that of the units of the supply, which, after all, is the important consideration in measuring the influence on value

¹ SECRIST, H., *op. cit.*, pp. 11-25. See also *Bull.* 9 in which he discusses the bearing of his data on the concept of the "representative firm."

from the side of supply. Instead of using the concept of a *representative firm* in explaining value, we should center attention upon a *cost condition* that tends to prevail, recognizing that value tends, in the long run, to coincide with the cost of producing the marginal increment of the supply, whether that increment comes from the least efficient or marginal producer, or from the least effective application of the factors of production by the more efficient producers.

Bulk Line Cost.—During the World War, as a result of the attempts to fix the prices of certain commodities, the expression "bulk line costs" came into use. It was generally conceded that if a price covered the expenses of the bulk of the supply, say 85 per cent, it constituted a fairly satisfactory basis on which the governmental policy of price fixing might proceed. Recent cost studies have thrown light on this problem. Professor

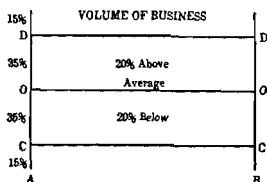


FIG. 18.

Secrist has shown for the merchandising of men's clothing that 70 per cent of the volume of business was done by firms whose expenses fell within a 20 per cent deviation above and below the average of the year. These facts are represented in Fig. 18. In this Figure, the vertical lines A and B represent the years 1916 and 1917. The horizontal line OO, represents the average cost of all firms, while the lines DD and CC represent a 20 per cent deviation above and below this average respectively. The percentages to the left of the vertical line A show the percentage of the volume of business transacted in each of the expense areas, that are represented to the right of the vertical line A. It is clear from these figures that 85 per cent of the business was transacted at an expense that did not deviate more than 20 per cent above the average. This position is represented by the line

DD in the Figure. The significant fact in this study is that this relation is fairly constant from year to year, so that we may safely conclude that in the merchandising of men's clothing 85 per cent of the volume of business is done at an expense that does not exceed a 20 per cent deviation above the average of all firms. The method used in this study seems to have established an effective plan for determining the bulk line costs. Whether the percentage found in these mercantile operations would hold for other industries or business enterprises is not known, but the presumption is that cost figures for other industries would doubtless show similar results if they were subjected to similar treatment.

But from the point of view of explaining value we are not concerned so much with costs at which a bulk of the supply can be secured, as we are with the expense that must be incurred in order to secure the marginal increment of the supply, as value, in the long run, tends to coincide with this expense. At this point, marginal utility or demand will just balance marginal costs or supply.

Having described the way value is determined under conditions of increasing costs, we may now inquire what industries conform to this condition. Agriculture, forestry, mining and, in fact, all extractive industries, show a tendency toward increasing unit costs. Improvements in the methods of production will show decreasing unit costs in these industries as in all others, but sooner or later these improvements will run into the permanent tendency of diminishing productivity which is characteristic of all extractive industries. Unless the improvements take place at a rate faster than the tendency to diminishing productivity, unit costs will increase. It can be shown that over a long period of time increasing productivity may be found in some of these industries and wherever it occurs, it causes lower unit costs. The cost of producing a bushel of wheat has doubtless decreased during the last hundred years as the result of the application of agricultural machinery and the introduction of new methods of cultivating the soil. These results have been obtained because scientific discovery and invention have more than kept pace with the tendency of the industry to show diminishing productivity and higher unit costs. If scientific methods do not keep constantly ahead of these normal tendencies, unit costs will increase.

Some of the extractive industries, like mines and forests, present special cases. In addition to the tendency toward increasing costs that characterizes the extractive industries, mining, lumbering and other similar industries have the added cost which results from the using up of the materials themselves. A coal mine, for instance, will show added unit costs as the coal has to be mined farther and farther from the mouth of the mine and, in addition, the stock of coal itself is gradually being depleted. There can be no restoring an exhausted coal mine, and even a depleted forest will take generations to restore. In agriculture the cultivation may continue in such a manner as to retain the fertility of the soil, but in mining, lumbering, and industries similarly situated from the point of view of production, operation means the using up of the valuable properties themselves. From the point of view of society, this problem appears as one of conservation of the natural resources.

The value principle applies to these industries, however, just as it does to all other extractive industries. The value of mineral, or other similar products, will tend to coincide with the costs of the most expensive portion of the supply.¹ But from the point of view of private management the depletion cost, or wasting of the valuable assets, creates a special problem. Unless care is exercised in the accounting methods and due recognition given to the depletion of the natural resource, a company may find that the basis for its whole valuable property rights has disappeared through an erroneous treatment of a portion of the return as earnings or profit, when in fact it is a part of its capital investment. The recognition of a depletion cost is significant from a social point of view in its relation to the general problem of conservation, and also from that of private management as a problem in determining what constitutes true earnings of the industry, but it has no bearing on the application of the value principle to the products of these industries. As stated above, the long-run value of these products tends to coincide with the expenses of the most expensive portion of the supply.

Decreasing Costs.—Third, supply may be secured under conditions of decreasing unit costs. This means that as an industry increases in size, its unit costs become lower and this makes for large-scale production. Our next problem is to inquire how the value of goods produced under these conditions is deter-

¹ Cf. Chap. XI for further treatment.

mined. The case of decreasing cost is illustrated by Fig. 19. In this Figure the amount produced is measured along the line OX and the cost along the line OY . The supply curve SS' indicates lower costs as the volume produced is increased. The unit cost of an amount OA would be AP' (OP), while the costs of an amount OB would be BE' (OE), which is less than AP' . With a demand equal to OA , the price will be AP' , for at this point supply and demand will be in equilibrium. But if demand in the market sense can be increased, the supply can be furnished at lower unit cost. In Fig. 19 the producers would be willing to furnish the quantity OA' at the unit cost $A'F$, which is less than

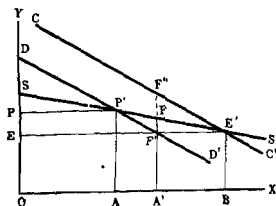


FIG. 19.

AP' . But the purchasers in this case would be unwilling to pay more than $A'F''$ for this volume of supply, hence, the market conditions as here represented would not warrant an expansion of production. If the decline in unit costs had kept pace with the diminishing utility of the added supply, a new equilibrium point would have been established at $A'F''$. If we assume that the demand has increased in the schedule sense, the purchasers would be willing to pay $A'F''$ for an amount OA' . Since the producers can furnish this supply at a cost of $A'F$, which is less than $A'F''$, there would be a tendency to expand the volume of output to B . At this point there would be an equilibrium between supply and demand at the price of BE' .

In this condition of production, supply and demand are both operating to determine value and competitive forces tend to fix long-run value at the costs of the least expensive portion of the supply. At any given time, in an industry producing at decreasing costs, there will be varying costs. Some plants will have installed superior methods of production, so that they will

have an advantage in their production costs. Under these circumstances, value will temporarily coincide with the expenses of the least efficient plants, as demand and supply will come into equilibrium at marginal costs. This situation, however, cannot continue indefinitely, for those producers who enjoy an advantage in their costs will be tempted to enlarge their plants, or the high-cost producers will be forced to install more efficient methods that will bring their costs down to those of the best equipped and best managed plants. In other words, there will be a tendency toward uniform costs in all plants—a condition that would prevail if competition worked perfectly. While, as we have seen, there are obstacles to the perfect operation of competition, and while some producers will install improved methods sooner than others, there is, nevertheless, a constant tendency for costs to move in the direction of uniform expenses for all producers in those industries that are operating under decreasing costs. In such industries both demand and supply are operating in the determination of value. While marginal utility may temporarily come into equilibrium with marginal costs to determine market price, in the long run, value, or the equilibrium price point, will tend to be determined by the costs of production at the plants most advantageously situated as regard costs.

In following this analysis, it should be apparent to the reader that an industry showing a tendency toward decreasing unit costs moves in the direction of uniform costs. In tracing the effect of this influence, distinction should be made between internal and external economies of production. An industry may show continuously lower unit costs due to improved methods of production lying outside of the industry itself, as when the cost of producing the machinery used in its operation is lowered, or when an invention or other change in the processes of production lower the costs of the raw materials used. These are called "external economies." External economies may be utilized by all establishments alike and seem to show some tendency toward constant cost.

But of more significance is the fact of "internal economies." What has just been said in connection with invention and discovery as applied to outside industries applies with equal force within any single industry. In addition, all the advantages that flow from large-scale production appear as internal economies. Larger plants, more economical methods of handling

materials and of manufacturing the products, economies in freight due to larger shipments, more effective division of labor, and, in fact, whatever methods lower unit costs as a result of large-scale production, come under the term "internal economies."¹ If internal economies due to increasing size of plant were to continue indefinitely, the long-run result would be complete monopoly, or control of production by one single management. It is an interesting fact that the tendency toward industrial monopolies has been to a large extent in industries that have shown internal economies from large-scale production.

The size of the producing concern may increase in either of two directions. There may be "horizontal combination," where plants of approximately the same size and producing the same type of goods are brought under one single management, or there may be "vertical combination," where the various processes of production, from the raw materials to the finished products, are brought under a single management.² Internal economies will result from each of these methods of combination, but there is increasing doubt as to how far the size of plant can be carried economically. The problem of supervision and direction becomes increasingly difficult. The farther the actual performance of a productive operation gets from the personal force of the executive, or the driving force in industry, the less is the efficiency. The possibility of loss due to the lack of good will, or to the inertia of human nature when it is not vitalized by a dominant and driving interest in the business, is tremendous.

Then extremely large business concerns find that it is difficult to adapt their methods when industrial conditions make such changes desirable. There is a larger degree of inertia in a big plant than in a small one. For instance, during the depression in 1921 it was said that the smaller plants in the packing industry were able to adjust themselves to the situation more readily than the larger concerns. Most of the larger plants had financial adjustments to make that were embarrassing. Armour and Company, and Wilson & Co. Inc. are well-known illustrations of financial embarrassment at that time. Adjustments to the various phases of the business cycle seem to be more difficult for the large plant to make than for the small one. The significance of this observation is the bearing it has on the application of the

¹ TAUSIG, F. W., "Principles," Vol. I, pp. 190-191.

² Ibid., "Principles," Vol. I, pp. 59-64.

principle of diminishing returns to managerial ability. It seems a fact that in certain types of industries, the size of the plant can become too large for effective operation and, hence, places a limit upon the scale of production. So far as the size of plant retards efficiency, it acts as a very positive limitation on the tendency toward the development of monopoly. We can conclude from this discussion, that so long as these economies appear, there will be a tendency for large-scale production to develop, but since there seems to be a point beyond which large-scale management shows diminishing productivity, there will arise at that point a positive check to the development of monopoly as the result of this cause.

From the preceding discussion, the conclusion has been drawn that, in the long run, the market price of reproducible goods sold in a competitive market tends to coincide with the expenses of producing them. While many instances can be found where existing market prices depart from costs of production, yet there is plenty of evidence to show that competitive market prices constantly approach costs. So generally has the public mind accepted this conclusion, namely, that competitive conditions will keep market price close to the expenses involved in producing a good, that any material departure from it is regarded as an unfair price. During the World War the term "profiteer" was freely used. The only real basis for the epithet were the instances where sellers took advantage of an unusual situation and sold at prices fixed by demand quite irrespective of cost. The wide-spread condemnation of any such practice is evidence that the general public has accepted as a standard of fair value, the market price that tends to coincide with the costs of producing the good.

Normal Value.—Up to this point in the present chapter and in the one that preceded it, we have been concerned with the principles that determine market value. At several points reference has been made to the long-run effect of competitive forces. It is now our purpose to analyze more in detail the time element in the determination of value. Market price is always a present price and is always affected by immediate influences.¹ The price of wheat today is affected by the volume offered, and the urgency

¹ Prices on future contracts are apparent exceptions, but as a matter of fact even these prices are present prices. They are offers in the present for future delivery of goods.

of the demand. But there are permanent influences that lie back of the immediate situation, which will eventually control the movement of market price. These permanent influences tend constantly to establish a price that will clear the market of the total supply. This long-run price, the one that takes the time element into account and allows sufficient time for the competitive forces to work out their full effect, is known as "normal value." Normal value is that value that tends to prevail during the period in which the whole supply of a good is being marketed.

Temporary fluctuations in demand and supply will cause market price to deviate from normal value, but the longer the time element involved, the more likely will market price conform to it. Normal value is the result of the permanent influences at work in the market and it may be said that it never exists in fact. In this regard, it is something like the level of the ocean, which is constantly disturbed by waves of varying heights. The water is constantly in motion up and down from a level which never exists but which the forces of nature are constantly endeavoring to establish. By analogy we may think of the forces operating in a competitive market, constantly pulling market prices (the crests and troughs of the waves) toward a normal level. This is the concept of normal value.

These long-run influences are affected by the conditions of production already referred to, namely, industries producing under conditions of constant, increasing, or decreasing costs. If an industry is operated under constant costs, we can say that the long-run influence of competition will be for normal value to coincide with these costs and for market price, due to the short-run influences of demand and supply, to fluctuate around them. But if the industry shows increasing costs, then normal value will tend to coincide with these costs, market value fluctuating around them. In an industry producing under decreasing costs, normal value will coincide with the declining costs, while market price will fluctuate around them. These three conditions can be illustrated by the following diagrams: In Fig. 20, *PP'* represents normal value and shows the full effects of the competitive forces where costs are constant. *MM'* represents the fluctuations of market price around the permanent price or normal value, as established in these conditions. It will be apparent that market price can not deviate far from the expenses of production but tends constantly toward it.

Figures 21 and 22 show the same market influences in the case of increasing and decreasing costs. It will be noted that, in the one case, there are permanent influences at work to make normal value coincide with increasing costs, while market prices

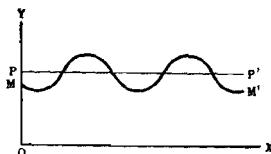


FIG. 20.

fluctuate around this norm. In the other, the situation as to the direction of the cost movement is reversed. Here normal value tends to coincide with the diminishing costs, market prices departing in the same general manner as shown in the preceding figures.

Joint Costs.—In the treatment of value thus far, it has been assumed that every good has a separate cost and that the long-run influences tend to bring about a market balance between supply and demand at the point of marginal costs. A little observation and reflection will make it plain that some productive processes result in two or more products. For instance,



FIG. 21.

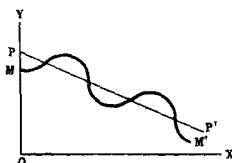


FIG. 22.

from one process we get beef and hides, from another, cotton fiber and cotton seed. Many other similar examples of joint costs and joint products are easily available. In such cases, it is impossible to separate the cost that is chargeable to any single product. Take the packing industry as an example. All parts of the animal are used. Innumerable articles are made,

such as meats, hides, athletic goods, pharmaceutical supplies, fertilizers, etc., all products from the original operation of slaughtering animals. Who can say what part of that first cost is chargeable to meats, what part to leather goods, or to any other of the commodities that are joint products in this industry? Any division of this cost is purely arbitrary because the one cost gives rise to two or more commodities. This condition of production creates a special problem that is of much significance, viz., the determination of value where there are "joint costs."

In a case of this kind the price of each product is determined by its marginal utility, or, in other words, demand for the product is controlling, and the normal value of the whole group of products must equal the long-run costs of producing them. To illustrate the operation of this principle, take the case of cotton fiber and cotton seed, as they constitute a simple example that shows the perfect operation of the joint-cost principle. Cotton fiber cannot be used until the seed is removed and the one operation gives both products. The value of cotton fiber will be determined by the marginal utility of the existing supply. The estimate placed upon cotton fiber by marginal buyers will tend to determine its price. If the crop is small, the value per unit will tend to rise and *vice versa*, assuming, of course, no change in the general desire for cotton fabrics. The same principle holds in the determination of the value of cotton seed, as its value will tend to equal its marginal utility to marginal buyers. Before new uses were found for the seed it was largely a waste product. Except for purposes of planting, there was little or no demand for the bulk of the seed, hence its utility was very small. As a result, most of the joint costs had to be charged against the main product, the cotton fiber. Under these circumstances we may conclude that the value of any joint product is determined by its marginal utility, and that the combined value of all joint products must equal the joint expenses of producing them.

It has been found that by further treating cotton seed, additional products can be made. Cotton-seed oil has become an important article and cotton-seed cake is now used for feed for cattle and hogs. The finding of these new uses for the seed has enhanced its value by increasing its marginal utility which has made it possible to assign a larger portion of the joint expenses to the seed, thus enhancing the return from the joint products.

To get these additional products from the seed gives rise to a new example of joint costs. To get cotton-seed oil, for instance, an additional expense had to be incurred because of the further operation necessary to secure the additional products. The combined value of cotton-seed oil and cotton-seed cake must be equal to the cost involved in separating the oil from the seed, or the task would not be undertaken. Wherever further processing is required to make one or more of a group of joint products usable, such costs are known as "separable expenses" and the additional processes will not be undertaken unless the utility of the resulting products is equal to the separable expenses necessary to bring them into existence.¹ If these expenses are not covered by the price offers, the additional processes will not be undertaken.

This question raises an interesting problem. How will an increase in the demand for one of the products affect the value of the other? Suppose an increase in the demand for cotton-seed oil. If this increased demand does not result in an increased supply of cotton fiber, the value of the fiber will not be affected, but the combined value of cotton fiber and cotton seed will exceed the joint expense and will thus yield larger earnings. This fact will tend to stimulate additional production because the returns to the cotton growers are now greater than previously. If the former return was adequate to attract capital and labor, the new situation will make the industry more profitable.

Under these circumstances, total production will tend to be stimulated by the increase in the demand for cotton-seed oil, but because of the larger cotton crop grown, the price of cotton fiber will fall. There will be an increased supply, but no increase in demand. Under such conditions, it is customary to shift a part of the joint expense to the joint product that has enjoyed the benefit of an increased demand. This dividing of the cost between the joint products is purely an arbitrary process that rests upon the idea that each joint product should bear a part of the joint expense in proportion to its relative value.

There are many industries in which the joint-cost principle operates, and its significance is especially great where a large fixed capital is used for several purposes, as in the case of railroads. Here, the same roadbed and right of way serve both the passenger and the freight business and, in the case of freight,

¹ TAUSSEIG, F. W., "Principles," Vol. I, pp. 214-218.

there is a vast variety of commodities carried. The division of costs in this industry (and others similarly situated) is an exceedingly complicated problem, but it follows the joint-cost principle. The value of the combined products or services must cover the joint expense. Any single product or service must equal the *separable expenses* involved in its production, but its value depends upon the marginal utility of the good.

A counterpart of joint costs is *joint demand*. There are some goods that are useful only in connection with one or more commodities, as in the case ink and paper. A demand for a dwelling house creates a demand for all of the materials used in its erection. It also creates a demand for the various kinds of labor used in the construction of houses. A joint demand affects the value of the several products differently, for the supply of the different products does not remain perfectly balanced. Some products or services may be easily increased, while others will take a longer time, as in the case of common labor and a skilled carpenter. It would be easy to find additional laborers who could do the pick-and-shovel work in building construction, but it is much more difficult to get skilled carpenters and bricklayers. The value of the product or service which responds slowly to an increase in demand will rise faster and to greater heights than that of commodities which can be produced or increased more readily. Lathers and plasterers have recently enjoyed extraordinary increases in wages because of the increase in the demand for buildings. The supply of this type of labor was scarce, due partly to union regulations limiting those entering the trade and partly to the unusual activity in the building industry. Hence, the value of the services of these men increased faster than that in the other trades.

Joint demand, in the long run, has a different effect on the value of the commodities affected by it than does joint supply. If the value of any commodity is increased as the result of joint demand, the conditions of supply, barring artificial restrictions, will be adjusted to the new conditions of demand. If the demand for buildings increases, more building materials will be produced, more men will enter the building trades, and a new balance will be established. The increase in value of any good from this cause is likely to be temporary. But in the case of joint supply, the effect of a change is likely to be much more permanent. An increase in the demand for cotton fiber, for instance, will tend to lower the

price of cotton seed permanently. This difference between the two cases should be carefully noted.

Speculation and Value.—We have seen that production is carried on primarily in anticipation of demand. It proceeds upon estimates of demand that are made by entrepreneurs or business executives. If these estimates prove to be accurate, there will be a balance between demand and supply, but there is always a large amount of uncertainty involved in these estimates and this uncertainty is one of the important causes of risk in industry. Speculation as a means of carrying risk has been treated in Chap. VI. It now remains to consider its effect upon market value, as uncertainty involves the price at which the producer can dispose of his goods.

Speculation consists in forecasting future prices and in buying and selling contracts to deliver some commodity on the basis of these estimates of prices. The simpler effects of speculation on market value of some commodity like wheat can be illustrated from the operations on the Board of Trade. In our discussions up to this point, we have treated demand and supply as present influences only, but in actual market transactions future conditions of both demand and supply react on prices. If the whole season's supply of wheat were dumped on the market immediately after the harvest, it would depress the price and stimulate uneconomical uses, which would enhance the price three to six months later. The traders on the Board, the "bulls" and the "bears" already described, are concerned not only with present demand and supply conditions, but they discount probable future movements of these two forces. If crop conditions are unfavorable and the prospects indicate a smaller yield, these facts will be taken into account several months before the crop is harvested. Higher prices for future delivery will stimulate the withholding of larger amounts of the existing supply from the present market, thus causing a rise in the price of wheat in the cash market. In this way, speculative buying of the kind described tends to equalize through a period of time the price of any commodity that is sold in a market similar to that of wheat.

In the hypothetical case cited, the price of wheat in the present or cash market is raised, while the price at a future date is lowered because of the additional volume that is held over from the previous supply. Speculative dealings of this character tend to establish a price that measures the seasonal estimate of

the significance of the whole supply, and tend to equalize present and future prices at a point that will cause an equilibrium of demand and supply during the period of time required to market the good. Normally, the future price will exceed the present price by an amount sufficient to cover all of the costs involved in carrying the product from the present to the date of delivery.

So far as speculative dealings consist of the simpler functions herein described, a very important social service is performed. The effect of the operation of an organized market is to give more stabilized price conditions than would exist if there were no means of discounting future conditions of supply and demand. But it would be incorrect to leave the impression that all transactions carried on under the guise of speculation result in a social service of this character. There may be attempts to manipulate a market and efforts may be made to *corner* the supply of some commodity with a view to securing unusually large private profits. However, contrary to popular impression, the amount of manipulation of this character that has evil social consequences is relatively small on the grain and produce exchanges. On these exchanges, the speculator's estimates must be fundamentally correct if large profits are realized, which means that the prices would have moved in the direction of his estimates, whether he had or had not undertaken to secure control of the available supply. No speculator can gain consistently unless his estimates are fundamentally correct.

There may also be gambling in connection with any organized exchange. The difference between gambling and speculation is that in the former there is no equivalence of service rendered, while in the latter there is a *quid pro quo*. In the former, what one man gains, the other loses; in the latter, the payments are for a service performed. A contract that partakes of the nature of gambling may be illustrated as follows: Suppose A makes a contract with B to deliver at some future date 1,000 bushels of wheat at \$1 per bushel, and at the maturity of the contract, if the price has risen, instead of delivering the wheat to B, A pays him the difference in the price of wheat prevailing at that time and at the date when the contract was negotiated. A contract of this kind could be made and completed without any movement of wheat. So long as such contracts are purely private matters as between A and B, they can have no influence on general price movements; but when contracts of this character

are handled through an organized exchange and run into millions of dollars the result may be an artificial stimulation of market conditions.

The speculative operations on the stock exchange are more likely to be of the nature of gambling than are those on the Boards of Trade or produce exchanges. The stock exchanges constitute an organized market for the sale and distribution of the listed securities. By means of this market, the securities of established industries find a ready sale. But the purchase and sale of securities, especially by means of margins, are peculiarly adapted to gambling operations. The gains and losses are very largely private matters and have less social significance than is sometimes thought. In the words of Taussig:

The economic loss arises primarily from the waste of much brains and energy on unproductive doings. The waste is more than that of the labor given directly, the labor of the brokers and their under-strappers and of the speculators themselves. It is increased by the demoralization of many men in the community who take no great direct share in speculation. Like all gambling, it distracts from the sober, continuous work on which the common welfare rests.¹

Both the boards of trade and the stock exchanges perform important functions as agencies for the marketing of the respective goods regularly offered for sale. In this capacity, they both constitute a part of the machinery by which the price of the various goods sold is established, and as agencies by which demand and supply are equalized. There are, however, possibilities of grave social evil in the operations of each of them. While there is a tendency for the so-called gambling contracts to nullify each other and, hence, to have small evil social consequences, the impression that speculative dealings are responsible for the widespread and recurring price depressions is so generally accepted that it behooves both of these agencies to watch their methods of operation and to eliminate so far as possible those forms of dealing that give a semblance of truth to the charge of manipulation. It is probably impossible to eliminate the gambling element entirely from speculative buying, but the exchanges can and have done a great deal, by means of rules and regulations governing the making and the fulfilment of contracts, to eliminate the worst features in the operation of these markets. As a means of correcting the evils, there is probably more to be

¹ TAUSSIG, F. W., "Principles," Vol. 1, p. 165.

hoped from regulations of this character, brought about, it may be, by the pressure of public opinion than from any attempts to control the exchanges by means of legislation.

Suggested References

- BYE, RAYMOND T., "Principles of Economics," Chap. XVI.
CLAY, HENRY, "Economics for the General Reader," Chap. XIV.
EDIE, LIONEL D., "Economics, Principles and Problems," Chap. IX.
ELY, RICHARD T., "Outlines of Economics," 4th Ed., Chap. XI.
FAIRCHILD, FURNISS, and BUCK, "Elementary Economics," Vol. I, Chap. XVII.
FISHER, IRVING, "Elementary Principles of Economics," Chaps. XVII and XVIII.
HAYES, H. GORDON, "Our Economic System," Vol. I, Chaps. XXI and XXII.
MARSHALL, ALFRED, "Principles of Economics," Book V, Chaps. I-V, and XIV.
RUPENER, L. A., "Principles of Economics," Chaps. VIII, IX, XI, XIII, and XIV.
SELIGMAN, EDWIN R. A., "Principles of Economics," Chap. XVI.
TAUSSIG, FRANK W., "Principles of Economics," 3d Ed., Chaps. XII-XIV.
TATLOR, FREDERICK M., "Principles of Economics," Chaps. XXI-XXXI.

CHAPTER X

MONOPOLY AND VALUE

Summary.—In the presentation of the theory of value up to this point, we have assumed a perfect market in which the forces of competition and mobility worked perfectly, *i.e.*, where the factors of production are promptly adjusted to the uses which yield the highest return or income. As is well known, actual conditions do not, and in fact, cannot, conform perfectly to this assumption. But while these forces do not work just as assumed, no one will deny the presence of competition and mobility in every market. So far as these forces are present, we may conclude that their effect is to cause value to be determined in accordance with the principles set forth in the preceding chapters.

However, in most markets there are other forces that have an influence in determining value. It will be our purpose now to examine some of the more important of these forces and to show their effect upon the determination of value. There are two types of obstacles to the perfect operation of competitive forces, namely, natural obstacles, or what has sometimes been called economic friction, and artificial obstacles, or conscious control of the productive forces. As illustrative of the former, we may cite the lack of information among both buyers and sellers concerning the actual market conditions in respect to both demand and supply, and the inability to adjust the factors of production to changed market conditions as rapidly as the changes in demand and supply occur. A little reflection will convince the reader that land, labor, and capital cannot be shifted with perfect ease from one industry to another. Any such obstacles prevent the perfect working of competition and tend to create a lag between causes and results.

To illustrate concretely, take the case of an invention that greatly reduces the costs of producing some good. There are many influences that prevent the immediate use of this invention by all producers. Even barring the possibility of a patent, which would give the patentee exclusive control over the use of the

invention, some producers will install the new method sooner than others. Some men have more initiative than others and on this account are more inclined to introduce new methods of production. Or, it may be that the machinery in one plant is more nearly worn out than in another, hence, in replacing the old equipment, the manager will take advantage of the latest and most efficient devices on the market at that time. There are a great variety of causes that may retard the perfect operation of the competitive forces, but enough has been said by way of illustration to indicate how these obstacles operate.

The result of these conditions is that the costs of different increments of the supply in the market will vary. But these differences in costs may not persist, for those having an advantage in their costs will tend in time to expand their production, or those having high costs will install improved methods, so that, in the long run, the competitive forces will cause value to be determined by a balance between marginal utility or demand influences and marginal costs or supply influences. The only modification of the conclusions based upon the assumption of a perfect market is, in this case, that variations in value will be greater as the obstacles are more permanent and persistent in character.

In some cases, the lag in the adjustment of the factors of production to changed market conditions will be greater than in others. Under such circumstances, with varying conditions of production and a tendency for different prices to exist for identical articles, there will be some who occupy unequal competitive positions. Some producers will have lower production costs than others, and some buyers may not be fully informed as to the different prices at which the same good is being offered for sale. This inequality will persist as long as these conditions exist. Nevertheless, there will be constant pressure both on the side of demand and of supply to equalize the market conditions. The more advantageously situated producers will, by advertising or by other methods, endeavor to secure a larger volume of sales. In order to maintain their own business, the other producers will strive to bring their costs down as a means of meeting the competitive prices of their rivals. Demand and supply will tend in this way to become equal at some price point. The *pull* of the competitive forces is in the direction of establishing a balance between marginal utility and cost in one of the three forms in which they have previously been discussed. While

obstacles of the kind indicated may cause a departure from this balance for a longer or shorter period, nevertheless, competition is constantly pressing to establish market value at the point that will equalize demand and supply. In the long run, obstacles of the kind described tend to give way to the competitive forces, so that the conclusions based on the conception of a competitive market are, in general, correct.

Monopoly.—In addition to the natural obstacles just described, there are artificial or conscious forces that greatly affect the actual determination of market value. The consciously directed forces operating in the market are commonly known as “monopoly.” Monopoly may be more precisely defined as that control exercised by one or more persons over the supply of some economic good that will enable them to fix the price of that good. Usually monopoly power is exercised through control over supply. By limiting the supply of the good offered for sale, the monopolist has large power to regulate the price to suit his own economic advantage. It should be noted that mere scarcity does not imply monopoly, for it has already been shown that only goods which are limited in supply in relation to human desires have economic consequence. In addition to relative scarcity, which is an attribute of all economic goods, there must be present a considerable degree of conscious control before it can be said that monopoly exists. If this control is exclusive, the monopoly is absolute. But since conscious control is not generally exclusive, monopoly power is more commonly partial in its extent.

Monopoly power can be exercised by virtue of the control which emanates from the rights of private property. The owner of any valuable goods has, within very broad limits, exclusive control over them. He can use them, sell them, or even destroy them, unless by so doing he endangers valuable goods of others. In other words, his control has to be exercised within the legal framework that governs all property rights. Within these limits, the owner of property may do with it as he sees fit. If a greater return can be secured by withholding a portion of the supply, or by destroying a part of it, or by selling it in different markets at different prices, the monopolist is in a more strategic position to take advantage of these circumstances than if the supply were not so controlled. The monopolist is likely to exercise his control in such manner as will give him a maximum net return from the disposal of his goods. This statement should not

be accepted too literally for most men are affected by other than strictly economic motives. What needs to be realized is that where monopoly exists, the power to dispose of goods in accordance with the generalization stated is present. How far that power will be exercised will depend upon a variety of influences but, in the main, it may be said to depend upon the extent to which the monopolist is dominated in his conduct by other than motives of private gain.

Historically, monopolies were at first direct grants of power by the king to favored subjects. These grants gave the favored individuals certain exclusive privileges or rights, such as those that were granted to the London and Plymouth companies, with which all students of the early attempts at colonization and settlement of our country are familiar. The idea of exclusive right or control has continued as the central idea of monopoly, even though the source of monopoly power has changed with the course of time. It is because of the special privileges exercised by favored individuals that monopolies have always been under the ban of public opinion.

In the definition of monopoly given, emphasis was placed upon the control of supply. This is, in fact, the way in which monopoly power is most commonly exercised, but control over demand may at times be sufficient to give to the buyers price-fixing power. So far as buyers exercise the power to fix prices, we may say that there is a buyers' monopoly. As an economic force, however, monopoly power exercised by buyers is much less significant than the exclusive control over supply.

Kinds of Monopolies.—As was indicated above, public opinion is extremely hostile to anything that has the appearance of monopoly power, and the conception of what constitutes the limits between monopoly and large-scale production is not always clearly drawn. Because an industry has grown to enormous proportions is not convincing evidence that it is a monopoly. A department store in a large city may transact an enormous volume of business and yet may be operating in the face of the most active competition with other stores of similar size, or with other methods of merchandising. Marshall Field & Company of Chicago, and Sears, Roebuck & Company are firms that have an extremely large volume of sales, yet both have to meet the competition of other similar companies and other methods of marketing goods. Because of the confusion

of thought concerning the nature of monopoly, and also because monopoly appears in so many different ways in our economic life, a classification of monopolies may add to the clarity of thought concerning them. The source of the monopoly power is used as the basis of the following classification because any attempts at regulation and control should be preceded by a careful study of the cause or source of the monopoly power in order that the remedy proposed may be fitted to the peculiarities of the case.

I. Social Monopolies:

A. General welfare monopolies.

1. Patents.
2. Copyrights.
3. Public consumption monopolies.
4. Fiscal monopolies.

B. Special privilege monopolies.

1. Those based on public favoritism.
2. Those based on private favoritism.

II. Natural Monopolies:

A. Those arising from special limitation of the supply of raw materials.

B. Those arising from secrecy.

C. Those arising from peculiar properties inherent in the business.

Social Monopolies.—A monopoly may be called a social monopoly when it depends for its existence upon some legislative enactment, or upon some other grant of power recognized by society. All such monopolies rest upon conditions that are consciously created. Among these we may cite "patents" and "copyrights." Here, exclusive control is granted by legislative enactment for a limited period of time as a stimulus to invention and discovery. Exclusive control in these two forms is based on the assumption that by this means an appeal can be made to the self-interest of the individual as a means of stimulating him to make some discovery or to write something that will contribute to the welfare of mankind. It frequently happens in the case of patents that the monopoly power does not work just as it was contemplated. An individual who makes a discovery may sell his patent rights to those who are in a more favorable

¹ ELY, R. T., "Outlines of Economics," 4th Ed., pp. 184-185. This text has an excellent presentation of the nature of monopoly, and the classification here given is taken from this source, from which I have drawn heavily in the treatment of this topic. Cf. also FERRELL, F. A., "Principles," Chap. XXXIII.

position to make use of them, and the special privileges granted by the patent may be enjoyed by the new owners for a longer period than is necessary to call forth invention or discovery. Yet, in spite of many apparent injustices in the operation of the patent laws, the granting of such rights still receives public approval.

In addition to patents and copyrights, other forms of monopoly grants of the same general character may be mentioned, namely, *public consumption monopolies* and *fiscal monopolies*. If a government wishes to regulate the use of some commodity, as drugs or liquor, it may grant a monopoly privilege to sell it—a privilege which would constitute a public consumption monopoly, the primary purpose of which would be the regulation of the use of the good. A fiscal monopoly is one in which the purpose of the special privilege is to aid the raising of revenue for defraying public expenditures. In France, the sale of tobacco is a public monopoly run for fiscal purposes. These two purposes, *i.e.*, the raising of revenue and the regulation of an industry, are often combined and account for many of the special privileges granted by governments to favored individuals.

Closely associated with the above forms of monopoly power are those in which the special privilege enjoyed flows from what has been aptly called "public favoritism."¹ A protective tariff which gives to some producers in a country a special privilege over other producers is an illustration of monopoly power based upon public favoritism. A further example that may be classed here is that of municipal franchises which give the favored individuals exclusive privileges to operate municipal railways, or to perform some other function of like nature. While these two examples are not precisely alike, they both represent cases of special privileges that have been granted to favored persons by means of public enactment.

In some cases, the special privilege may arise from *private favoritism*, as when a railroad grants rebates or other special favors to some producers but does not make these privileges open to all shippers. Other forms of private favoritism may give rise to monopoly power, as where the price to one customer is distinctly lower than to others. An example of such a price advantage frequently occurs where one customer purchases larger quantities than others, and the difference in price is justi-

¹ ELY, R. T., "Outlines," *op. cit.*, p. 187.

fied on the grounds of difference in costs of supplying goods in large lots. If the difference in price just offsets the difference in costs, then no favoritism exists, but if the difference exceeds this sum, then a special privilege is granted. There are many obstacles of this character to the free exercise of competition in modern business. The granting of discounts from the sale price is frequently used to give some customers an advantage over others. All such obstacles to free competition are of the nature of monopoly based upon private favoritism.

Natural Monopoly.—A natural monopoly is one that depends for its existence upon natural forces, or upon a limitation of the supply by nature. In some instances, the supply of raw material is narrowly limited. The deposits of anthracite coal in this country are confined to a comparatively small area. This fact is an important contributing cause to the monopoly enjoyed by the owners of the anthracite mines. Some localities enjoy a monopoly because of the limitation of situation. The owners of city lots bordering a lake shore or public park may be said to enjoy a monopoly privilege due to limitation of supply. In this connection we might cite also the limitation of personal ability, such as that of a Caruso, a "Babe" Ruth, a Steinmetz, or any other individual possessing rare personal abilities who can render services that few others can perform. Such instances may well be thought of as natural monopolies. Many other instances of monopoly due to the limitation of supply by nature can be found.

In some cases, monopoly is due to secrecy, although this is not an important cause in modern industry. If a firm has a process of producing some article and finds that it is impossible to patent it or, if patented, that the method can be easily imitated, the firm may decide to rely upon secrecy as the most effective way of protecting its special privilege.¹ Exclusive control can arise from such a cause and should be regarded as a monopoly based on secrecy.

The largest source of monopoly power due to natural forces is found in the inherent nature of the business itself. Among such monopolies, we may cite streets, roads, canals, rivers,

¹ In visiting a large firm manufacturing pool and billiard tables and supplies, the author was told that he would be taken through all parts of the plant except where the billiard balls were made. No visitors were allowed in this department because of a secret process.

bridges, docks, harbors, railways, telegraph companies, public schools, post offices, and other public utilities, such as water, gas, electric lighting, and street railways. Often in the case of docks, harbors, railway terminals, or rights of way through mountain passes, etc., the monopoly is due to especially advantageous location that can not be duplicated, or, if capable of duplication, only at an expense that is prohibitive. In this respect such monopolies are akin to those depending upon the limitation of supply that were described above.

In most of the instances cited, however, there are conditions within the industry that tend to eliminate competition, the most important of which is the tendency for costs to diminish as the scale of production increases. So far as this condition is present, competition tends to work in the direction of a uniform cost and toward concentration of control, or monopoly. If the tendency of costs to decline as the scale of operation increases were to continue indefinitely, then all such industries would eventually become monopolies. There seem to be limits to this tendency, however, due to the difficulties in management. The principle of diminishing productivity seems to apply to managerial ability as well as to other factors of production.¹ What this means is that the size of an industry may become too large for the most effective results to be obtained. Of course, this tendency will vary as between industries but, where present, we can say that, if the scale of operation continues to increase beyond a certain point, the cost will begin to increase and thus check the tendency toward monopoly. So long, however, as the costs decrease with enlarging scale of operation, competition will work in the direction of monopoly control of the whole industry.

In some industries, the nature of the service makes for monopoly. Competing highways, public schools, post offices, telegraph, telephone, street railways, etc., are incompatible with the most effective service to society. The most adequate service to telephone users is a system that brings them into communication with the largest number of persons. Competing companies would require duplication of plant and equipment as well as the annoyance in locating a party at the time of a call. Similar statements can be made concerning street railways, electric light, gas, and waterworks. Because of the duplication in plant

¹ E. H. HARRIMAN is alleged to have said that 15,000 miles of railroad were all that any president of a railroad company could manage effectively.

and equipment and the inconvenience occasioned thereby, we may say that the nature of the service rendered by the industries in this group makes it desirable that they be operated as monopolies.

Another contributing cause of monopoly is the relatively high fixed expenses in the operation of such an industry. Take the case of railways. If competing railroads should undertake to cut rates, the inevitable result, as history has shown, would be that they would be forced into bankruptcy or into agreement as to rates. Competition in such industries is self-destructive, because the relatively high proportion of fixed expenses puts a premium on rate cutting, for a loss of business would not reduce proportionately the fixed expenses. When management finds itself faced with a situation of this character, the temptation to secure business at any cost above the direct cost in handling the goods is so great as to be well-nigh irresistible. The fixed expenses will continue whether there is any business done or not, and any surplus above the direct costs growing out of such transactions is likely to be regarded as desirable, because such a surplus will help to absorb the fixed expenses.

In the foregoing treatment, it should be clearly understood that the forces described tend to develop conscious control and monopoly in industry. They are positive forces that prevent the perfect operation of competition. It now becomes our task to examine the effect of these forces in the determination of value.

Monopoly Value.—The chief difference in the determination of monopoly as compared with competitive value is in the control exercised over supply. In the disposal of goods, the monopolist can so regulate the amounts offered for sale as to accomplish his own private purposes to an extent that is impossible for those producing under conditions of competition. His primary purpose as a monopolist is to avoid the constant wearing down of business profits through the force of competition and to secure for himself maximum net returns. The only method of accomplishing this end is to regulate the amounts offered for sale so that the supply and demand influences will establish a price high enough to yield him maximum net returns.

But it should not be inferred that the monopolist is free to charge any price he wishes for his goods, because, under ordinary circumstances, the purchasers can find substitutes and, besides, they can always economize to some extent in their use of the

goods. But once the supply has been offered on the market, the determination of its price is not different from that of a similar amount sold under competitive conditions. In other words, monopoly presents no new problem in the determination of the value of goods actually offered for sale. Their value will be determined in accordance with the principles developed in connection with a fixed supply. There will be some price that will clear the market of the amount offered, and that price is determined by the marginal utility of the good to marginal buyers.

In the exercise of his control over supply, the monopolist encounters different conditions of both production and demand. It will now be our purpose to examine into these different circumstances under which monopoly may exist and to inquire how they affect the exercise of monopoly power. Let us assume, first, that the supply of goods is definitely fixed and ask how a monopolist might proceed in the disposal of that supply. Since the monopolist has no control over demand in the case assumed, market forces will operate naturally. The demand for the good may be either elastic or inelastic. If the demand were inelastic, the monopolist might find that half the supply would bring more than if the whole amount were sold. Under such circumstances, it would be to the interest of the monopolist to destroy half of the supply (or some other portion, if that were found more profitable) and offer the balance for sale. If the demand is elastic, he is more likely to find his highest profits accruing from the sale of the whole amount, for any attempt to raise price will be followed by a falling off in the amounts purchased and a loss in the gross returns received. It follows from this analysis, that limitation of supply is more likely to be found when demand is inelastic than when it is elastic.

The above argument shows how elasticity of demand affects monopoly on the assumption that there are no costs to be considered. Seldom would such an assumption correspond with actual conditions, for most goods require a cost to produce them and to put them on the market. When the monopolist has to keep in mind his expenses as well as his gross returns, he will still endeavor to control supply so that his gross return, less his expenses, will be a maximum.

Since costs may be either constant, increasing, or decreasing as the supply is increased, there is a threefold problem for the

monopolist to consider. If costs are constant, he has only to adjust supply so that his net returns are a maximum. This condition can be illustrated as follows: In Fig. 23, the line CC''' represents constant costs per unit of output. Under these circumstances, if the monopolist should decide to sell OA quantity, the price would be AG . The area $HCC'G$ would represent his net returns. But if he thought that he would receive larger net returns by selling a larger volume, he might carry production to the point B . In this case the price BE would be established, and the area $FCC''E$ would represent the net returns. Whichever of these two areas, $HCC'G$ or $FCC''E$, is the larger, will tend to determine the policy of the monopolist, or if some point between A and B should be found to yield larger net returns, the monopolist would adjust supply accordingly.

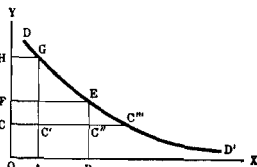


Fig. 23.

If the costs increased with increasing supply the problem of the monopolist would be more complicated. In Fig. 24, increasing cost is represented by the curve CC''' . Should the amount put on the market be represented by OA , then the market price would be AF and the costs AC' . The area $KC'FI$ would be the monopoly profit.

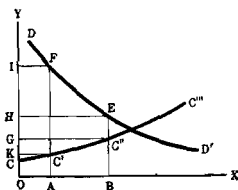


Fig. 24.

If the monopolist undertook to increase his output to OB , the unit costs would increase to BC'' , while market price would fall to BE . Monopoly profits would now be represented by the area $HGC''E$. Under these conditions, the monopolist would be under a twofold influence to limit supply, for a smaller product would tend to decrease his unit costs and to increase the market price. There are fewer instances of monopoly in the case of increasing costs than those where costs tend to decline with increasing supply. One of the nearest approaches to this kind of monopoly is that found in the anthracite coal industry. Control over the anthracite deposits is nar-

rowly confined so that, barring regulation, the development of monopoly tends to become a significant social problem in this industry. The demand for the coal is fairly inelastic, which makes the temptation to limit supply very strong. No doubt the fear of government interference, because coal is such an essential commodity both for domestic and industrial uses, has prevented monopoly from developing to a greater extent in this industry.

If the monopoly is in an industry in which the costs tend to decrease, the problem is again different, as can be seen in Fig. 25. The decreasing costs are represented by the curve CC'' . In this case, if the supply offered were OA , the costs would be AC' , and

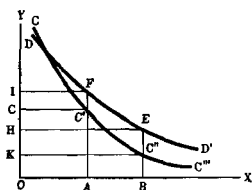


FIG. 25.

the market price AF . Monopoly profits would be represented by the area $ICC'F$. But if the supply were increased to OB , the costs would fall to BC'' , and market price would fall to BE . Profit would now be represented by the area $HKC''E$. A monopolist would benefit by decreasing costs, pro-

viding the costs declined faster than the fall in market price. In this situation, the monopolist would tend to increase his supply and sell on a lower margin per unit, for he would enjoy in this way the largest net yield. This would be particularly true, if the demand for the good is elastic. But even if the demand were inelastic, the temptation to limit supply would be less than in the case of increasing costs. Although market price would rise if a smaller supply were put on the market, so also would the expenses. In determining the policy of a monopoly much would depend in an actual situation upon the relative rate of increase of costs with the shrinking of demand.

Class Price.—In the previous discussion, the conclusion was drawn that the sale of a definite supply of goods under monopoly presents no new problem of value. One modification of that general statement has to be made. In making the statement, it was assumed that the whole supply was offered in a single market. The conclusion as stated would be substantially correct if the goods were sold as assumed. But by means of control over

supply, the monopolist can determine the method of marketing his good in a way not open to those operating under competition. Under competition there tends to be one price per unit for the whole supply offered for sale, but a monopolist may classify the price and sell to different customers at different rates. He will find it to his advantage to follow this practice if there is a considerable variation in the income of his customers.¹ The recognition of consumers' surplus, which has already been explained, will stimulate the monopolist to divide his customers on the basis of income into different groups or markets and to sell the same good at different prices to the various groups of customers. Thus, by classifying his price, he is enabled to absorb some of that surplus and add it to his net profits. This constitutes what is generally known as "class price" and it may be regarded as a special case of monopoly value, as it could not exist unless there were some obstacles to the perfect operation of competition.

A common method of classifying price is by selling the same article to different customers at different prices, or by giving a somewhat different service at a price higher than offsets the difference in the cost of the service rendered. For instance, a soap manufacturer may sell an unwrapped cake of soap at three for 10 cents, while he will charge 5 cents per cake if the soap is wrapped. If the wrapping is fancier and the soap is scented, he may charge 10 cents. Then he may box the soap in a fancy box and sell it three for 50 cents. In each case the difference in service is more than offset by the difference in price. There are many ways by which the price of an article may be classified. Open classification may cause criticism so that variation in price is usually disguised by some modification in service. Another common method is to put out a commodity in the early part of the season at a high price and, as soon as the demand has been satisfied at this price, the dealer will cut the price and continue to do so until his entire stock is sold. This is a common method of merchandising, and there frequently is present an element of classification of price to fit persons of different incomes. The

¹ This practice of classifying price is frequently employed by firms that are operating in a competitive field, but the ability to classify price in such instances is due to the presence of obstacles to perfect competition. The phenomenon of class price, therefore, is not a characteristic of a competitive market, but is found only where there is monopoly, or, where there are obstacles to the perfect operation of competition.

purchasing at the high price may not be an accurate index of income, as there are always some persons in all income groups who must buy the last article demanded by fashion irrespective of price. Nevertheless, such a method of selling does tend to approximate a class price.

It is frequently asserted that department stores sell the same article in the bargain basements at a lower price than is charged on the upper floors. We also know that railroads, especially in Europe, and steamship lines regularly classify their passenger fares. A second-class fare on the railroads in this country deprives the passenger of the privileges of the Pullman cars, but otherwise the privileges are apparently the same. In Europe the difference in service is more pronounced, as the different classes of passengers are accommodated in different coaches. Another example of classification goes under the name of "dumping." This consists of selling articles at home at one price and in foreign countries at another. Manufacturers, who have a larger productive capacity than the domestic market can absorb at what is regarded as a paying price, will, if the domestic price can be maintained, sell the balance in foreign markets on such a basis that will cover the direct expenses involved in manufacturing and transporting the goods to the foreign market. If anything above these expenses is received, it will help to absorb part of the costs of producing the goods that the domestic supply would otherwise have to bear.

Industries producing under a condition of decreasing costs would secure the advantage of lower unit costs as the volume produced increased and if, in addition, the demand tended to fall off rapidly at certain points, the monopolist would have a strong inducement to make his foreign prices low in order to enjoy the lower unit costs upon the domestic as well as upon the foreign supply. When the domestic market can be protected by a tariff, as it frequently is, then it is easier for the monopolist to classify his price as between domestic and foreign purchasers. If there is a heavy duty on the importation of this good, the favored foreign purchaser is not tempted to reship and become a competitor in the domestic market. With the domestic price thus protected, the monopolist can regulate the quantity sold so that the price will be high enough to give him a profit but not so high as to invite foreign competition. The more complete the

monopoly, the more likely it is that dumping will take place, and the greater will be the variation in prices to fit the conditions of sale in both domestic and foreign markets.

Complete monopoly is rare except in the cases of certain public-welfare monopolies, such as patents, franchises, or fiscal monopolies, and some natural monopolies. In most instances where monopoly is present and has an influence on price, there are limitations that prevent complete control over price. First, there is the possibility of substituting some other good. If the monopolist undertakes to regulate supply so that price becomes burdensome, it puts a premium upon the finding of a substitute. Hence, the monopolist is not independent of the consumer's attitude in the market. He cannot charge any price he sees fit, for he ordinarily has no power to compel the purchaser to buy.

In the second place, he has to avoid public condemnation and legal interference. If he exercises his monopoly power too vigorously, he may create a hostile public opinion which may result in legal restraint upon the freedom of his action. No more significant movement has taken place during the last forty years than the legislation enacted against the exercise of monopoly power as found in the control over railroads and other public utilities. The popular indignation against monopoly, that found expression in such bodies as the Interstate Commerce Commission, state public utility commissions, and other regulatory bodies, is evidence that the monopolist must reckon with this limitation, even though he may forestall for a time the effects of such legislation by resisting it in the legislative halls and in the courts.

There is, finally, the possibility of competition. Few monopolists can prevent the success of any business venture from becoming known. If the market price becomes so high as to give large returns to the capital invested, other capitalists are likely to be attracted. Even though large sums may be required to launch a competing concern, if the prospect for profits looks attractive, the necessary funds can ordinarily be found. The possibility of competition, then, is generally present to limit the power of monopoly over price.

Monopoly Price as High Price.—One further question requires consideration at this time. In the minds of most people, monopoly price is regarded as a high price. How much truth is there in this belief, and how does monopoly price differ from competitive price? Monopolists often point to the fact that the price of a

good is now lower than it was before the monopoly was formed. Kerosene is cheaper now than before the Standard Oil Company became such a powerful influence in the oil industry. Other illustrations of like character are cited to show that, after the monopoly has developed, the price is actually lower than when the article was sold under competitive conditions. But such statements do not meet the issue squarely. Competition works constantly to bring market price down to the cost of producing the good or service, while monopoly price tends to be fixed at the point which will yield the highest net returns, irrespective of the cost of production. The difference between the two prices can be illustrated in the following table:

Total sales	Price per unit, cents	Gross income	Total expense	Net earnings
2,000,000	15	\$30,000	\$36,000	-\$ 6,000
4,000,000	12	48,000	42,000	+ 6,000
6,000,000	10	60,000	48,000	12,000
8,000,000	9	72,000	54,000	18,000
10,000,000	8	80,000	60,000	24,000
12,000,000	7	84,000	66,000	18,000
14,000,000	6	84,000	72,000	12,000
16,000,000	5	80,000	78,000	2,000
18,000,000	4	72,000	84,000	- 12,000

It will be noted from the above table that a very different price situation would exist in a monopolized than in a competitive market. If there were a strong monopoly control exercised, the price per unit would tend to be fixed at 8 cents, for at this figure the net profits would be at the maximum. But if competition were active, then a larger volume would likely be sold and the price would approach 5 cents, for at this price there would still be a profit over all expenses involved in bringing the good on the market.

This table, while clearly artificial, illustrates the difference between monopoly and competition as market influences. If the actual figures of sales under the two conditions could be tabulated, they would show the same kind of differences as is indicated in this table. Competition would always work toward a market price that just balanced the costs of production, while monopoly would stop the influence of competition at that point which would

yield the maximum net returns. While it may be true that historically prices may be lower under monopoly than they formerly were under competition, yet at any given time monopoly price is likely to be a high price as compared with the price that competition would tend to establish at that same time.

While it is true that monopoly price is a high price in the sense in which it has just been explained, there is one generally accepted belief concerning monopoly that is not supported by careful analysis. It is commonly believed that a monopolist can secure a higher price for a given supply than if the same amount were sold by a group of competitors. Assuming that the whole supply is sold and that there is no attempt to classify price, then it is probably true that a monopolist cannot get a higher price than if the same quantity were sold in a competitive market. The estimate of the buyers as to the utility of the supply would be the same in each case, and unless the monopolist were able to classify the price, so that he could divert to himself some of the consumers' surplus enjoyed by a part of the buyers, he would be unable to charge more than the marginal purchasers were willing to pay. The possibility of classifying the market is much greater under monopoly than where competition prevails and, if the market is classified, a given stock can be sold for more than in a competitive market.

The "Corner" as a Monopoly.—This argument bears directly upon that form of monopoly that is commonly known as a "corner." Frequently, speculative dealers in grains attempt to corner the market as has been previously described. They buy all the supply that is offered, or an amount that is sufficient to give them power over the market. When a corner is secured, it is generally believed that the dealer can get a higher price for the grain than would have been possible if control over the same supply had not been thus concentrated. The impression needs careful analysis before acceptance. Take the case of a dealer who is endeavoring to get control over the supply of some commodity, such as wheat. He proceeds upon some theory as to the course that the market price of that article will take. Let us suppose that he believes that the price of wheat will be higher in three months than it is now, while most other dealers believe it will be lower. Under these circumstances he will buy wheat now for future delivery. If his theory concerning the direction of wheat prices is correct, he will make a handsome profit when he

comes to sell his holdings. But if he should make a mistake in his forecast of the market, and the prices fell, he would then be the loser. The large profit which he obtained was not because of his monopoly control over the supply, but because of his foresight in accurately gauging the fundamental forces of demand and supply. The advance in price came in spite of concentrated ownership, so that monopoly in this case can not be regarded as the cause of the high price. It is only when the monopolist reduces the supply by dumping or by other means, or classifies the price, that he can actually command more for a given amount than if that amount were sold under competitive conditions.

Suggested References

- ELY, RICHARD T., "Outlines of Economics," 4th Ed., Chap. XII.
——— "Monopoly and Trusts."
FAIRCHILD, FURNISS, and BUCK, "Elementary Economics," Vol. I, Chap. XVIII.
HAYES, H. GORDON, "Our Economic System," Vol. I, Chap. XXIII.
MARSHALL, ALFRED, "Principles of Economics," Book V, Chaps. VI, VII, and XIII.
FIGOU, A. C., "Wealth and Welfare," Part II, Chaps. IX-XII.
RUFENER, L. A., "Principles of Economics," Chap. XV.
TAUSSIG, FRANK W., "Principles of Economics," 3d Ed., Vol. I, Chaps. XV and XVI.
VANDERBLUE, HOMER B., "Economic Principles," Chaps. VIII and IX.

CHAPTER XI

MONEY AND THE MECHANISM OF EXCHANGE

Division of Labor and Money.—In the three preceding chapters, value has been presented as a market phenomenon that grows out of the trading and exchange which result from the division of labor. In modern economic society, production is primarily for sale, but before goods can be traded a basis of exchange must be found. We have seen that the basis for trading goods is established by the evaluation of those goods that takes place in the minds of buyers and sellers. It is this process of evaluation that determines the value of the goods. It will be observed that, in our treatment of value, the subject of money was not introduced as an integral part of the discussion. The reason for this is because the phenomenon of value not only exists prior to the use of money, as in the case of barter, but it can be explained without bringing into the discussion the subject of money. Although the phenomenon of value may appear quite independent of the existence of money, yet in modern trading the value of goods is nearly always thought of in terms of money. The relative significance of goods in a market is generally expressed in terms of money at a definite price per unit. Thus, money becomes an easy and convenient means of expressing the value of goods and of effecting exchanges. In the actual operations of our economic life, money plays such an important part that it now becomes necessary to analyze the mechanism of exchange and to apply the principles of value to money.

The Disadvantages of Barter.—The simplest form of trading is a direct exchange of goods, commonly known as "barter." In primitive societies, exchange begins by a swapping of goods. One person may have a surplus of grain while another may have a surplus of fish or game. Before a trade can take place in such a community, it is necessary for the man with the surplus of grain to find some one who has a surplus of some other commodity which the first man desires. Besides, the second man must desire the grain which the first man is willing to trade. This situation has

been called the "double coincidence in barter," and constitutes one of the chief difficulties of trading by means of barter. It is evident that the obstacles to barter become increasingly great as the number of exchanges increases. The larger the number of those who specialize in the production of goods, the more difficult it is to find this double coincidence of the possession of goods and the willingness to trade. Because of these inconveniences, barter very early gave way to the use of some form of money as a means of facilitating exchange. When some one commodity became generally accepted in making purchases or sales, society passed from a *barter* to a *money* economy.

Definition of Money.—From the preceding statement, it will be apparent that whenever any commodity comes to be generally used as a means of effecting exchange it becomes known as "money." Money, therefore, may be defined as any good that is generally accepted and passes freely from hand to hand in making exchanges. While it is not possible to run a sharp line distinguishing between those things which are, and those which are not, money, this definition will assist in clarifying our conception of the essential nature of money. For instance, according to this definition a check is not money, even though it may be used to perform money work, because it does not pass freely from hand to hand in making exchanges. While checks are extremely important in making payments, as will be shown later, they have only a limited circulation, and for this reason should not be classed as money. They are merely a convenient means of economizing in the use of money. On the other hand, a bank note, which does pass readily from hand to hand in making exchanges, is by this definition money. It will be apparent that this definition of money is more restricted than the usage of the term among business men. Reflecting this usage, the financial columns of the newspapers include under the term "money" all forms of transactions in terms of money." Reference is made to the "money market" to "easy money," etc., in such manner as to include all forms of credit transactions. While this usage serves the ordinary purposes of business, it is not sufficiently exact for scientific analysis, hence, the term will be used in this text in the more restricted sense given in the definition above.

The Functions of Money.—From the above definition, it might be inferred that the sole function of money was to effect exchanges, but a little reflection will reveal the fact that money

performs a number of functions, the most important of which may be listed as follows: (1) To serve as a medium of exchange; (2) to act as a common denominator of value; (3) to serve as a measure of value; (4) to serve as the means of saving and accumulation; and (5) to act as a standard of deferred payments. A word of explanation will make plain each of these functions.

1. The first function is evident. The inconveniences that one would encounter in making exchanges without the aid of money have already been shown and will be apparent to all persons who have had even a limited experience in buying or selling goods. Because money is so universally used as the means of securing the things which people want, the impression is very frequently held that the accumulation of money rather than of wealth is the end of economic activity.

However, we seldom find money serving in the capacity of a medium of exchange alone. But during and after the World War, the extremely rapid depreciation of the paper monies of Germany, Austria, and Russia led to a situation in which the monies of these countries lost practically all other functions except that of effecting exchanges. Marks and rubles were of no use except to spend and the more quickly they were spent the less the loss from depreciation to the holder. Wages fixed at the beginning of the week might fall 50 per cent or more in purchasing power by the end of the week and the working man had no recourse. Merchants are said to have repriced their goods several times a day in an effort to keep pace with the depreciation. So completely did the monies of these countries lose the ordinary functions of money, that people again resorted to the accumulation of consumers' goods, such as flour, sugar, etc., as a means of saving. The modern world has never seen the depreciation of money on such a scale as took place in these countries, but their experience illustrates how the functions of money may be separated and how a commodity used as money may perform only one of its many functions.

2. Trading can take place if the value of the articles can be expressed in such terms that a ratio of exchange can be found. Money serves this purpose. We may quote the value of corn in terms of money, by saying that a bushel of corn is worth \$1. In like manner we may say that the value of a bushel of wheat is \$1.50. We have here reduced the value of these two commodities to a *common denominator* and have established a ratio of

exchange. In this instance, it would require one and one-half bushels of corn to purchase or exchange for one bushel of wheat. These two commodities could be exchanged in this ratio without the use of actual money, simply by expressing the value of each commodity in terms of money. When the farmer takes butter and eggs to the country store and buys sugar and coffee, these transactions will be in terms of money. The farmer's produce will be sold at some price and sugar and coffee purchased at a price, then these two sums will be balanced. If the amounts are just equal, no money is used in the buying and selling of the goods, so that money serves in such cases solely as the means for comparing the value of the commodities traded. Should there be a balance due one way or the other, money may be used to settle the difference. While cancellations of this kind are frequent enough in rural communities, the volume of such business performed in this way is relatively insignificant. But this method of *check off*, or cancellation, becomes immensely important where banking has been extensively developed. As will be described more in detail later, banks regularly trade checks held by and against each other and thus reduce the actual volume of money that is transferred in making settlements with one another. All such transactions are conducted in terms of money or, in other words, money is serving as the common denominator of value.

3. Closely akin to the function just described is that of the *measure of value*. How much is a good worth? The answer to this question implies the existence of a *standard measure of value*. We are familiar with standard units of measure in other lines, such as units of length, of weight, or quantity, etc. Unless we had the yard, the mile, or some other similar unit, we would have great difficulty in expressing length or distance. Without the gallon, or bushel, or pound, we would have similar difficulties in expressing quantity and weight. These are all familiar units of measurement. It is just as essential in our economic life to have a unit for measuring the value of commodities as it is to measure their quantity or weight. Money serves in this capacity as a standard measure of value. The particular monetary unit chosen depends upon legislative enactment, hence, monetary units vary from country to country. In the United States, the unit is the dollar; in England, it is the pound sterling; in France, the franc; in Germany, the mark; and so on. Whatever mone-

tary unit is adopted, thereafter the value, or social significance, of a quantity of wealth may be expressed in terms of this unit. For instance, we may say that the value of a house is \$25,000, the dollars serving here as a means of expressing quantitatively the amount of value which the house has at a given time. Thus, money serves this function of measuring value.

4. The fourth function of money is to serve as a means of *saving and of accumulation*. In primitive times the only means of saving was to store away a stock of wealth. But as soon as money came to be generally used, instead of storing wealth, purchasing power in the form of money was stored. So long as the value of the monetary units does not change greatly from one time to another, money will serve as a convenient means of saving. This may be done by hoarding where the actual coins are stored away to be used at a future date, or by means of deposits in a bank where banking has been highly developed. By depositing gold coins in the bank, the saver creates a right to receive gold coins on demand at some future date. Furthermore, we may accumulate a right to receive an income by purchasing stocks or bonds, which are expressed in terms of money. In modern times in a country like the United States, much of the saving of individuals is done by the purchase of industrial securities. Dollars, or the right to receive dollars, are traded for the security which gives to its owner the right to receive income when earned. In this way, money becomes a very important means by which saving and accumulation are accomplished.

5. The last function of money, mentioned above, was to serve as a *standard of deferred payments*. When one person borrows from another, the contract is most likely to be expressed in terms of money. For instance, A borrows \$500 from B to be paid back five years from date. A receives \$500 in money and has the use of it in his business for the time specified in his contract. When his note falls due, A will pay B the \$500 called for in the note. Deferred payments are likely to be expressed in some such manner as this. If the value of the dollar does not change during the interval covered by the note, A will pay B as much purchasing power as he borrowed, but, as we shall see later, should the value of the dollar change in this interval, then either A or B will gain thereby. Nevertheless, money is the regular means by which contracts involving deferred payments are made. Thus we may conclude that money performs the important functions of serving

as a medium of exchange, as a common denominator of value, as a measure of value, as a means of saving and accumulation, and as a standard of deferred payments.

Characteristics of Good Money Material.—The manner in which money performs its functions depends upon the character of the money itself. Some commodities are better adapted to monetary uses than others, although historically, a great many different commodities have served in this capacity. We find—just to mention a few examples—that cattle, sheep, hides, copper, iron, tobacco, wampum, shells, gold, and silver have all served as money at one time or another. By a kind of survival of the fittest, gold and silver have become the chief money metals of the civilized world. Before any commodity will be generally used as money, it must be desired by a large number of persons. In explaining why gold and silver have survived other commodities as money, Professor Taussig says:

The chief reason why gold and silver became the money metals was that they satisfied the craving for adornment. Things that minister to the deep-rooted love of display are in unfailing demand; and any commodity that is in unfailing demand may perform passably the functions of a medium of exchange.¹

General acceptability, then, may be regarded as the first quality of a good money material. It must have utility in the sense that many people desire it. A second quality is *durability*. To perform its functions well, the money material must be capable of withstanding the wear and tear of constant use. While paper money, which is widely used in this country, seems to be an exception to this rule, it must be remembered that its use is justified primarily on the basis of convenience and that it is not the fundamental money of the country. In fact, the expense involved in keeping paper money in circulation has led to an attempt to put a larger volume of silver dollars into use in the place of silver certificates. The cost of reissuing these silver certificates is large and the treasury department endeavored to substitute in our circulating medium 40,000,000 silver dollars as an economy measure. This attempt failed because people prefer a more convenient form of money than the silver dollar; its bulkiness is a serious handicap to its general use.

Another characteristic of good money is that of *uniformity*. To serve the functions of money well, a commodity should

¹ TAUSSIG, F. W., "Principles," Vol. I, p. 224.

always be of the same quality so that equal amounts of it will have the same value. In defining monetary units, it is customary to specify a definite weight and fineness of the money material, such as gold or silver, and it is highly essential that these units should be uniform in quality. Closely associated with this quality is that of *divisibility*, i.e., the ability to be divided easily into units without loss of value. Unless the commodity used as money can be subdivided without loss of value, the making of coins of different denominations, a condition so essential for the effective trading of goods, would be a costly process. A fifth quality of the material selected as money is ease of recognition. Great inconvenience and loss would be entailed if the money material could be easily counterfeited. *Cognizability*, therefore, is another important quality of a good money.

Lastly, the money material should be *convenient*. It should be neither too valuable nor too cheap. If the money material is too valuable, the standard unit will be so small as to be inconvenient. In fact, the gold dollar does not conform to this quality of money, as it is too small for general use, and is, therefore, no longer coined. While the gold dollar is the standard unit of our monetary system, other kinds of money, as silver and paper, are used in suitable denominations for the sake of convenience. On the other hand, if the material is too cheap, the size of the monetary unit will be too large for convenience. The size and weight of the silver dollar militate against its usage. The silver certificate, a form of paper money, takes its place in our circulating medium primarily because of convenience. If we were to use copper, nickel, iron, or any of the baser metals, the size of the coins would be so unwieldy and inconvenient as to make the carrying of money in large quantities extremely difficult. To comply with the test of convenience, the money material should be sufficiently valuable as to enable a relatively large sum of value to be embodied in a unit that can be easily carried. We find, in fact, that experience has dictated the use of more than one form of money in meeting this test—a fact to be further elaborated under the subject of the monetary system. In general, we may conclude that the qualities of a good money material are acceptability, durability, uniformity, divisibility, cognizability, and convenience.

In addition to the qualities already mentioned, there are two others that deserve recognition. A good money material is

one whose value is relatively stable. One of the difficulties in using money as a standard measure of value is the fact that the standard itself fluctuates in the particular quality which it is intended to measure, *viz.*, that of value. On this account, money is not an accurate measure of the value of commodities. Hence, it is not always easy to distinguish between the changes that affect the value of the goods and those that affect the value of the money itself. During the World War every one found that the dollar would purchase only about one-half as much as formerly. The significant question to ask in circumstances like these is, whether the change has come on the side of the commodities, or on that of the money. Therefore, in selecting a money material "relative stability" of value should be given due consideration.¹ Gold is superior in this respect to many other materials, because the yearly additions to the world's stock of gold is relatively small, hence, the value of gold is affected less by changes in the yearly output than is true of most other commodities. But, as we shall see later, the advantage of gold in this respect is only relative and some writers have advocated the establishment of a different standard of measurement.

Another quality of good money, that applies to the monetary system as a whole rather than to the money material, is *elasticity*, *i.e.*, the capacity to expand and contract in accordance with the needs of business. When the volume of business transactions is large, there is need for more money than when this volume is small. The best monetary system will include an elastic element for this purpose and in most countries this elastic quality is provided, as we shall see, by means of some form of bank credit.

The Monetary System.—The system of money in use in any country is in modern times the result of legislation. The first step in the development of that system is the adoption of a monetary unit. The purpose of the unit is to make it possible to express quantities of money, and thus to enable money to perform effectively its functions of a medium of exchange and a measure of value. Each country fixes its own unit and the fractional parts, or the multiples thereof. In England, the unit is the pound sterling; in France, the franc; in Germany, the mark; and so on. In the United States, the unit is the dollar, and provision has been made for fractional parts and multiples

¹ Edm., L. D., "Economics, Principles and Problems," p. 496-497.

of this unit. The other denominations of our present monetary system are the cent, the nickel, the dime, the quarter, the half-dollar, and the quarter-eagle, the half-eagle, the eagle, and the double eagle. By means of these denominations it is possible for the people in this country to express any quantity of money which may be involved in the trading of goods.

The second step in the development of a monetary system is the determination of the monetary standard, the purpose of which is to fix the size and the material of the monetary unit. This task is comparable with that confronted in every other scientific field, and consists of a method of expressing quantity. In all sciences it is necessary to select and define an arbitrary unit of measurement as a means of conveying the notion of quantity. In the case of weight, bulk, and distance, the federal government has established units of definite size, as the pound, the bushel, the gallon, and the yard, which it undertakes to maintain through a bureau of standards. The character of such a standard unit may be illustrated by means of the gallon which is defined as a container that holds 8.33 pounds of pure water under definitely specified conditions of temperature and air pressure. Any container that deviates from this standard unit is not a true gallon. In like manner specifications similar in character are established for other standard units. So also in the case of money, we have the standard unit, the dollar, which was established by the Mint Act of 1792.

This Act, which followed substantially the recommendations of Alexander Hamilton, the first Secretary of the Treasury, established a bimetallic standard. It provided that the dollar should consist of 24.75 grains of pure gold, or 371.25 grains of pure silver. Thus a definitely specified weight of gold or silver was made the monetary standard. Hamilton thought that this mint ratio in weight of gold and silver corresponded to the market ratio of the value of the two metals. In other words, he thought that 15 grains of silver was equal in value to 1 grain of gold. This mint ratio was changed in 1834 (slightly modified in 1837) by reducing the weight of the gold dollar to 23.22 grains of pure gold, or 25.8 grains of gold nine-tenths fine. The amount of silver remained unchanged so that the new ratio was approximately 16 to 1, a ratio that has been retained until the present time. Thus, the legislation referred to established a bimetallic monetary standard consisting of either gold or silver in the quan-

tities specified above. This situation was changed in 1900 when Congress made gold the standard money of the country, so that now the gold dollar of the size and fineness mentioned above is the standard monetary unit.¹ There are other forms of money in circulation, but, as in the case of liquid measure, unless a container conforms to the standard, it is not a true gallon, so other forms of money must be kept at a parity with the standard dollar in order to be true dollars.

Attention is here called to a fact that is often overlooked, namely, that the coin value and the bullion value of a gold dollar of standard weight are always the same.² In other words, a gold dollar of standard weight is worth as much in bullion as it is in a coin. There can, therefore, be no variation between the market and the coin values of gold, for should gold be worth more as coin than as bullion, it will be taken to the mint and made into coins until the coin value and bullion value are equal. This is not true of any other form in which the dollar appears. The silver bullion in a silver dollar is now worth much less as bullion than as a coin, and of course the paper in a greenback or a bank note is worth scarcely anything as paper. How these latter forms of money are maintained at a parity with gold will be explained below. For the present it is sufficient to recognize that the gold dollar is the standard unit in the monetary system of the United States.

The Coinage System.—It is not sufficient for a government to establish the denominations to be used in its monetary system, or to determine the size and material of its standard money. Experience has shown that the coins in circulation must be uniform in character. In order to secure this uniformity, governments have found that it is necessary to maintain a monopoly control over the issuance of coins used within their respective boundaries. Coins are merely pieces of metal that are stamped and certified by the government in respect to weight and fineness. In order to avoid clipping, shaving, or other means of debasing coins, they are now stamped on each side with appropriate designs and the edges are milled or lettered. Modern machinery can turn out coins in such form that attempts at reducing the metal content can be easily detected. Coins are seldom made of pure

¹ Dewey, D. R., "Financial History of the United States," 2nd Ed., pp. 101-104; 210-212; 468-469.

² Except for the slight deviation mentioned below, due to the loss of interest for the time required in minting the bullion.

metal. Gold and silver are too soft to withstand constant usage so that an alloy is used to render them more durable. In most countries the coins are made nine-tenths fine, that is, in every ten parts one will be alloy and nine pure metal. This is the fineness of the coins of the United States. In Great Britain the alloy is in the ratio of 1 in 12 or $916\frac{2}{3}$ fine.

Coinage may be either free or limited. In the case of free coinage, every holder of bullion may take it to the mint and have it converted into coins. If the cost of converting the bullion into coins is borne by the government, then coinage is said to be *gratuitous*. But if a charge is made to cover the expense of minting, or to make a profit to the government, this charge is known as *seigniorage*. If this charge merely covers the expense of minting, the term *brassage* is ordinarily used, and seigniorage would then mean charges in excess of the cost of manufacturing the coins. In the past, governments resorted quite generally to the practice of charging more than the cost of minting the coins as a means of securing a profit for the treasury. It was argued that the *government stamp* gave to the coins their value rather than the metallic content, so that, if a government placed its stamp on a piece of metal, it would circulate at its "stamped value" even though this were above its market value. The amount of bullion thus kept out could then be used to help defray the expenses of the government. Experience has shown conclusively that the value of a coin, when freely issued, depends upon the value of the metallic content rather than upon the government stamp. This fact has become so generally recognized that all modern governments have abandoned the practice of charging a seigniorage for the purposes of a profit.

In the United States, gold is the only kind of monetary metal that is freely coined. One may take gold bullion of standard fineness to the mint and exchange it for coins on the basis of weight. If the gold bullion presented is not of standard fineness, however, a small charge is made to cover assaying, refining, and the alloy. Ordinarily a period of six weeks is required for the minting of gold bullion, hence, there is a loss of interest for the use of the money during this period, which may cause a slight difference in the bullion and coin values of gold. Otherwise it is correct to say that the bullion value and coin value of standard gold coins are equal. This is not true of any other coins in our monetary system.

In the case of all silver coins, including the silver dollar, there is a limited coinage. Only such amounts of silver bullion may be minted into coins as have been authorized by Congress. The bullion is purchased in the bullion market at commercial prices and coined into denominations of the metallic content specified by law. The government enjoys a considerable profit in the coinage of silver. However, no modern country looks upon the coinage of money as a source of profit, but rather as a function that can best be performed by the government for the convenience of the people.

Kinds of Money.—The money in circulation in the United States can be classified as follows:

1. *Metallic Money.*—This consists of the standard gold coins, the quarter eagle, the half eagle, the eagle and the double eagle, and of the silver dollar, and subsidiary coins to be described later. The silver dollar holds an anomalous position in our circulating medium. For years it was in law the equal of the gold dollar but since 1900 it has lost its legal position as a standard unit. The value of the silver dollar is kept at a parity with that of gold although it is not directly interchangeable with gold. It has full legal tender value, that is, it is acceptable in meeting all obligations both public and private in the absence of a contract to the contrary; but, its coinage is limited to the amounts specifically provided by Congress. Its coin value is distinctly above the value of the bullion contained in it, and in this respect it is like subsidiary coins. The situation of the silver dollar would be much clarified if Congress frankly recognized it as a subsidiary coin and passed legislation to establish it in the class of coins to which, in fact, it belongs.

2. *Paper Money.*—This appears in many forms in our circulating medium. First, there are the *greenbacks* which have continued as a part of our monetary system from the period of the Civil War. They were issued by the government and were at that time simply promissory notes of the government. In other words, the greenbacks were an inconvertible paper currency, and since they were at first merely promises of the government to pay, these notes constituted at that time an example of what is commonly known as "fiat money." They were made legal tender for all public and private obligations, in the absence of a contract, except the payment of import duties and interest on the public debt. The Resumption Act of 1875 gave authority to the Secre-

tary of the Treasury to accumulate gold for the purpose of resuming specie payments but no provision was made for a definite gold reserve. When specie payments were resumed in 1879, the volume of greenbacks outstanding was in round numbers \$346,000,000 and this amount is still in circulation. From 1879 to 1900 no special fund was kept for redeeming the greenbacks but the Treasury used any surplus funds on hand for this purpose. In fact, there was some doubt during this period as to the responsibility for redemption, but this doubt was removed by the Monetary Act of 1900 which provided for a gold reserve of \$150,000,000 for the redemption of greenbacks when presented. This act, then, has made these notes convertible paper money and as such they circulate alongside of gold and silver as a part of our money medium. There is, of course, a credit element still present, which amounts to the difference between the total volume of greenbacks outstanding and the gold reserve held specifically for their redemption.

Gold and silver certificates constitute another form of paper money in circulation. They are in reality only certificates of deposit, certifying that a definite amount of gold or silver has been deposited in the Treasury of the United States. They are, therefore, only a device for facilitating the circulation of specie and have been called "representative money."¹ Both gold and silver are inconvenient as forms of money, especially when handled in large amounts. The wear and tear of constant use is large, especially on gold, and the inconvenience of carrying large quantities of either metal puts a limit on their general usage. It was found during the decade of the 'eighties, that a comparatively limited volume of silver dollars could be placed in circulation, but when provision was made for the silver certificate in convenient denominations, the volume of silver used was greatly increased. On July 1, 1927, the Comptroller of the Currency reported a total stock of \$424,935,426 of silver in circulation, but of this amount only \$49,145,184, or about 11.5 per cent was in silver dollars, the balance, or \$375,790,242 was in silver certificates. In the case of gold the proportion of gold to gold certificates was higher, the percentages being about 27.7 and 72.3 per cent.² These percentages vary quite widely from time to time,

¹ FAIRCHILD, FURNISS, and BUCK, "Elementary Economics," Vol. I, pp. 383, 401.

² Cf. table on p. 272 for the amounts of money in circulation.

but it is safe to conclude from these facts that gold and silver circulate chiefly by means of gold and silver certificates.

The stock of money in the country and the amount in circulation as of July 1, 1927 is given in the following table:

STOCK OF MONEY AND MONEY IN CIRCULATION IN THE UNITED STATES AS OF JULY 1, 1927¹

Kinds of money	Stock of money	Money held by U. S. Treasury	Money in circulation	
			Amount	Per capita
Gold coin and bullion.....	\$4,565,070,147	\$3,650,974,055	\$ 386,460,785	\$ 3.31
Gold certificates (1,625,285,099) ²			1,007,081,189	8.61
Silver dollars...	537,948,084	475,681,351	49,145,184	0.42
Silver certificates.....	(469,591,901) ²		375,790,242	3.21
Treasury notes of 1890.....	(1,326,804) ²		1,326,804	0.01
Subsidiary silver United States notes.....	295,818,732	5,347,024	275,733,058	2.36
Federal Reserve notes.....	346,681,016	3,235,483	292,200,153	2.50
Federal Reserve bank notes...	2,077,473,195	979,355	1,702,823,642	14.56
National bank notes.....	4,854,238	192,906	4,605,575	0.04
	704,146,267	19,029,816	650,055,936	5.56
Total.....	\$8,531,991,679	\$4,155,439,990	\$4,745,222,568	\$40.58

¹ Taken from the 1927 Report of the Comptroller of the Currency, p. 119.

² These amounts are not included in the total as they have already been included under the items of gold coin and bullion and standard silver dollars, respectively.

The most important form of paper currency in our monetary system is the *bank note*, which appears, in three different forms, the national bank note, the federal reserve bank note and the federal reserve note.¹ Only a brief statement concerning these three forms of money will be made here as each will be treated more fully below. The national bank notes were provided during the Civil War as a part of the system of national banks

¹ The federal reserve notes, though an obligation of the United States government, are essentially bank notes of the normal type and are secured by commercial paper and gold.

that was established at that time. These notes were issued by the national banks under the authority granted by the National Bank Act of 1863. They are promises of the issuing bank to pay on demand and they circulate from hand to hand in settlement of private obligations. The note holder is protected by the requirement that a bank must deposit with the Treasurer of the United States government bonds equal in market value to the volume of notes issued. There was outstanding \$650,000,000 of national bank notes in circulation on July 1, 1927.

The *federal reserve bank note* is a special form of note that was provided by the Federal Reserve Act of 1913. Provision was made for retiring the old national bank notes but fear existed that they might be retired so fast as to affect general prices.¹ Authority was granted the Federal Reserve Board to issue federal reserve bank notes on the basis of government debt or other definitely specified securities to replace the national bank notes retired, if in the judgment of the Board there were need for an additional volume of a circulating medium. During the World War, a considerable volume of these notes was issued, but since then they have been retired until there was outstanding on July 1, 1927 only \$4,605,575. The need for such notes is doubtful and it would seem advisable for the sake of unity and simplicity to eliminate this form of currency from our monetary system.

Finally we have the *federal reserve note* which was provided by the Federal Reserve Act in 1913. These notes are issued by the Federal Reserve Board, through the federal reserve banks upon demand by member banks for a circulating medium. They are a direct obligation of the federal government even though they are issued as a form of bank credit. The federal reserve banks must maintain a gold reserve of at least 40 per cent against all outstanding federal reserve notes which, as will be seen shortly, arise from the deposit with the federal reserve agent of certain types of paper that the reserve banks hold on their own account, or have rediscounted for the member banks.² The volume of these notes increased from \$9,000,000 in 1915 to \$3,195,000,000 in

¹ For a discussion of the relation between the quantity of money and the general level of prices, see pp. 280, 281, 301-304, 322-3.

² For a complete statement of the steps involved in the issue of federal reserve notes, the reader is referred to any standard work on banking in the United States. A convenient reference on this subject is E. W. Kemmerer's "The A B C of the Federal Reserve System."

1920. Since that time the volume has been contracted so that on July 1, 1927 there was only \$1,702,823,642 outstanding. The federal reserve note is a very elastic element in our monetary system and expands and contracts readily with the fluctuating demands of the country for a circulating medium, hence, the figures quoted will vary from month to month.

These, then, are the chief forms of paper currency now present in our monetary system. The only other form of paper currency is a small and insignificant item of \$1,326,804 of treasury notes of 1890. A more complete and critical study of paper money and of bank credit will be set forth in a subsequent chapter. At this point all that has been attempted is to give the student an acquaintance with the chief kinds of money that make up our monetary system.

3. *Subsidiary Coins.*—The third and last form of money composing our circulating medium is the subsidiary coins, sometimes called "token money."¹ The term applies to our pennies, nickels, dimes, quarters, half-dollars, and, also, at the present price of silver, the silver dollar. This form of money has been issued as a matter of convenience in effecting small transactions. There are literally thousands of transactions which would cause great inconvenience if it were not for the small coins. The first characteristic, then, of subsidiary coins is the fact that they are issued in small denominations. Second, these coins are either made from metal of a low bullion value and are circulated at a higher value, or, if standard metal is used, they contain less than their fractional part of this metal. Our pennies and nickels are made from inferior metals, and hence the coin value is much higher than the bullion value. In the case of silver, the dimes, quarters and half-dollars formerly contained the exact fractional amount of silver, viz., the dime was composed of one-tenth of the silver that was contained in the silver dollar, etc. Experience showed that, should the bullion value become greater than the coin value, the coins would be melted or exported and thus cause a shortage of coins of small denomination. In order to overcome this, Congress passed the Subsidiary Coinage Act of 1853, which reduced the bullion weight of these coins below the fractional parts of the dollar that they represented.² So long as the bullion

¹ Strictly speaking, token coins are those containing less than their normal value of metal and redeemable in standard money.

² The small coins contain 345.6 grains to the dollar. TAUSSIG, F. W., "Principles," Vol. 3, p. 268.

in these coins is worth more as coins than as bullion they will remain in circulation.

A third characteristic is that subsidiary coins are limited in the quantity issued. The government issues these small coins only when there is evidence of a need for them and then only in specified quantities. If the coinage were not limited there would be such a profit in coining bullion into subsidiary coins as to cause a redundancy of this form of money. In the United States these coins are kept at a parity with standard money by providing that they can be redeemed at the Treasury in standard money. Unless the amounts issued were limited, the Treasury would be unable to redeem all that could be presented. Finally, subsidiary coins have limited legal tender quality, now fixed in the United States at ten dollars for all the subsidiary coins except the five- and one-cent pieces, which are limited to twenty-five cents. The purpose of this provision is to prevent an individual from being burdened with this form of money and to prevent subsidiary coins from displacing standard money, which might happen if there were unlimited coinage and the coins had full legal tender quality. There is an active need for coins of this character and this need is protected by limited coinage, limited legal tender quality, and redeemability.

Legal Tender.—The various kinds of money constituting our monetary system have been described in the foregoing paragraphs. They all circulate and are generally accepted in the settlement of purchases, yet the legal obligation to accept these several kinds of money is not the same. Legal tender may be 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."¹ The forms of money which are legal tender in the United States are gold coins, silver dollars, treasury notes of 1890, United States notes,² and gold certificates. Silver certificates are not legal tender, but since they represent silver dollars, the practical effect is to make them equivalent to legal tender. Federal reserve notes, federal reserve bank notes, and national bank notes are not legal tender but are receivable by the government in the payment of public dues. The legal tender quality of subsidiary coins was given in

¹ FAIRCHILD, FURNISS, and BUCK, *op. cit.*, Vol. I, p. 381.

² The United States notes are full legal tender, except in payment of duties on imports and the payment of interest on the public debt.

the previous paragraph. In the absence of a contract calling for a particular kind of money, the legal tender law is controlling and the kinds of money here listed as legal tender can be demanded in the settlement of an obligation. Since most contracts are expressed in terms of dollars, the legal tender law is controlling in the vast majority of business transactions. In fact, business transactions are most commonly settled by means of checks or bank notes. The general acceptability of these forms of money in normal times is sufficient to meet the needs of business. But in an emergency, the legal tender law would authorize a creditor to demand in the settlement of debts due him, only those forms of money that have been made legal tender. He is thus protected against the acceptance of a depreciated currency.

The Value of Money.—Our next problem is to examine what determines the value of money. Since all other forms of money are kept either directly or indirectly at a parity with standard money, the problem consists in the explanation of what determines the value of standard money. In the United States this question means, What determines the value of gold? Since gold mining is an industry producing under conditions of increasing costs, one might conclude that the value of gold would be determined like that of any other commodity that was produced under similar circumstances, *viz.*, that its value would tend to coincide with the cost of producing the most expensive portion of the supply put on the market. But there are some peculiarities in the determination of the value of gold that require special consideration.

In the first place, gold is a highly durable commodity, and hence the annual production is always a relatively small proportion of the total existing supply. It was estimated that the world's supply of gold in 1900 was \$9,000,000,000, about half of which was in use in the arts. At the end of the calendar year 1926, the world's monetary stock of gold was estimated at \$9,621,645,000.¹ The annual average gold production between 1900 and 1920 was a little more than \$401,000,000 of which from one-third to one-fourth has gone into the arts and the remainder has been added to the world's monetary supply. In the case of most goods, the production of previous years has little or no effect on the value of goods currently produced. Many of them, such as food, clothing, etc., are consumed immediately and the

¹ Report U. S. Comptroller of the Currency, 1927, pp. 127-130.

only stock that affects the value is the supply which is flowing from the current sources of production. Even in the case of those goods which have greater durability, such as iron, steel, etc., only a small proportion of the existing supply has come from the production in previous years. Hence, the value of such goods tends to conform to the cost of production of the most expensive portion of the supply put upon the market. But not so in the case of gold. It is true that there is some loss sustained in the use of gold. There is, first, the wear and tear on that part that is used as money, which is small compared with the wear and tear on other metals, such as iron and steel. Then, that portion that goes into the arts is mainly lost for monetary purposes. While some old gold in the form of plate or jewelry is melted and added to the current supply, the amount from this source is relatively small. So far as Occidental countries are concerned, the drain of specie, both gold and silver, to the Orient is a permanent loss. The habits and customs in Oriental countries result in the use of enormous quantities of both of these metals. In the trade of these countries, metallic money is used almost exclusively and, in addition, large quantities are used for ornaments and for hoarding, so that for ages these countries have absorbed gold and silver in large amounts, which sums have been permanently lost to the gold supply of the rest of the world.

With respect to the losses in the arts, gold behaves like other commodities, the losses having to be made up from new supplies. But the situation is different in connection with the monetary uses of gold, especially in western countries. Here the loss is exceedingly small, being confined mainly to the wear and tear of the coins in actual use, which in many countries is greatly reduced by the use of paper money substitutes. Therefore, there has been an ever increasing stock of gold to which the annual production is added. The yearly addition to the monetary supply, even during years of greatest gold production, does not represent more than one-twentieth to one-thirtieth of the total existing stock. It will thus appear evident that instead of the cost of production determining the value of gold, it is perhaps more nearly correct to say that the value of the existing supply determines what costs can profitably be incurred in the operation of mines. For this reason, the determination of the value of gold presents a special problem, as it does not strictly

conform to the ordinary rule of value applicable to other commodities produced under similar conditions.

From the discussion above, it is apparent that there are two primary uses for gold, namely, its commercial use, or its use in the arts, and its use as money. Assuming the existence of free coinage of gold into money, we may then conclude that whenever it is worth more in the arts than as money, it will flow into the arts until its commercial value just equals its coin value. Hence, demand or marginal utility determines not only the amount of gold used in the arts but also the value of that portion so used. It is the marginal utility of the ring, or watch, or what not, that determines the value of the gold used in their production. Its value in such uses cannot exceed the estimate placed upon such goods by the marginal consumers. It should be observed, however, that gold products, as jewelry, are affected in a peculiar way by changes in the value of gold. For instance, when gold is depreciating in value, gold products will rise in price but not to the same extent as other commodities, for the reason that the gold content of the commodity falls in value along with all other gold and only the cost of manufacturing will rise in price, whereas the depreciation of gold will affect the prices of both the material and the manufacturing costs of all other commodities. With this exception, the value of gold products tends to conform to the general principles of value as applied to other commodities.

The real problem lies in the application of the principles of value to that portion of gold that is used for monetary purposes. The value of money can be expressed only in terms of the prices of other commodities. If a unit of gold will exchange for a large amount of other goods, its value is high and prices are correspondingly low, and *vice versa*. The problem, then, is to determine what makes the value of gold money high or low. At first it may appear that there is no new problem here, and that the value of money is determined like other goods by the principle of marginal utility and marginal cost. Further consideration will convince one that this principle does not fully explain the value of money. The peculiarity lies in what constitutes the demand for money.

In the case of consumers' goods, demand, as we have seen, means the quantity of those goods which the consumers stand ready to take from the market at a price. Other things remaining constant, fluctuations in supply will be followed by inverse fluctuations in price, but the change in price will not be in pro-

portion to the change in the supply. By doubling the supply the price will not be cut exactly in half because of the elasticity of demand, which differs with different goods. It would be correct to conclude, under the conditions specified, that the price of any good would fall, but unless supported by statistics based upon past experience, it would be unsafe to predict the amount of the decline. In the case of indirect or producers' goods this same principle holds true. An increase in their supply will cause a fall in value, but the decline will not be in exact proportion to the fluctuations in supply. They, too, are affected by the elasticity of demand.

But the situation is different in regard to gold used as money. Assuming that gold is the only money material in use, the demand for money at any given time may be said to be constant. It consists of the volume of exchanges that are being made in the market. Whether this volume be large or small, all of the money available to make exchanges at that time will be used for this purpose. This condition does not exist in connection with any other good, and it constitutes the peculiarity of the demand for money. There is no corresponding fluctuation in the use of money as its value changes, as is the case in connection with all other commodities.¹

It should be apparent from this discussion that the principles of diminishing and marginal utility do not govern the demand for money, as is the case in connection with the demand for other commodities. It should be clear that the demand for gold as money remains constant in the sense that the whole available supply will be used in performing the money work to be done at any given time. This peculiarity in the demand for money creates the special problem in the explanation of its value.

Proceeding now to a consideration of the more specific questions in the determination of the value of money, we should recall that it has already been stated that there is no method of expressing the value of money except in terms of the prices of the commodities it will buy. A gold dollar is worth what 25.8 grains of

¹ An exception to this general statement will have to be made in the case of the demand for credit. For some purposes, at least, the demand for credit is affected by the discount rate charged for its use, hence it is not strictly correct to say that all of the money there is in the form of credit will be used in making exchanges. This topic will receive further consideration in the chapter on Banking and Credit.

standard gold will purchase. If the value of this weight of gold is high, then it will exchange for a relatively large volume of commodities and commodity prices will be low; but if it will buy only a small volume of goods, then its value is low and commodity prices are high. In other words, the value of standard money can be expressed only in terms of the general level of commodity prices.

This statement of the case brings us up to the first principle of the value of money, *viz.*, assuming that all exchanges are made by means of standard money, *i.e.*, gold, its value will be in exact inverse ratio to its quantity. By doubling its quantity without changing the number of transactions to be effected, the value of money will be halved. Or, to put the case from the side of the exchanges, if the number of transactions to be effected by means of a definite quantity of gold should be doubled, the value of the gold would be doubled. From this statement, we conclude that, in accordance with the circumstances assumed, the value of gold used as money varies inversely with its supply and directly with the volume of exchanges to be performed. This is what is commonly known as the *quantity theory* of money.

The relation between the quantity and the value of a commodity, as here expressed, does not hold true for any other good, because in all other cases there is greater or less elasticity of demand. In some cases, doubling the supply of a commodity would cause its value to fall more than half because purchasers could not be found for the additional amounts, while in other cases an increase in supply might find a large number of purchasers at a slightly lower price. But in the case of money, all of the supply that is available for use in making exchanges will be so used. Hence, we say, that the value of money tends to fluctuate inversely with its supply.

A corollary that is deducible from this principle, one that has already been suggested, is that the general level of prices varies directly with the volume of money. Since the whole available supply of money tends to be used in effecting the transactions, then these exchanges can and will be effected on a plane of either high or low prices. This means that when the supply of money is relatively large to the volume of exchanges, the level of prices will be high, and *vice versa*. Buyers and sellers tend to adjust themselves to the general price conditions that surround them.

The student should note carefully the assumption upon which the above conclusion is based. It is assumed that all exchanges are made by means of standard money. But this assumption does not conform to the actual conditions of trading, hence, some modifications of this statement must be made. The first qualification is that money should be regarded as a *flow* rather than a *fund*. There is a flow of commodities into the market, but they are all eventually sold and are purchased by the flow of money into the market. This flow of money has been called the "rapidity of circulation" or the "turnover" of money. It is a matter of common knowledge that the turnover of money is more rapid in cities and urban communities than in rural districts. The faster the flow of money or the more frequent the turnover, the smaller the volume of money that will be required to perform a given amount of money work. If a dollar is turned twice in an interval of time in one instance and but once in another, the absolute number of dollars is the same in the two cases but the number of transactions is twice as large in one as in the other instance. Instead of this increase in the number of transactions causing an increase in the demand for money and a consequent fall in prices, as would be the case if the rate of turnover of money remained the same, the increase in the rapidity of circulation enables the transactions to be performed in the same interval of time and at the same level of prices. Hence, the rapidity of circulation has the same effect on the value of money as an increase in its supply. These relations may now be expressed by what is known as the equation of exchange.

Equation of Exchange.—In accordance with the assumption thus far presented we may say that the value of money can be expressed by the equation $MV = PT$, in which M equals the volume of gold money in circulation, V the velocity or rapidity of circulation, P the price level, and T the total number of transactions. It will be clear from this equation that if the quantity of money, interpreting quantity now to mean not only the actual number of dollars in use but also the rate at which these dollars are turning over, is increased without a corresponding change in the volume of transactions to be performed, the only effect that is possible is an increase in general prices.¹ This, then, represents our first modification of the quantity theory of money. Further

¹ FISHER, I., "Elementary Principles of Economics," Chaps. VIII-XII, gives a full treatment of this formula and the relation of money to prices.

qualifications will have to be made to take account of other forms of money, but if the exchanges were effected solely by means of standard money the above statement would explain fully its value.

Bimetallism.—All civilized countries use more than one form of money, so that our next question is to inquire into the effect of these different forms of money on its value. Next to gold, silver is the metal most commonly used for monetary purposes. For centuries silver was freely coined and circulated on an equal basis with gold as a monetary material, but during the nineteenth century most countries passed to a single gold-standard basis. Bimetallism is a monetary system in which two metals, as gold and silver, are equally recognized as standard money. The monetary system in the United States was at first bimetallic, as the unit that was adopted consisted of a definite amount of gold or silver to be coined at a fixed ratio of 15 to 1. Provision was made for the free and unlimited coinage of both metals. The subject of bimetallism has in the past been actively discussed both as a theoretical issue and as a question of practical policy. It has been argued that bimetallism will result in greater stability of prices because the value of the two metals does not always move in the same direction. The conditions of production are not the same, and should the value of the two metals move in opposite directions the effect would be to give greater uniformity and stability to general prices.

The opponents of bimetallism have argued that it is impossible to maintain the two metals in circulation on an equal basis because of the difficulty of maintaining a parity between the mint and market ratios. The mint ratio must be fixed while the market ratio fluctuates with changes in the supply and demand for the two metals. If the market ratio deviated from the mint ratio, then the only metal that would be presented for coinage would be the one that was overvalued at the mint. If the market ratio were 17 to 1 and the mint ratio were 16 to 1, silver alone would be presented for coinage. A monetary principle of general recognition and of long standing would begin to operate, *i.e.*, the cheaper money would tend to drive the dearer money out of circulation. This principle is known as "Gresham's law,"¹ and

¹ This monetary principle was named by H. D. Macleod for Sir Thomas Gresham, a London merchant of the sixteenth century, who assisted the Crown in financial difficulties for a number of years. Macleod gave this principle Gresham's name, under the impression that he first gave expression to it in 1558. It had been well expressed by earlier writers, especially Oresme and Copernicus.

it rests on the observation that the debtor, in case of an option, will use the cheaper of two kinds of money in making payments. The result is that only the cheaper money will remain in circulation and the dearer will be hoarded. In the case here assumed, silver, being the cheaper, will displace gold and become the monetary medium.

The advocates of bimetallism admit the difficulty of maintaining the parity between the mint and market ratios, but they argue that the adoption of a bimetallic standard tends to equalize the value of the two metals at the fixed ratio. They claim that whichever metal is overvalued at the mint will cause coins of the other to be melted and thrown on the market, thus tending to bring its market value to the mint ratio. This compensating action under bimetallism, it is claimed, will prevent the market ratio of either metal from departing far from the mint ratio. The natural result from such a condition, if it were accomplished, would be greater stability of prices.

International bimetallism, established as the result of treaty agreements among the leading nations of the world, would doubtless show much greater effects in this direction than if a double standard were undertaken by a single nation. But even if one of the leading nations should adopt bimetallism, the first effect would doubtless be to bring the market and mint ratios of the two metals closer together. As evidence in support of this conclusion, the experience in the United States in 1896 and again in 1900 may be cited, when, under the influence of the hope of the success of the 16 to 1 campaigns of William Jennings Bryan, the value of silver was actually enhanced temporarily and the mint and market ratios brought closer together. The ultimate effect, had this policy been adopted in the United States, would have been the displacement of gold by silver, for the higher mint value of silver would have stimulated its production and the cheaper money would have eventually supplanted the dearer.

Historically, it can be shown from the experience of the United States that this situation has actually happened more than once. In 1792 the ratio adopted was 15 to 1, but this ratio undervalued gold so that little gold was brought to the mint. In 1834, the ratio was changed to a trifle over 16 to 1, but this time the mint ratio was higher than the market ratio so that only a small amount of silver was brought to the mint and silver began to disappear from circulation. In 1837, the ratio was again changed, this

time to 15.988 to 1, a ratio that has continued to the present time. The discovery of gold in California and Australia brought great quantities of gold to the market and much of it was brought to the mints of the world. In the United States, silver began to disappear and there was a great scarcity of small coins. In order to correct the inconvenience occasioned by this shortage, the Subsidiary Coinage Act was passed in 1853, which reduced the amount of silver in these smaller coins below the fractional part which each coin represented. Immediately the small coins came back into circulation for there was now no inducement to melt them and sell them as bullion. The coin value was now greater than the bullion value. It thus appears from the experience in the United States that the two metals cannot be maintained in circulation on an equal basis at a fixed ratio because of the difficulty of keeping the mint and market ratios together. Whichever metal is overvalued at the mint has become the actual money in circulation.

But, aside from the feasibility of bimetallism, the use of both silver and gold as money does not change the monetary principle already stated. If silver were standard money alongside of gold, it would simply add to the quantity of money available to make exchanges. Silver has occupied a peculiar position in our monetary system for a long period of time. In the previous paragraph the effect of the changes in the mint ratio in 1834 and 1837 was pointed out. The fact that silver was worth more as bullion than as money after the ratio had been changed prevented it from being coined. This situation continued until 1873 when, in a general revision of the coinage laws, the standard silver dollar was dropped from list of coins to be freely minted. Between 1834 and 1873 practically no silver was minted except into subsidiary coins.

The "Limping" Standard.—As the legislation in 1873 did not modify the legal tender quality of the silver dollar, the effect was to establish what has been called a "limping" standard. The silver dollar could no longer be freely coined but it was legal tender in the settlement of both public and private debts. In the Gold Standard Act of 1900, gold was specifically made the standard money of this country but no change was made in respect to the legal tender quality of silver dollars. As there are over \$49,000,000 in silver dollars in our monetary system, they still constitute a considerable problem. The silver

dollar occupies this anomalous situation. It can no longer be coined, except by specific act of Congress, yet it retains its full legal tender quality, but it is not exchangeable by law with gold. The only provision bearing on the relation between the two metals is that portion of the Gold Standard Act which requires the Treasury to maintain all coins at a parity with gold. This provision and the fact of limited coinage have the effect of maintaining the silver dollar in circulation alongside of gold, even though the bullion content is not worth much more than half that amount. So much for the nature of the silver dollar as a part of our monetary system.

Returning now to the main question, namely, the forces determining the value of money, we can conclude that the use of silver can have but one effect upon the value of gold as money. In so far as silver is used as a substitute for gold, it tends to depress the value of gold. If all of the exchanges were accomplished by means of gold, it would require a larger volume to effect those exchanges on the basis of the existing price level. If silver were eliminated, the value of gold would appreciate and the price level would decline. The existing supply of silver money can therefore be regarded simply as an addition to the stock of gold money and hence it conforms to the principle already stated.

Irredeemable Paper Money.—There are three forms of paper money in circulation in the United States, *viz.*, gold and silver certificates, bank notes, and government notes, or greenbacks.¹ Gold and silver certificates are in reality simply warehouse receipts, certifying that a definite amount of gold or silver has been deposited in the Treasury of the United States. In this form, paper money is only a substitute for the gold and silver held in deposit in the Treasury. They constitute no special problem in determining the value of money, as they simply represent an equivalent amount of gold and silver for which they circulate. The several varieties of bank notes, however, do constitute a special problem, the treatment of which is postponed until the next chapter, after the whole question of bank credit as a form of money has been explained. This leaves for present treatment only the government notes, or promises by the government to pay. The greenbacks, issued during the Civil War, are the

¹ The Treasury Notes of 1890 are omitted here because their volume is insignificant.

principal form of money of this character in this country. In the past, most countries have had experience with an irredeemable currency, but at no time in the world's history have countries gone to the extent in issuing this kind of money as during the World War, particularly in Germany, Austria, and Russia. In February, 1923, Germany had in circulation the sum of 3,500,000,000 paper marks while Russia in the same month had paper rubles in circulation to the enormous amount of 3,200,000,000,000,000. These figures convey some impression of the incredible extent to which the issue of government paper money may go when a country is pushed to the limit of its existence.

The problem with money of this character is to determine what gives value to these promises of the government to pay, and in answering this question we shall use the experience of the United States with the greenbacks. The cheap money advocates of the Civil War period (and it may be recorded that advocates of paper money have used similar arguments in the past in many countries) argued that all that was necessary to give value to any commodity used for monetary purposes was the will or fiat of the government. If the government stamped a piece of paper and said that it was a dollar, it would be accepted at its face value in exchange. Many people still believe that the government stamp is all that is necessary to give value to a commodity, although the experience of Russia and Germany has given most advocates of fiat money a shock from which they are not likely to recover for a generation at least. Notwithstanding the recent experiences, we know that irredeemable paper has been and is used as money, so that the question as to what determines its value is pertinent.

There are several conditions which will make paper currency desired even though it is not redeemable in standard money. First, if the government will receive it in payment of taxes or other obligations, the currency will have some value. Then, if it is made a legal tender in settling private obligations, its value is still further enhanced. With these conditions present, and if issued in limited quantities, promises of the government to pay will circulate, even if there is no obligation to redeem them in standard money. In the United States, the greenbacks were given full legal tender value except in payment of custom duties and in the payment of interest on the public debt. Thus, one of the influences which makes paper money acceptable was present. A second cause is that of limited supply. If the

volume issued does not exceed the normal needs of business, or, putting the same idea in another form, if the volume is not so great as to distort the normal exchange relations, irredeemable paper money will be accepted alongside of standard money and will circulate at a parity with gold. It simply adds to the volume of money in circulation, thus increasing the supply of money in relation to the volume of transactions to be performed. While such money will circulate within the country issuing it, it will not be accepted in the settlement of obligations with other countries.

It is sometimes argued that paper money should be adopted because it is inexpensive to issue and its volume does not depend upon the exigencies of discovery and of production as in the case of gold and silver; also because its volume can be kept stable through control exercised by the appropriate governmental agency. However, experience has demonstrated conclusively that it is much easier to expand than to contract paper currency. Expansion permits larger public expenditures, and temporarily money issued by the government will take the place of taxation. In addition, the larger volume of money will tend to raise prices and thus act temporarily as a stimulus to business, whereas, a contraction of the currency will mean economy with possibly increased taxes on the part of the government, and falling prices with the tendency for business to be curtailed. Few government officials have either the courage or the vision to contract currency under pressure of this character, and history has shown repeatedly that it is much easier to expand than to contract paper money. If the quantity of such money is not large, the business of a country may grow up to the volume outstanding and it will circulate on a parity with standard money. Growth in the volume of business transactions was certainly one of the influences that resulted in the appreciation of the greenbacks and thus made the resumption of specie payments easy in 1879. The adoption of an irredeemable paper currency because it is cheap is fraught with consequences that history has shown repeatedly to be disastrous to the country issuing it. The recent experiences of Germany, Austria, and Russia should serve as illustrious examples of the evils of cheap money, and dissuade other countries from embarking upon a policy of irredeemable paper currency.

We have in the United States at the present time slightly in excess of \$346,000,000 of greenbacks, against which the Treasury

holds a reserve of \$150,000,000 in gold. The difference between these two sums is credit money and represents a net addition to the supply of money in the country. We may ask two questions concerning this money. First, how is its value determined? And second, how does its existence affect the value of gold? In answer to the first question, we may say that its value is dependent upon its legal tender qualities plus the fact of a limited supply. Its utility exists because of its use in making settlements, which in turn is possible because its supply is limited. People will accept greenbacks now, whereas, if the quantity were being constantly increased, they would discriminate against their use. In answer to the second question, we may reply as in the case of silver. The sum in excess of the gold reserve held against the greenbacks represents a net addition to our total money supply and therefore tends to depress the value of gold. If all exchanges now effected by this form of money were effected by gold, it would enhance its value, so that the presence of this sum in excess of the gold reserve has the effect of depreciating the value of gold. Thus, in the case of both the silver and the greenbacks we need to make no essential change in our general principle of money, recognizing that the supply of these two forms of money represents a net addition, except as noted, to the volume of money in circulation.

Suggested References

- BYE, RAYMOND T., "Principles of Economics," Chap. XI.
 CARLILE, WILLIAM W., "The Evolution of Modern Money."
 CONANT, CHARLES A., "The Principles of Money and Banking," Book I, Chaps. I-IX.
 EDIE, LIONEL D., "Economics, Principles and Problems," Chap. XXV.
 ELY, RICHARD T., "Outlines of Economics," 4th Ed., Chap. XIV.
 FAIRCHILD, FURNISS, and BUCK, "Elementary Economics," Vol. I, Chaps. XXI and XXII.
 FISHER, IRVING, "Elementary Principles of Economics," Chap. VIII.
 FOSTER and CATCHINGS, "Money."
 GIDE, CHARLES, "Principles of Political Economy," Rev. Ed., Book II, Chaps. I-IV.
 HAWTREY, R. G., "Currency and Credit."
 HAYES, H. GORDON, "Our Economic System," Vol. I, Chaps. XXVI-XXXI.
 HOLDSWORTH, JOHN T., "Money and Banking," 3d Ed., Chaps. I-VI.
 JEVONS, W. STANLEY, "Money and the Mechanism of Exchange."
 RUFENER, L. A., "Principles of Economics," Chaps. XXV-XXVII.
 TAUBENIG, FRANK W., "Principles of Economics," 3d Ed., Vol. I, Chaps. XVII-IX.

CHAPTER XII

BANK CREDIT AND THE CIRCULATING MEDIUM

Thus far in our discussion of the value of money we have been concerned primarily with standard money, except in so far as attention has been given to silver and government paper money. If paper money is directly convertible into standard money, it acts simply as a substitute for standard money as is the case with gold and silver certificates, the silver represented by the certificates being kept at a parity with gold by limited coinage and by indirect convertibility. The case is different with inconvertible paper. Its value depends upon its quantity. If the supply be limited and the demand for money great enough, it may circulate at par with gold as do our greenbacks. This form of currency introduces into the circulating medium a credit element which in this case is a credit of the government.¹ These credit instruments pass from hand to hand and perform money work, and to that extent they have the same effect on the value of money as an increase in the quantity of standard money. But private credit in this country is much more important as a means of effecting exchanges than government credit. It will now be our task to examine into the nature of credit and to see how credit instruments may be made to perform money work and thus to economize in the use of standard money.

Nature of Credit.—Credit is a relation that involves both a creditor and a debtor. A debt is an obligation to pay money or other valuable considerations; a credit is a right to receive money or other valuable considerations. Wherever there is a credit there is also a debt, and these relations are usually expressed in terms of money. For instance, I may buy a bill of goods and give my promissory note in terms of dollars on the date of sale. This note does not pay for the goods but is merely paper evidence that I have agreed to pay money, or some other

¹ The credit element in the case of greenbacks is limited to the difference between the total volume in circulation and the gold reserve maintained against them.

valuable goods in terms of money, at some future date. To the creditor, the note is evidence of the right to receive money; to the debtor, an obligation to pay money. This simple but very common example emphasizes the twofold relation that is involved in every credit transaction. The primary difference between a credit transaction of this character and a cash transaction is the introduction of the time element. Credit transactions of the character described are merely a postponement of the use of money.

Kinds of Credit.—Credit exists in two principal forms, *personal credit* and *bank credit*. The simplest form of personal credit is "book credit," or a "charge account." When we buy groceries daily and have them charged, a book credit is established. It is perfectly evident that such purchases are merely a postponement of the use of money. At the end of two weeks or a month it becomes necessary for the purchaser to pay money. Credit in this form does not increase the volume of transactions that are made in the interim of time covered by the credit. At the time the purchase is made, the right to buy "on credit" has the same effect on the supply and demand for commodities as if money were used in payment, but when the obligation has to be met, the actual payment of money has the effect of reducing the possible transactions by the same amount as the credit purchases increased them. Thus for the entire credit period, purchases on charge account do not enlarge the volume of business done. If this were the only form in which credit appeared in our economic transactions, its use would have no appreciable effect on the value of money.

Personal credit exists also in the form of the promissory note. As indicated above, I may give the storekeeper a written promise to pay on demand, or at the expiration of a definitely stipulated time, a specified amount. The mere fact that the note is given adds nothing to the credit relation, and if the store keeper holds the note till its maturity the situation is identical with book credit. The promissory note bearing the signature of the purchaser makes it easier to collect, if any dispute should arise concerning the obligation. But the note is a negotiable instrument and may be transferred to a third party. Suppose the storekeeper has an opportunity to buy additional goods to replenish his stock and instead of paying cash offers to sign over (endorse) this note in payment thereof. If the seller is willing to accept it, then this

credit instrument, a right to receive money, has performed money work. When the note is due, instead of paying the original storekeeper in whose favor it was drawn, I am now obligated to pay the present note holder. My obligation is no greater than before and the sum of money which I have to exchange is the same, but during the credit period, by means of transferring the note, twice as much money work has been performed.¹ If the note was for \$100, my money not only pays for the goods which I purchased but, by means of transferring the note, another bill of goods of equal amount was purchased. In this way, the credit instrument has performed money work and has the same effect as an increase in the quantity of money.

Banks and Bank Credit.—A promissory note of an individual will not be accepted generally because the drawer is not likely to be well known and his ability to pay is uncertain. Consequently, the amount of money work that can be done by a note of this character is very small. But a note that is issued by an institution that is widely and generally known may be readily accepted in payment of obligations. In fact, such a note may circulate from hand to hand in the payment of debts just as standard money. This brings us to the subject of banks and bank credit. Classified in accordance with their functions, there are several types of banks. The three main types are *investment banks*, *savings banks*, and *commercial banks*. Investment banks are institutions that are normally engaged in selling stocks and bonds that have been issued by industrial concerns. Such institutions underwrite, *i.e.*, agree to finance those business undertakings that require long-time loans. They examine into the nature of a proposed business venture and determine whether a loan will be made. The probable success of the venture is the controlling factor. If it is decided to underwrite the business, the investment banker agrees to take the stocks and bonds of the company, which he then sells to his customers. A firm such as J. P. Morgan & Company, or Halsey Stuart & Co., may be taken as illustrative of this type of banks.

A savings bank is an institution that collects the savings ordinarily from persons with medium or small incomes. Such

¹ In this illustration, it is assumed that the note is accepted at its face value. It usually happens that such notes are discounted, but this fact will modify the general statement only to the extent that the total addition to the money supply is the face of the note less the discount.

institutions pay the savers a low rate of return on their deposits, which sums are then usually invested in safe securities that yield a higher rate. Since security is of greater importance than income to the depositors, the operation of these banks and the kinds of investments they can make are very strictly regulated in most states. Most commercial banks also operate a savings department, and so familiar are these departments that it is unnecessary to describe them here in more detail.

Commercial banks are those that are engaged in lending their credit mainly to business men for short periods of time. The nature of their functions will be explained more fully below.

Two other financial institutions should be mentioned in this connection, namely, insurance companies and trust companies. Insurance companies collect from individuals enormous sums in the form of premiums on insurance policies. They invest these sums in various types of securities, and thus constitute a very important influence in the general investment and money markets.

Trust companies perform a number of functions, the most important of which are to act as administrators of estates and to invest and care for trust funds. They perform, also, other similar functions, which need not engage our attention at this time. Many of the commercial banks operate trust departments through which large sums are controlled. Both the insurance companies and the trust companies are very important financial institutions and play a significant part in the monetary and financial operations of modern business.

But we are concerned here primarily with the operation and effects of commercial banks which are institutions that serve two main purposes. They furnish a place of safe deposit for idle money and lend their credit, *i.e.*, transfer personal credit into bank credit. A bank may lend its credit in two ways. It may issue bank notes, which are merely promises of the bank to pay on demand the face value of the note. Except for the general regulations governing the issuing of bank notes, the only difference between an ordinary promissory note and a bank note is that the former is drawn in favor of a specified person or firm and requires his (or its) endorsement before it can be accepted by a third party, while a bank note is made payable to the bearer on demand. Both are private credit instruments but the bank note will usually pass freely in the payment of obligations.

The second way a commercial bank may lend its credit is by means of what is sometimes called a "created deposit." When a customer deposits a sum of money in a bank, the bank assumes the responsibility to pay back the full amount on demand. The item designated as "deposits" in a bank statement is, therefore, a demand liability of the bank to pay cash. But the customer may, instead of depositing money, ask for a loan. Let us suppose that he wishes to borrow \$1,000 for six months. If he is known at the bank, he may secure the loan by simply signing a promissory note which obligates him to pay its face value plus the interest when due. This is what is known as "single-name" paper. Or, the bank may require him to have an endorser, *i.e.*, someone who will sign the note besides the borrower. In the latter case, both persons become liable to pay the note, and such notes are sometimes called "two-name" paper. In other instances the borrower may deposit bonds or other securities as a pledge that the loan will be paid when due. Such a note is collaterally secured and the bonds or securities pledged are called "collateral," and should the borrower fail to pay the loan when it is due, the securities can be sold by the bank.

In whichever form the loan is taken out, the borrower will have to sign a note which obligates him to pay the full \$1,000, plus the interest, at the expiration of six months. Let us suppose the bank charges 6 per cent interest, or, as we shall soon see, discounts the note at that rate. The borrower will receive in cash the \$1,000 less the interest on this sum for six months at 6 per cent or \$970, or he may have this sum added to his deposit account against which he can draw checks which the bank will honor in sums not in excess of this amount. It should be evident in this instance that the deposit arises out of a loan made by the bank and that the borrower can draw checks against this deposit account in payment of his obligations in precisely the same manner as he does when he has made a deposit in standard money. Thus a bank may lend its credit either by issuing bank notes or by created deposits as just described. We shall see shortly how bank credit performs a vast amount of money work and constitutes an enormous addition to the circulating medium.

Discount.—In the paragraph above, the charge made by the bank for the use of its credit was called "interest" or "discount." The difference between a discount and the charge for a loan is more a matter of procedure than it is of essence. Suppose Mar-

shall Field & Co., of Chicago, sells a bill of goods to Lord's Department Store on 60 or 90 days' time.¹ Lord's Store would likely give Marshall Field & Co. a promissory note for the amount of the purchase. Marshall Field & Co. might then hold this note until it was due and collect direct from Lord's Store, or the company might take the note to a bank and discount it at whatever rate the bank charges for carrying notes of this character. If the note is discounted, Marshall Field & Co. would have to endorse it, thus becoming liable to the bank should Lord's Store fail to pay the obligation when it becomes due. If the rate charged were 6 per cent, the bank would discount the principal at that rate for the time the note had to run and pass the discounted amount to the credit of Marshall Field & Co., keeping the discount for its service in carrying the note.

One might ask why Marshall Field & Co. would care to accept less for the note than its face value, if the firm was certain that the full amount would be paid at maturity. This query can be answered by stating that so long as the merchandising profit of Marshall Field & Co. is greater than the charge or discount rate made by the bank for the service rendered, it will be advantageous to the firm to discount such notes. The firm will have immediately available a deposit credit at the bank which it can use to better advantage in restocking its shelves with other goods. Notes such as these are known as "commercial paper." There is an enormous volume of business done in this way. A large number of smaller banks in industrial and commercial centers regularly purchase commercial paper of this character, for by so doing they can arrange to have their idle funds invested and at the same time arrange these investments in such a manner as to have them maturing day by day and thus enable them to meet their own demand obligations as they are presented.

But how, it may be asked, does discounting or the lending of bank credit affect the volume of the circulating medium? When Marshall Field & Co., for instance, draws a check against the deposits that result from a discounted bill, the effect is the same as if a promissory note were transferred in payment of a purchase, except that in this case Marshall Field & Co. can by means of the check divide the payment made into amounts that will suit the convenience of the store, whereas the promissory note is transferred at its face value. The bank not only holds the note until

¹ This is a department store in Evanston, Ill.

due but it transfers the personal credit instrument of Lord's Store, which could have but limited circulation, into bank credit which can be used in convenient form in the payment of other obligations. In this way bank credit is used to perform money work.

The Bank Statement.—But this is not a complete explanation of the case, and to understand the relation of bank credit to the volume of money it will now be necessary to examine the simpler operations of a bank. For this purpose a simplified bank statement is here presented:

Assets		Liabilities	
Loan and Discounts.....	\$150,000	Capital Stock.....	\$100,000
Stocks and Bonds.....	35,000	Surplus.....	25,000
Furniture and Fixtures...	10,000	Deposits.....	120,000
Cash.....	77,000	Undivided Profits.....	2,000
		Notes.....	25,000
	\$272,000		\$272,000

An examination of any bank statement will show in some form the items given above. It will be well to examine these items and understand their essential character. On the side of assets there appears the item, "loans and discounts," which represent rights to receive money, or amounts due the bank; "stocks and bonds" represent investments of the bank. If a bank has idle money which it cannot lend to individuals or firms, it may purchase securities which will yield an interest return. Then, in addition, the bank will have some money tied up in "furniture and fixtures," the necessary tools for conducting its work. Finally, there must be a sum of money or cash which is essential to meet the demand of its customers for a circulating medium. The cash held may for the present be regarded as "reserves." The reserve item is the amount of lawful money held by the bank to meet its demand liabilities in the form of deposits. It is out of this item that all promises of the bank to pay cash on demand are met.

The liabilities of a bank consist of the "capital stock," "the surplus," the "deposits," "bank notes," and "undivided profits." One might at first wonder why the capital stock of a bank is a liability rather than an asset. Since a bank is usually a corporation, the capital stock represents what the bank as a legal entity owes to its stockholders who have paid into the bank an amount of money represented by the par value of the stock which each one

holds. In addition, it is quite common for a bank to create a surplus at the time of its organization as additional strength. The surplus in such cases also comes from the payments made by the stockholders and is, therefore, a liability of the bank. After a bank has become established, further additions to the surplus are made by setting aside from the item, undivided profits, a sum that is determined by the board of directors. If the earnings of a bank are high, the directors may decide to keep part of the earnings in the business, and the customary way of doing this is to pass the sum so determined to the surplus account, which enhances the investment of the stockholders.

The item "undivided profits" represents the earnings of the bank from charges made for making loans, from discounting, or from any other service that the bank may render. It is from this item that dividends are declared. Next are the "notes," which are the promises of the bank to pay on demand and are issued in accordance with the regulations governing note issue. The notes are a liability of the bank to the note holder. The conditions under which notes may be issued will be explained later. Finally, the item "deposits" represents the liability of the bank to its customers, or to those who make deposits with it. As has been explained already, deposits may arise either from a deposit of money, or from an extension of a loan by the bank to a customer, or by discounting a bill for him. In any case, the bank obligates itself to pay cash on demand to the amount of its deposits and notes outstanding. Thus, it will be seen that in the simple form of bank statement given, the liabilities of the bank are either to the stockholders or to the customers and that the liability to the latter is always a demand liability. The bank must always be prepared to pay money on demand to note holders and to depositors. A few simple examples will illustrate this type of bank operations.

Suppose that the above bank makes a loan of \$1,000 for a period of six months at 6 per cent. How would the bank statement be affected by this transaction? First, the loans and discounts would be increased by the full amount of the loan; the undivided profits will be increased by the amount of the interest for the time the loan runs; and finally, the way the other items in the statement will be affected will depend upon the disposition of the loan by the customer. If he wishes cash immediately, the cash item will be drawn down by the amount of the loan less

the interest, but if he wishes to increase his deposits, he can have the discounted amount added to his deposit account. The statement will then stand in either one of two ways, depending upon the choice of the customer.

Assets		Liabilities	
Loans and Discounts.....	\$151,000	Capital.....	\$100,000
Stocks and Bonds.....	35,000	Surplus.....	25,000
Furniture, etc.....	10,000	Deposits.....	120,000
Cash.....	76,030	Notes.....	25,000
		Undivided Profits.....	2,030
	\$272,030		\$272,030

or

Assets		Liabilities	
Loans and Discounts.....	\$151,000	Capital.....	\$100,000
Stocks and Bonds.....	35,000	Surplus.....	25,000
Furniture, etc.....	10,000	Deposits.....	120,970
Cash.....	77,000	Notes.....	25,000
		Undivided Profits.....	2,030
	\$273,000		\$273,000

The corrected statements show the items as they would appear if the customer drew out cash, in the first place; or, if he accepted a deposit credit at the bank, in the second. Whenever he exercises his right to draw checks on his account, the cash may be drawn down by the amount of the checks, or if the persons receiving the checks are also customers of this bank, the transactions will be registered in the bank by mere transfers from one account to another. If A draws a check in favor of B, and B deposits the check, then A's account will be diminished and B's increased by the amount of the check. The bank statement as a whole will remain unchanged by such transactions.

It should be noted that the volume of cash which the bank holds is much less than its demand liability. Reverting to the original statement, it will be seen that the bank has only \$77,000 in cash, whereas, the notes and deposits amount to \$145,000. If all of these liabilities were presented in one day, the bank would be unable to meet its obligation to pay on demand. In fact, there is no bank, however conservatively operated, that could meet its demand liabilities if they were all presented at one time. Bankers have found by experience that all of their customers do not call for their deposits at once, and out of a total

volume of deposits only a certain percentage will be drawn out each day, the amount depending upon the habits of the people and the industrial conditions of the community served. The rapidity with which checks are presented is much greater in an urban and commercial center than in a rural and agricultural community. By arranging its loans so that they will mature day by day, a bank can have a flow of payments into it that will equal the demand for money. The recognition of this fact has led bankers to extend credit in the form of loans or discounts, until their liability to pay cash equals the payments due them, with such a margin as experience has shown is a necessary protection against unforeseen demands.

In this way a dollar in the bank is capable of doing from \$7 to \$13 worth of money work. This can best be shown by reverting again to the original bank statement and asking, How much can this bank safely loan? We shall see later that the amount of reserves a bank must carry is regulated by law, but at this point let us suppose that the bankers are following past experience which has shown them that they must carry a reserve of 10 per cent, *i.e.*, they must have \$1 in cash for every \$10 in deposit liability. Under these circumstances, the bank may safely make additional loans up to \$650,000 providing they are arranged so that they mature day by day. Assuming that it is safe for them to operate on the basis of a 10 per cent reserve and that each loan runs for six months and the rate of discount is 6 per cent, then the corrected statement will stand as follows:

Assets		Liabilities	
Loans and Discounts.....	\$800,000	Capital.....	\$100,000
Stocks and Bonds.....	35,000	Surplus.....	25,000
Furniture, etc.....	10,000	Deposits.....	750,500
Cash.....	77,000	Notes.....	25,000
		Undivided Profits.....	21,500
	\$922,000		\$922,000

By comparing the ratio of cash with the deposits, it will be seen that the bank has slightly more than 10 per cent of its deposit liability in cash and, if past experience has shown that this is adequate, it will then be safe for the bank to extend loans to this amount. Since bankers are in business to make profit, and since most of their profits come from the lending of their credit, they will extend credit to this amount, providing their

customers have need for such a volume of loans.¹ Someone may ask, How can a bank loan more than it has? The answer is that this is the distinctive function of the banking business, namely, the turning of personal credit into bank credit, and as long as there is a demand for bank credit, a bank can continue to discount or lend its credit until its demand liabilities begin to encroach upon its cash reserves. Assuming the bank has made no mistakes in making its loans, *i.e.*, in the choice of personal credits that it has assumed, it has ample resources which, given time to allow the loans to mature, will enable the bank to meet all of its obligations. From these few simple illustrations the reader should perceive how important bank credit is in performing the business transactions of a country like the United States, and how bank credit in the form of deposits and circulating in the form of checks adds tremendously to the volume of money.

Clearing Houses and Check-off System.—The methods of economizing in the use of money have been developed by the banks still further by means of a *clearing house*. A clearing house is an association of banks for the purpose of exchanging checks held by each against members of the association. Suppose, for instance, that three banks, A, B, and C are serving a local community. During the course of the day checks drawn against accounts in each bank will find their way into the other banks so that each bank will hold claims against the other banks. Let us suppose that bank A owes B \$10,000 and C \$5,000; that bank B owes C \$12,000 and A \$8,000; and bank C

¹ The relation of reserves to deposit liability is stated boldly here in order to show how important bank credit is in our circulating medium and how this bank credit comes into existence. The statement of this relation needs some qualification to bring it more nearly into accord with the way in which additional credit is actually extended. The statement as presented would not hold for an individual bank, as is implied, but it would hold for the system of banks. If, for instance, any individual bank should receive a deposit of a definite sum of gold, say \$1,000, it would probably be unable to build up its deposit liability through an extension of loans by ten times this sum. The reason for this is, that checks drawn against its "created deposits," *i.e.*, those resulting from the loans, would doubtless be followed by a loss in the reserve item. While an individual bank would probably not be able to extend its credit by the amount specified in the body of the text, the deposit of the \$1,000 in gold would enable the banks of the country, as a whole, to expand credit to the extent fixed by the legal ratio of reserves to deposits. For further treatment of this phase of credit, the reader is referred to Chester A. Phillips, "Bank Credit."

owes A \$13,000 and B \$6,000. It would be possible for each bank to pay the other banks its full obligations in cash, which would require A to pay out \$15,000, B \$20,000, and C \$19,000. Instead of making payments in cash, the banks meet and *clear* the checks, or *check off* the accounts held against each other and pay in cash only the balances due.

In the cases cited the accounts of the three banks will stand as follows:

A	A	B	B	C	C
Credits	Debits	Credits	Debits	Credits	Debits
\$ 8,000	\$10,000	\$10,000	\$12,000	\$ 5,000	\$13,000
13,000	5,000	6,000	8,000	12,000	6,000
\$21,000	\$15,000	\$16,000	\$20,000	\$17,000	\$19,000
Credit balance	\$ 6,000	Debit balance	\$ 4,000	Debit balance	\$ 2,000

By the process of clearing checks instead of requiring \$54,000 in money to settle the claims against each other, *i.e.*, the combined sums owed by the three banks, they can adjust their claims with \$6,000. The method shown in this simple illustration is representative of what takes place on a much greater scale in all large centers like New York, Chicago, and Boston. In the larger cities the banks have organized clearing-house associations for the purpose of clearing checks, and banks that are not members of the association arrange to clear through some bank that is a member, so that in this way the total business of a city is transacted by means of a minimum of actual money changing hands.

Some impression of the significance of credit money as a part of the total money of the country may be obtained from the following figures: The total stock of gold money in the banks and public treasuries of all the countries of the world at the end of the calendar year 1926 was given by the United States Comptroller of the Currency as \$9,621,645,000.¹ This figure does not include the amount in circulation, but, except for a few countries as India, the United States, Canada, and Australia, no large amounts of gold are outside of the public treasuries and the central banks. Making due allowance for this item, it probably would be liberal to estimate the total world stock of gold money

¹ Report of U. S. Comptroller of the currency, 1927, p. 130. The Federal Reserve Board estimated the holdings of 40 countries in 1926 as \$9,180,000,000. This figure does not include gold in circulation, *i.e.*, the sums held by commercial banks, business concerns, and private individuals. *Federal Reserve Bull.*, April, 1927, p. 276.

at from \$10,000,000,000 to \$11,000,000,000. Turning to the statistics of the United States, which are more accurate, we find that on July 1, 1927, the Comptroller of the Currency reported the total stock of all kinds of money in this country as \$3,531,991,679, and of this sum \$4,565,000,000 was in gold coins and bullion, \$537,000,000 in silver dollars, and \$295,000,000 in subsidiary silver, or a total of \$5,397,000,000 metallic money. On the same date the total deposits of the 27,061 state, private, and national banks, including 1,656 trust companies, were \$51,100,000,000.¹

It would thus appear that the deposits in the banks of this country alone exceed the world's supply of gold money by \$41,500,000,000, or are more than five times that supply. Put in another way, these deposits are more than three times the total world production of gold from 1860 to 1925, inclusive.² But before we have the total volume of bank credit in use in this country there must be added the stock of bank notes outstanding. In 1927, there were \$2,077,473,195 federal reserve notes, \$4,854,238 federal reserve bank notes, and \$704,146,267 national bank notes, or a total of \$2,700,000,000 in bank notes which, together with the deposits, would bring the total bank credit of the country to \$53,800,000,000. These figures make it perfectly clear how important bank credit is in our total circulating medium.

Bank Credit and the Value of Money.—The significance of credit instruments in making exchanges has an important bearing upon the value of money. We have already seen that silver and paper money have the same effect on its value as an increase in the supply of standard money. In the same way, bank credit, whether in the form of deposits or bank notes, has the effect of increasing the total volume of purchasing power in terms of money. Because of this fact, the equation of exchange must now be restated to include bank credit as a part of the money medium of a country. Let M' equal the total volume of bank credit and V' the rapidity of the turnover of bank credit, then the equation of exchange will be $MV + M'V' = PT$. The total volume of money now means the total quantity of gold times the rate of its turnover, plus the total quantity of bank credit times its rate of turnover. The quantity theory may now

¹ The comptroller of the Currency estimates that returns were had from about two-thirds of the private banks, but those reporting were doubtless the larger ones, so that complete figures would not modify very greatly those given. Cf. Report of U. S. Comptroller of the Currency, 1927, p. 113.

² The total world production of gold from 1860 to 1925 inclusive was in round numbers \$15,500,000,000.

be stated as follows: The value of money varies inversely with the total quantity of money in circulation. This volume consists of the total stock of standard money, together with both the silver and irredeemable paper money and the amount of bank credit in terms of money.

If the number of transactions increases at the same time and at the same rate as the quantity of money, as defined above, the effect will be to prevent the value of money from falling. It may happen that bank credit will expand as a result of an expansion in transactions, in which case the expansion of credit may be regarded as the result of an increase of business rather than as an independent cause. Some have argued that bank credit expands only as business expands and, therefore, bank credit can be disregarded as an influence on the general price level and that the value of money depends solely upon the quantity and turnover of gold. This point of view leaves out of account the way in which credit may be used to expand the total amount of purchasing power in terms of money without increasing the volume of commodities. Suppose A and B are manufacturers and they believe that market conditions are favorable for increased production. Suppose, also, that the banks have idle reserves, i.e., they may safely make additional loans. If each of these men secures a loan of \$10,000, the extension of these loans, disregarding the discount, increases the purchasing power by \$20,000. The banks need to hold, in support of this volume of credit, from \$1,400 to \$2,600 in gold as reserve, depending upon the class of city in which the banks are located. The mere fact of making these loans does not in itself increase the supply of labor or the quantity of commodities available for productive purposes, so that A and B begin to bid against each other to control the available supplies of both labor and commodities in order to carry forward their plans for increasing production. The result is that there is relatively more purchasing power in terms of money than there is of services and commodities. In this way, an extension of credit tends to raise the price level. An increase of bank credit used as indicated has the same effect upon the value of money as an increase in its quantity.¹

¹ When there is no demand for credit for general commercial and industrial purposes and there is "idle credit," it seeks the stock market and security prices rise. Put in another way, the market value of the securities is determined by a lower rate of capitalisation. The rise of security prices during 1927-1928 can be explained very largely by this influence.

Relation of the Volume of Gold and Bank Credit.—Another question of importance is, What relation, if any, is there between the volume of gold and bank credit? Is an increase in the volume of gold followed by an increase in bank credits? To answer this question completely would require a careful analysis of statistical data on the volume of gold, total deposits, and the general level of commodity prices, that would carry us beyond the range of present purposes, but the simple relationship can be briefly stated. Additional supplies of gold, whether from imports or new gold produced at the mines, is divided between monetary and industrial uses in the ratio of about 3 or 4 to 1. The additional volume going into monetary uses will, unless specifically withheld by the government, find its way into the banks and result in an increase of gold reserves. High reserve ratios mean idle money to the banks, which in turn means reduced earnings. As a means of stimulating borrowing, therefore, the discount rates are usually lowered when the reserve ratio is high. Increased borrowing will result in increased bank credit in the form of deposits which can be used, as shown in the illustration above, as a means of securing labor services and commodities to further the processes of production. The competitive bidding of borrowers will work in the direction of higher prices. Students of this question have been able to show that there has been in the past a close correspondence between an increase of gold reserves and the volume of bank deposits,¹ thus giving faith to the conclusion that additional supplies of gold will result in an expansion of bank credit, even though the expansion may not be immediate nor in exact proportion to the legal provisions for reserves.² The effect of this increase of both M and M' is to

¹ ELY, R. T., "Outlines," p. 295, footnote.

² There are sharp differences of opinion on this question and the student who is interested should follow the general theories of money, banking, and credit given here by a more comprehensive study of the subject. It has recently been held that the United States had received large gold imports from Europe and, at the same time, commodity prices were falling and these two facts were presented as proof that other influences than the increase of gold were responsible for price movements. But it has been suggested in reply that the stock and bond prices have, during this same interval, moved upward at a time when a reverse movement was generally expected and, furthermore, the banks have greatly increased their investment in securities during this same time. This evidence supports the general theory stated in the text. The beginning student should recognize that these are complex questions that require careful marshalling and interpreting of facts before

enhance prices. Higher prices require more money to effect a given volume of business.

The sequence seems to be as follows; more gold makes for larger bank reserves, larger bank reserves stimulate lending and an increase of bank deposits which result in more purchasing power without a corresponding increase in commodities and services. The extension of bank credit stimulates competition for the control of commodities, and services, and this competition results in higher prices. The higher prices require a larger volume of money in circulation to sustain these prices and business tends to become adjusted to this higher price level. In this way an increase of gold and bank credit tends to lessen the value of the whole volume of purchasing power in terms of money.

Suggested References

- ANDERSON, B. M., "The Value of Money," Chaps. VI-IX, XII, XIII and XIV-XVIII.
- BAGEHOT, WALTER, "Lombard Street."
- BYE, RAYMOND T., "Principles of Economics," Chap. XII.
- CONANT, CHARLES A., "The Principles of Money and Banking," Book II, Chap. I-V.
- DUNBAR, C. F., "Chapters on the Theory and History of Banking."
- EDIE, LIONEL D., "Economics, Principles and Problems," Chap. XXVII.
- ELY, RICHARD T., "Outlines of Economics," 4th Ed., Chap. XV.
- FAIRCHILD, FURNISS, and BUCK, "Elementary Economics," Vol. I, Chaps. XXII-XXIV.
- HAWTREY, R. G., "Currency and Credit," Chaps. I, III, V, VIII and XII.
- HAYES, H. GORDON, "Our Economic System," Vol. I, Chap. XXX.
- HOLDSWORTH, JOHN T., "Money and Banking," 3d Ed., Chaps. VII, X, XIII and XVI.
- KEMMERER, EDWIN W., "The A B C of the Federal Reserve System."
- PHILLIPS, CHESTER A., "Bank Credit."
- RUFENER, L. A., "Principles of Economics," Chaps. XXVIII-XXIX.
- TAUSSIG, FRANK W., "Principles of Economics," 3d Ed., Vol. I, Chaps. XXIII-XXVII.
- WILLIS, PARKER, "The Federal Reserve System."

dogmatic judgments can be given. The author is of the opinion that the general theory given above will best serve as the basis for further study along these lines. Such qualifications as are necessary should be made only after the student has gone farther than is possible in this course into both the facts of banking and the theories of bank credit.

CHAPTER XIII

THE BANKING SYSTEM OF THE UNITED STATES

A complete picture of the part that bank credit plays in modern economic life requires some understanding of the present banking system. In order to appreciate these conditions fully, it will be well to trace the main steps in the development of our present system of banks. Prior to the Civil War, two attempts had been made to establish central banks in this country, namely, the First Bank of the United States, which secured a charter in 1791 and continued until 1811 when it failed to be rechartered; and the Second Bank of the United States which was chartered in 1816 and continued in operation until 1836, when it likewise failed to be rechartered. Both of these banks served useful purposes and were a great aid in stabilizing the banking and currency conditions of the periods during which they were operating. A great deal of opposition developed against the Second Bank on the ground that its officers were playing politics and were showing favoritism in the extension of credit. In addition, there were those affected by the prevailing sentiment who feared the monopoly power which such an institution might exercise. These and other causes were sufficient to swing public opinion against the bank, and it failed to be rechartered in 1832, when it was made a distinct political issue. President Jackson and his followers were vigorous opponents of the bank, while Henry Clay and his friends championed its cause and fought for a renewal of the charter. The failure to recharter was a great stimulus to the development of state and private banks, as can be seen from the following table on page 306.

The number of banks increased in the interval, 1829-1838, from 329 to 829, or 250 per cent; circulation increased more than 300 per cent to the peak in 1837, and loans by more than 380 per cent at the same date. These figures reveal the rapidity of the growth of bank credit that took place during these years.

THE GROWTH OF BANKS AND BANK CREDIT 1829-1838¹

Year	Number of banks	Note circulation in millions	Loans in millions
1829.....	329	\$48.2	\$137.0
1834.....	506	94.8	324.1
1835.....	704	103.7	365.2
1836.....	713	140.3	457.5
1837.....	788	149.2	525.1
1838.....	829	116.1	485.6

¹ DEWEY, D. R., "Financial History of the United States," p. 225.

State and Private Banks.—The state and private banks that were organized so rapidly at this time were regulated, if at all, by state laws. Many of the states, particularly in the South and West, had very lax laws governing bank operations. As a result of inadequate regulation, these private banks issued large volumes of notes and put them into circulation, often at considerable distances from the point of issue. Because of the difficulty of travel and communication at this time, these notes were not promptly presented for redemption, but continued in circulation for a long period of time. They were used in the speculative buying of land during the early part of the decade of the 'thirties, and contributed to the rapid rise of commodity prices which culminated in the serious financial and industrial crisis of 1837.

In some sections of the country, as in New York and New England, local methods were developed which made it necessary for the banks to redeem their notes regularly, so that notes of many of these banks "were as good as gold," while in other sections, where effective methods for forcing redemption were not developed, the notes were worthless. No uniformity of control over note issue was developed prior to the Civil War, and the situation was so bad that there were at that time some 7,000 different kinds of bank notes in circulation and some 5,500 of these were spurious, altered, or counterfeit.¹ Even good notes might not circulate at par, so that it became necessary to publish "bank-note reporters" and "counterfeit detectors" as a means of protection against fraud and loss due to the bank notes then in existence. Imagine a business man of today having to consult

¹ DEWEY, D. R., "Financial History of the United States," p. 322.

a price list when a customer offered a bank note in payment for a bill of goods in order to know at what rate he could accept the note, if at all! This was the situation that existed prior to the passage of the National Bank Act in 1863 which established for the first time a general system of banking in this country.

National Banking System.—In 1863, the country was in the midst of the Civil War. The finances of the federal government were in straitened circumstances. It was under these conditions that the National Bank Act was passed. Among other things to be accomplished, Secretary Chase wished to develop a market for the bonds which the government found necessary to issue in order to find funds to prosecute the war. The act establishing the national banks required that each bank should invest one-third of its capital in government bonds. Funds so invested would not be available for general banking purposes unless some special provisions were enacted to take account of this fact, consequently, the law provided that a bank could deposit these bonds with the Secretary of the Treasury and issue bank notes up to 90 per cent of the current market value, but not exceeding 90 per cent of the par value of the bonds deposited.¹ This enabled the banks to secure interest on the government bonds, and at the same time to earn interest on the loans of the bank notes as well.

In 1866, the national banks were given what amounted to a monopoly of bank-note issue by the imposition of a tax of 10 per cent per annum on the note issue of all state and private banks. From this date until the passage of the Federal Reserve Act in 1913, we had in this country a uniform bank note, circulating from one end of the country to the other. Instead of the condition existing prior to the passage of the National Bank Act, when such a large variety of bank notes were in use, we had but one form of bank note. In addition to uniformity there was established at the same time security to the users of bank notes, which added greatly to the confidence in banks and bank credit. The note holder was protected by the deposit of government bonds with the Secretary of the Treasury to the full amount of the notes outstanding. The notes were as good as the credit or ability of the government to meet its obligations. The development of a uniform bank note that was safe and acceptable in all

¹ The banks were permitted by the Monetary Act of 1900 to issue notes to the par value of the bonds.

parts of the country was a great improvement in the banking conditions of that time.

The National Bank Act has been amended many times, but from 1900 to 1913 it stood as follows: The stockholders must contribute a minimum capital, varying according to population from \$25,000 in the smaller places to \$200,000 in cities of more than 50,000 people. The notes must not exceed the capital stock, and government bonds equal in amount to the notes issued must be deposited with the United States Treasurer as a protection to the note holder,¹ and each bank must maintain in the Treasury of the United States a redemption fund in lawful money of 5 per cent of the note issue. The banks were also required to maintain a minimum reserve which varied according to the class of city in which a bank was located. The cities of the country were classified into three groups,—central reserve cities, consisting of New York, Chicago, and St. Louis; reserve cities or those with a population of 50,000 or more, in which three-fourths of the national banks therein had applied to be so designated; and other cities or towns. The national banks in the central reserve cities had to maintain in their own tills a reserve of 25 per cent of their deposit liabilities. Banks in the reserve cities had to maintain a 25 per cent reserve against their deposit liabilities but could redeposit one-half of this sum with the banks in central reserve cities. Finally, the banks in the third class of cities had to maintain a 15 per cent reserve, three-fifths of which might be redeposited in banks of either the reserve or central reserve cities.

The Strength and Weakness of the System.—The national banking system was a great improvement over the banking conditions that had previously existed. It gave a safe and uniform bank note that would circulate throughout the country. The note holder was amply protected by the deposit of government bonds. The supervision that was given to bank operations made bankers more careful concerning their investments and extension of loans. But notwithstanding the merits of the system it had certain serious defects that the country was slow to admit.

¹ An amendment after the Civil War required that banks with a capital of more than \$150,000 had to invest \$50,000 in government bonds, while banks with a capital of \$150,000 or less had to invest at least one-fourth of their capital in these bonds.

Inelasticity of Note Issue.—First, the note issue was inelastic. The need of the business of a country for money is not for a constant sum. It fluctuates from season to season and from year to year. During some seasons, as when the agricultural crops are being harvested and marketed, farming sections need large volumes of money, but during other parts of the year the demand is much less. Business activity generally is much greater during some years than in others. Many factors affect this need for money which an effective banking system should be able to supply through an extension of credit in the form of bank notes.

For a currency to be elastic, banks should be in a position to expand their notes when the commercial and industrial needs of the country demand it. Since this demand is first felt at the bank where customers regularly register their needs, the initiative for expanding bank notes should rest with the banks. To be truly elastic, currency should contract as readily as it expands. An effective system should provide bankers with a motive to retire their notes as readily as they issue them. If provision is made for rapid redemption of the notes by the issuing bank, they will not remain in circulation beyond the period of actual need but will be deposited in a bank as soon as the need for currency declines. If the bank is unable to reissue the notes or to put them into circulation again, the currency will contract as the business needs of the country fall off.

The national bank notes were not elastic, either in the direction of expansion or contraction. As explained above, the banks were required to invest part of their capital in government bonds and few bankers were satisfied with the return on these bonds. The only way they could increase the earnings on this part of their resources was to deposit these bonds, take out bank notes, and loan their credit in the form of bank notes. In this way, the total earnings of the bank could be increased. But beyond this amount, any additional extension of notes depended quite as much upon the price of government bonds and the rate of return from them as upon the commercial needs of the country, so that the motive to expand notes in accordance with business needs was not as potent as it should have been. Then, once the notes were paid out, they tended to remain in circulation, as the inducement to retire them was insufficient. Banks were inclined to pay out these notes and keep gold and greenbacks in their tills as these forms of money were lawful reserves. When once

issued therefore, the national bank notes tended to be a permanent part of the circulating medium. In addition, the total volume of notes could not exceed the amount of government bonds that had the circulation privilege, which fixed a maximum limit on the sum that could be issued and, as the government paid off its obligations, it reduced by that amount the bank-note currency that could be outstanding. From the point of view of the commercial needs of the country, this was a serious defect in the system.

The Use of Surplus Reserves.—The second defect involved the use of the surplus reserves. The law provided that the banks should maintain a minimum of reserves, but it permitted reserve city banks and country banks to count as reserves certain amounts on deposit in banks in reserve or central reserve cities. Whenever a bank had more money than it could loan to its local customers, it would transfer these sums to its correspondent banks in reserve and central reserve cities, i.e., to banks in other cities with which it regularly dealt. These sums so transferred were always subject to be drawn upon whenever the local bank had need for additional funds but, in the meantime, they would be earning something even though the rate paid on such deposits was small. This provision resulted in idle money flowing into the main financial centers and especially into the larger banks in New York City.

Three Monetary Drains.—To understand the effect of the piling up of these reserves on the whole financial structure of the country, it will be necessary to explain how certain financial problems of the country caused definite monetary movements and to show how these movements affected the institutions into which the surplus reserves tended to flow. There were three main drains upon the money markets of the country, the burdens of which fell most heavily upon the New York banks. First, was the demand for crop-moving purposes. The country has always had a large agricultural product. The South and West, and particularly the Central West, are the regions in which farming is the predominant industry, and agricultural crops are by nature seasonal. The marketing of these products annually causes a heavy drain of money into these sections of the country. Beginning about mid-summer and continuing during the fall months, the banks in the farming sections were drawing down their balances in the larger banks in financial centers and particularly in New York. This

drain of money for the moving of crops always caused a strain on the New York money market.

The second drain was to the United States Treasury. Taxes are usually paid at stated intervals. Tariffs are paid when goods are imported and importation is heavier during some seasons of the year than during others. Then, the payments by the government are not uniform. Certain expenses are fairly stable but interest payments and expenditures for public works are not constant sums. The important point to note is that there was not a constant flow of money either into or out of the Treasury. From the date of the founding of the Independent Treasury in 1846, the revenues of the government were kept out of the channels of trade and were returned only by disbursements of the government in payment of its obligations. During some periods, the Treasury accumulated large surplus funds, which were withdrawn from the money markets of the country. These payments to the government caused a withdrawal of money from the banks, and this drain also fell heavily upon the New York banks.

Finally, the settlement of international balances also constituted a monetary drain. As will be shown later, in the course of trade between countries, there come times when there must be a movement of specie in the settlement of international trade and, in this case, the money of international exchange is gold. These payments are not constant. The ordinary settlements are made by bills of exchange, and if imports just balance exports there is no need for the movement of money. But it is impossible for a perfect balance to be maintained and, as a result, one country becomes indebted to another. Gold is then demanded to make these settlements and the banks are called upon to provide the money for this purpose, with the result that the reserves of the banks engaged in this form of business were drawn down. These three drains, for crop moving, to the Independent Treasury, and for the settlement of international balances were, under the National Bank Act, a source of continual trouble. If any unusual financial problem developed at the time that the banks were adjusting themselves to these drains, there might, and usually did, result a serious financial stringency. The banks of New York into which the surplus reserves of the country flowed bore the shock of the financial strain due to these and other causes.

The Banks and the Speculative Market.—Since the surplus reserves transferred to correspondent banks were subject to call, they were likely to remain but a short time in the New York banks. As a consequence there were few opportunities for the investment of such funds. To keep their assets liquid, it was necessary for these banks to carry a large amount of loans in such form as would enable them to meet these drains on short notice. Hence, the banks were inclined to loan heavily on call. The principal demand for loans of this character comes from speculators operating on the stock exchange. In this way the banks of the larger centers, and especially of New York, became more closely identified with the speculative markets than was wholesome for the country as a whole. All of these facts have a bearing on subsequent developments.

A Decentralized System.—A third defect of the national banking system was the lack of cooperation among the banks. A few simple facts concerning banking operations will make the significance of this statement apparent. Bankers are in business to make profits. Their profits come from discounting bills and from making loans, i.e., from transferring personal credit into bank credit. The resources of a bank are made up largely of bills and notes for which the bank has exchanged its credit. These credit instruments have varying intervals of time to run, but the banks aim to arrange the loans so that the daily payments into the bank will equal the daily demands on the bank. As the banks under the National Bank Act were required to maintain a definite reserve ratio, and since it was to their advantage to keep loans up to the full legal limit, no bank would carry a larger surplus reserve than was required. The reserves usually exceeded the legal limit only at times when the business community had need for a small amount of credit. As a bank's business increased in volume, the ratio between its reserves and demand liabilities tended to decrease. This occurred from two causes. First, an increase in loans and discounts was generally followed by an increase in deposits, as many borrowers accepted deposit credit instead of drawing money from the bank. This increase in deposits at once decreased the reserve ratio. Then, if one bank loaned more rapidly than others, the volume of checks drawn against it exceeded those drawn against other banks, and it soon found that it had adverse balances at the clearing house which tended to draw down its reserves. In this way the banks

of the country tended to expand their loans to the full legal limit, whenever the business conditions of the country demanded additional credit.

If, at such times, any unusual conditions occurred that required a larger volume of money than was normally needed, the banks would find their reserves withdrawn and there was no method of replenishing them. One bank could not rediscount for another however good the bills and notes held by it might be, nor could the banks in one part of the country come to the relief of those in another. Each bank had to assume its own responsibilities, which fostered periods of rapid change in bank accommodations. Whenever a bank saw its reserves being depleted, the only thing it could do in self-defense was to raise its discount rate, thereby retarding the desire to borrow and, if this was not sufficient, then shut down on further extensions of credit and allow its bills and notes to mature. It might resort, as a last measure, to the sale of the bonds and other securities which it held, but such a method would not be used unless the strain was very great, and even then the bank would likely lose heavily, for security prices would fall rapidly at such periods of financial stringency. The most effective way to protect itself at such times was to curb sharply the extension of credit, which would help the bank but would not cure the strain for the community affected. The decentralization in the system made deposit credit inelastic at times when the country needed a greater volume of personal credit transferred into bank credit. There was neither the concerted action that is needed to discourage undue expansion of business, nor was there a possibility of pooling the banking resources of the country to carry the financial stress that was so frequently experienced. In fact, the banks in self-defense had to follow a practice that enhanced rather than relieved the financial pressure that accompanied unusual periods of monetary need.

Proposed Remedies.—The defects mentioned here were repeatedly brought to the attention of the people of the country, but not quite so forcibly as in the panic of 1907. The failure of a large company in this year started the banks to curbing their credits and finally brought on a country-wide financial panic. Up to this time most of the proposed remedies had to do with the securing of greater elasticity in note issue. It was thought that if the banks could issue notes more freely it would be easy to satisfy the demand for additional currency. If the people

found that they could get money when they wanted it, they would not want it in unusual quantities. Fear, it was thought, caused the runs on banks, and these runs depleted the reserves. If the banks could easily increase the volume of bank notes, this demand for money would soon be satisfied. Those who argued thus advocated that notes be issued on other securities than government bonds. They held that if the banks could transform their discounted bills into bank notes, the needs of the country could be more satisfactorily met. But inelasticity was not confined to notes, as has been shown, as a part of the difficulty was the rigid reserve provisions and the lack of any system of cooperation among the banks, which also made deposit credit inelastic.

Some students of banking advocated the establishment of a central bank along the lines found in almost every European country. These banks hold the reserves of the country, are the depositories of the government funds, and usually have a monopoly of note issue. The temper of the people of this country was not favorable to the establishment of a central bank. In many quarters the belief was firmly held that Wall Street, which was the synonym of the moneyed interests of the country, controlled not only credit but also manipulated commodity prices. To establish a central bank with a monopoly of note issue and controlling the banking reserves of the country, meant to many persons the tying up of the banking system with the speculative interests of the New York money market. Closely associated with this belief, and as a means of freeing the balance of the country from the effects of the "money trust," it was argued that each section of the country should have its own supply of money and credit. In the presence of such widely accepted views, together with the opposition inherited from the days of the Second Bank of the United States, the movement to establish a central bank could make little headway. The most important step in this direction was the recommendation of the National Monetary Commission.

The panic of 1907 led to a wide spread use of clearing-house certificates, issued by the clearing houses in the larger centers, like New York and Chicago, as a temporary means of relief. These certificates were virtually the joint obligations of the banks composing the clearing house, and were issued to individual member banks upon the deposit of adequate security. In the

main, these certificates were used to settle balances between the banks, but in some communities (Chicago) they were issued in small denominations and passed into general circulation. The large department stores and public service companies in Chicago advertised that they would accept these certificates from their customers in payment of purchases. It should be noted that these certificates were extra-legal, *i.e.*, they were not a part of our regular monetary system provided by law. They were accepted and circulated, however, because of the shortage of the regular forms of money.

As a result of this panic, Congress passed the Aldrich-Vreeland Act, in 1908. Provision was made by this act for emergency currency in times of financial stress. The national banks were permitted to issue notes on other security than government bonds, but such issues were heavily taxed so that the banks would retire them as soon as the emergency had passed. This same Act provided for a National Monetary Commission to investigate and recommend a plan of general revision of the banking system. After a comprehensive study of the banking systems in all leading countries, and an investigation of the special problems in the United States, this commission recommended a system of coordinated banks with a National Reserve Association to serve as a central bank for the whole system. The system was to be controlled by the banks rather than by the government because, it was argued, the property invested in the business was bankers' property and therefore bankers should have the controlling voice in determining the use to which their property should be put.

Because of the wide prevalence of the belief that the banks should be removed as far as possible from the speculative markets in New York and that each section of the country should be independent in its own supply of money and credit, the plan of reform proposed by the National Monetary Commission failed of enactment. The country had become convinced, however, that a general revision of the banking system was imperative, and the discussion of this plan did much to inform the public concerning the weaknesses of the old system and of the simpler phases of banking principles. The discussions tended to crystallize the opinion of the country as to the general direction which the reform should take. A presidential election brought a different political party into control and it became necessary for the

leaders of this party to develop a definite line of policy on this question. It was, therefore, not until 1913 that Congress passed the Federal Reserve Act. The following year the Federal Reserve system was put into operation.

The Federal Reserve System.—The two popular beliefs held concerning the operation of banks and the nature of money and credit, mentioned above, had their influence upon the new system. Instead of a central bank, the country was divided into twelve districts, and a central bank, known as a federal reserve bank, was provided in each district. The system is under the general supervision and control of the government. The central authority is the Federal Reserve Board, composed originally of seven, but now of eight, persons of whom the Secretary of the Treasury and the Comptroller of the Currency are members *ex officio*, and the other members are appointed for long terms by the President. This board has large powers over the reserve banks, having authority not only to examine their accounts but to remove officers and directors, to require them to rediscount bills for one another, to suspend reserve requirements, and to exercise general supervision.

The Reserve Banks are Bankers' Banks.—The stock of these reserve banks is owned by the local banks in the respective districts. National banks are required by law to be members of the system or lose their national charters and other banks can join under certain specified conditions. Each reserve bank is governed by a board of nine directors, six elected by its member banks and three appointed by the Federal Reserve Board. The business of the reserve banks is mainly with the member banks. They do not accept deposits from an individual nor do they lend directly to individuals. Their relations to member banks may be set forth under the three following heads:

Reserve Provisions.—First, the reserve banks now hold the reserves for the member banks. Although at first member banks were required to hold part of their reserves in their own vaults, since 1917 all reserves have been kept in the reserve banks in the form of deposit credit. The old reserve classification of cities has been retained in the law, but new reserve provisions were provided and the banks must now maintain 13, 10, and 7 per cent of their demand deposits as reserve, according to the class of city in which the bank is located.¹ A reserve of 3 per cent must be

¹ St. Louis has been removed from the central reserve city group.

kept by all banks against time deposits. The reserves, then, are now held by the reserve banks themselves. The only money that a member bank needs to keep in its tills is just enough to meet the daily demands of its customers. Thus the reserves have become much more highly centralized than under the old National Banking law.

The law requires the reserve banks to hold in gold, or lawful money, not less than 35 per cent of their deposit liability, which arises from the discounting of loans and notes for the member banks. In addition, these banks have to carry a 40 per cent reserve in gold against the federal reserve notes. However, the Federal Reserve Board has authority to suspend these reserve provisions temporarily, imposing a progressive tax upon the amounts by which the reserves fall below the legal limit. It is expected that the reserve banks will maintain a much larger reserve than that required. Thus, by centralizing reserves and by maintaining surplus reserves, it was thought that greater elasticity in this form of currency would be obtained and that one of the serious evils of the old system would be removed.

Rediscounting.—Second, the law provides that the reserve banks may rediscount for member banks certain types of notes or bills of exchange that have been drawn for agricultural, industrial, or commercial purposes. Notes or bills growing out of trading in stocks, bonds (other than government bonds), and other investment securities cannot be rediscounted. This provision was to prevent the reserve banks from being used to finance speculative dealings. The way the rediscounting provision works can be illustrated as follows: Suppose a bank found that its reserve ratio was falling below its legal limit. It can take out of its own loans any paper that is legally rediscountable and present it to the reserve bank. The reserve bank will discount this paper at the rate that it is charging for this service, pass the discounted amount to the deposit credit of the member bank, and credit its own profit account with the amount of the discount. If a member bank has a greater need for a circulating medium, it can request federal reserve notes (to be described shortly) instead of accepting a credit at the reserve bank. In this way, a member bank can continue to meet the needs of the community which it serves, and the credits received as the result of the rediscount can be used to maintain the reserves of the member bank, or the notes can be used to satisfy the demand for a circulating

medium. Of course, expanding loans by the member banks will cause the reserve bank to hold a larger volume of gold or lawful money against its own obligations. The reserve bank can discourage borrowing by member banks by raising the rediscount rate, or by bringing pressure on the member banks to reduce their credits at the reserve bank, which in turn makes it more difficult for their customers to borrow. This kind of pressure usually takes the form of requiring customers of member banks to reduce their indebtedness to the banks.

The federal reserve system has been criticized because it has not used its power over general credit conditions by raising its rediscount rate above the market discount rate, when its own reserves are approaching the legal limit. It is argued that so long as the member banks can loan at rates higher than the reserve bank charges for rediscount, it will be profitable for the member banks to continue expanding their loans quite irrespective of the fundamental needs of the country. In this way it is held that the federal reserve system has contributed to credit inflation. Without entering into a full discussion of the merits of this controversy, it is apparent that there is a real danger at this point and the student should get clearly in mind that easy credit does not always mean sound financial conditions. The extension of credit itself does not increase the volume of commodities and may mean simply an addition to the total volume of money in circulation and thus cause higher prices. But, notwithstanding this danger, the rediscount provision has made the credit facilities of the banks much more elastic than formerly, and even though this fact causes new problems it has corrected one of the serious evils of the old system. For further consideration of these problems the student is referred to the literature on banking and credit, particularly that portion dealing with the operation of the federal reserve system.

Note Issue.—Third, the law provides for two new kinds of notes, *federal reserve bank notes* and *federal reserve notes*. It will be remembered that the national bank notes required the deposit of government bonds before they could be issued. To supplant these notes by an entirely new issue would have caused undue hardship. The bonds against which these notes were issued all bore very low interest rates, so that one factor supporting their market value was the privilege of using them as a basis for issuing these notes. If this privilege had been withdrawn, the price of

the bonds would have fallen sharply and the resulting loss would have fallen mainly on the banks that owned them. Recognizing the force of this argument against the retirement of the old bank notes, Congress made provision for a gradual retirement of the notes and at the same time provided that the member banks could sell their bonds to the reserve banks at par plus accrued interest. This provision protected the banks against loss of value in their investments in government bonds and paved the way for the issuance of the federal reserve note.

There was fear, however, that there might be a too rapid reduction of the number of national bank notes, so it was provided that the reserve banks could use the bonds that they had purchased for issuing federal reserve bank notes, which are in fact essentially like the old national bank notes, except that the decision to issue them now rests with the reserve banks. During the World War there was a large volume of federal reserve bank notes put into circulation, but these have been gradually withdrawn until there is now less than \$5,000,000 outstanding. It was the purpose of the framers of the Federal Reserve Act that eventually all of the old bank notes would be withdrawn leaving only the new type of note that will now be described.

Of much greater significance is the provision for the federal reserve note. These notes are issued by the Federal Reserve Agents on the security of eligible paper and gold which the federal reserve banks have purchased or rediscounted for member banks.¹ When this rediscounted paper matures, it must be replaced by paper similar in character, or by gold, if the notes are kept in circulation. In this way, certain of the loans and discounts that grow out of the normal commercial and industrial operations of the country become the basis for issuing these new notes. The kinds of paper that are acceptable for this purpose are specified in the law. Thus, instead of a bank note based upon government bonds, the available amount of which is likely to be

¹ The federal reserve notes have back of them at least 100 per cent of high-grade collateral. "This collateral may consist of: (1) Paper endorsed by member banks and drawn for strictly commercial, industrial, or agricultural purposes; or for the purpose of carrying or trading in securities of the United States Government . . . (2) Bills of exchange endorsed by a member bank and banker's acceptances bought by the federal reserve bank in the open market. (3) Gold and gold certificates." In addition, with a minor exception, the federal reserve bank must maintain not less than a 40 per cent gold reserve against its outstanding notes.

definitely fixed over any period of time, we have a note based upon certain specified assets of the member banks, the amount of which fluctuates with the industrial and commercial activity of the country. The member banks can in this manner change their assets in the form of legally discountable paper, into a form of bank note that circulates from hand to hand.

Although the federal reserve notes are not issued by the government, they are guaranteed by it and are made a direct "obligation of the United States," and are redeemable at the holder's option at the United States Treasury. In addition, the reserve banks must hold a gold reserve of 40 per cent of the notes outstanding and the notes are given priority over other claims against the banks issuing them. These provisions give the note holder ample security, so that the notes are readily accepted in settling private obligations. While they are not legal tender, they are receivable in the payment of federal taxes. These notes now constitute the largest single item in our actual circulating medium, amounting on December 31, 1928, to \$1,808,034,000.

It was shown above how a member bank could get relief through the reserve system when its reserves were running low. Since a bank lends its credit both in the form of deposits and in notes, it now remains to show how a bank can get aid when its customers are demanding a larger volume of circulating notes. If a member bank should find that its supply of "till" money was running low, it could rediscount some of its eligible paper and take the proceeds in the form of federal reserve notes which it could pay out to its customers on demand. So long as its customers call for a money that will circulate from hand to hand, a member bank can satisfy that demand by transferring its rediscountable paper into federal reserve notes.

Thus in either of these two ways, that is, by taking the rediscounts in the form of deposit credit with the reserve bank, or by requesting federal reserve notes, the reserve system has made it easy for member banks to keep their credit in the form most convenient for their customers. The notes, like deposits, tend to be retired as soon as customers find it advantageous to reduce their bank indebtedness. In the two ways described, bank credit has become much more elastic than under the national banking system. The consolidation of the gold reserves of the country and the rediscounting of notes have given much greater

security against sudden financial crises, as well as greater flexibility in meeting the banking needs of the country. Since the member banks do not have to maintain reserves in their own vaults, most of the gold and gold certificates of the country have been gathered into the reserve banks. The result has been to centralize the gold or standard money, so that it now can be used more effectively in supporting the whole credit structure of the country.

Other Functions of the Federal Reserve Banks.—One of the difficulties that was regularly experienced, prior to the adoption of the federal reserve system, was the withdrawal of money from the ordinary channels of trade into the Independent Treasury. The Federal Reserve Act attempted to remedy this situation by making it possible for the government to deposit any or all of its funds in the reserve banks. So completely has the Treasury Department made use of the reserve banks as depositories for public funds that it closed the subtreasuries in 1920. This is evidence of the efficiency with which the reserve banks have performed these functions for the government. These banks may act, also, in other capacities as the fiscal agents of the government. During the war the Treasury found that the reserve banks could be used to great advantage in supporting the governmental financial policy. They became the medium through which government bonds and certificates of indebtedness were sold to other banks and to individuals. In these and other ways the reserve banks perform important financial functions for the government.

The reserve banks are also authorized to perform certain open-market transactions. They may buy and sell gold coin and bullion, securities of the Federal government, and short-time obligations of state and local governments.¹ While the rediscounting privilege is limited to member banks, there are no restrictions designating from whom the banks may buy or to whom they may sell commercial paper that is eligible for rediscount. The difference between purchasing a bill of exchange and rediscounting consists more in the effects than in essence. Member banks usually accept the proceeds of rediscounted bills in deposit credit or in notes, while purchases of bills from firms not members of the reserve system result in a direct drain upon the reserves of the federal reserve banks. These open-market operations of the federal

¹ Federal Reserve Act. Sect. 14.

reserve banks are a controverted issue. Many bankers argue against the policy of the reserve banks in entering into competition with member banks for this paper. Those who favor the policy of the reserve banks in continuing this form of credit dealings argue that in this way the reserve banks can exercise a steadying influence on the general credit and monetary situation of the country. It is held that many strong banking institutions are not members of the reserve system, that there is no well-organized bill market, and that changes in the rediscount rate are not a sufficient influence to guide credit dealings, especially since many banks may at times carry enough reserves in their own vaults to affect the general market conditions. Rediscounts must wait upon the initiative of member banks, and they will not undertake to rediscount until the demand for credit exceeds the supply. Credit may by this time have become so expanded as to make the general financial conditions unsound. But if the reserve banks on their own initiative can buy and sell, not primarily for profit but to relieve a financial strain and to discourage too rapid expansion, they may exercise a steadying influence on the whole credit situation of the country. Whatever may be the outcome of the discussion on this phase of the policy of the federal reserve banks, there can be little difference of opinion on the merits of the system as compared with the banking conditions under the National Bank Act.

Before leaving this subject, emphasis should again be placed upon the effect of the federal reserve system upon bank credit and through bank credit on the monetary and price problems of the country. We have already seen how much money work is performed by means of bank credit. The reserve system, by means of the provision for rediscount, has made bank credit much more elastic than it was before the passage of the Federal Reserve Law. Much of the personal credit that grows out of normal business transactions and is accepted by the member banks can now be rediscounted at the reserve banks and be made the basis for issuing bank notes, whenever there is need for additional circulating medium. This new feature, together with the concentration and efficiency in the use of gold as reserve, has made it possible for the banks of the country to support a larger structure of credit than was possible under the old banking laws. This greater credit structure may have an important influence on the general level of prices. It should be observed that $M'V'$

is not only more elastic than formerly, but may increase faster than the volume of transactions. In this way the additional credit may have the effect of lowering the value of money by increasing its supply faster than is necessary to perform the exchanges of the market. Should this expansion of credit occur, the effect will be shown in higher commodity prices.

But, while this effect is possible, the system has clearly made for greater stability of prices and has thus created more favorable conditions for the general operation of industry. There is now no necessity for sudden and sharp fluctuations in the conditions of the banks which formerly caused recurrent periods of strain in the financial and monetary markets of the country. The resources of individual banks need not now become so tied up that the demand of their customers for a circulating medium cannot be easily satisfied, unless the banks have followed an unwise policy in extending their loans. So long as the resources of the reserve system have not been unduly expanded, the banks can accommodate their customers and furnish them with sufficient funds to carry on their industrial and commercial transactions. This has been a great accomplishment to the credit of the reserve system which is a marked improvement over the situation under the National Banking Law.

However, it is incumbent on the Federal Reserve Board to keep its eyes open to the general business situation and not to permit a gradual overexpansion of credit. So long as there is a possibility to extend credit, *i.e.*, so long as there are idle reserves, member banks will be disposed to make loans, if they can find safe personal credit to discount or other borrowers seeking bank accommodations. There are always some persons willing to borrow for the conduct of their operations, providing the banks will lend and the rates charged are not too high. Because of the desire on the part of member banks to invest their idle funds and the willingness of some to borrow, there is always the possibility of an undue expansion of bank credit. When such conditions arrive, there is still the possibility of financial panics and resulting industrial depressions. This danger, however, is much less now than formerly, and so long as the Federal Reserve Board exercises reasonably intelligent judgment in the development of its policy, such periods of financial and monetary strain should be largely prevented.

Suggested References

BURGESS, W. RANDOLPH, "The Reserve Banks and The Money Market."

DUNBAR, C. F., "Chapters on the Theory and History of Banking."

HEPBURN, A. BARTON, "History of Currency of the United States."

HOLDSWORTH, JOHN T., "Money and Banking," 3d Ed., Chaps. XI, XXI, and XXII.

KEMMERER, EDWIN, "The A B C of the Federal Reserve System."

PHILLIPS, CHESTER A., "Bank Credit."

RUFENER, L. A., "Principles of Economics," Chap. XXX.

WILLIS, H. PARKER, "The Federal Reserve System."

See references at the end of the preceding chapter.

CHAPTER XIV

PRICE MOVEMENTS, FINANCIAL PANICS AND THE BUSINESS CYCLE

In the three chapters just preceding, there has been set forth the general principle that the value of money varies inversely with its quantity, including as a part of the idea of quantity bank credit, both in the form of deposit credit and bank notes. It is obvious from this principle, that the value of money can be expressed only in terms of a general movement of prices. When the prices of commodities are generally high, the value of money is low and *vice versa*. This is our general principle used in explaining general price movements.

But a little observation reveals that the prices of individual commodities are subject to change also, some moving upward at the same time that others are falling. During an interval of time the economic significance of a good may change, either because of a variation in its supply or in the intensity of the desire for it. In either case, the marginal utility has changed, which will cause a change in its value. These influences affecting the urgency of the desire for a good, whether they appear on the side of supply or on that of demand, will cause changes in the value of individual commodities, but they cannot cause a general movement of commodity prices.

Price changes, therefore, may result from either or both of two distinct influences, *viz.*, a change in the intensity of the desire for a good on the part of its users, or a change in the value of money. *The former influence affects the prices of individual commodities, while the latter causes a change in the general level of prices.* If, then, commodity prices are generally moving in a common direction, the conclusion is that the cause for this movement is to be found in a change in the value of money rather than in any change in the economic significance of the goods themselves. It now becomes necessary to inquire how to determine whether price changes are the result of influences affecting the value of commodities or are due to fluctuations in the value of money.

Index Numbers.—This problem is complicated by the fact that both influences are working simultaneously. The supply and demand for individual commodities vary at the same time that changes occur in the value of money. As a result, some commodity prices are rising and others are falling. The difficulty of the problem, then, is to find a method of separating these influences sufficiently so that a judgment may be formed as to the effect of each on price movements. In dealing with this problem, the quantity theory of money is accepted as the basis for explaining the general movement of prices. It is then assumed that the variations in the prices of individual commodities, due to the fluctuation in the demand for them, will tend to offset each other, *i.e.*, that a rise in the price of some commodities will be balanced by the decline in that of others. If, then, by averaging the prices of a group of commodities at a given period of time, and comparing this average with a similar average of identical commodities at a different date, a change is observable, it is concluded that this change has been caused by a variation in the value of money rather than by a variation in the value of the individual commodities. These averages are known as "index numbers."

An index number, in the most general sense, may be defined as a series of numbers "by which changes in the magnitude of a phenomenon are measured from time to time or from place to place."¹ It will thus appear that an index number is a statistical device by which the variations of a group of varying items are measured. In the problem under consideration, an index number is a device to measure the variations in the average of prices of a group of commodities rather than the variations in the price of individual commodities composing the group. The way in which this is done may be illustrated by taking the prices of a few commodities at a given date and by comparing them with the prices of identical commodities at a subsequent date, as appears in the following table:

¹ SECAIST, H., "Introduction to Statistical Methods," Rev. Ed., p. 469.

Commodities	Jan. 1, 1926		Jan. 1, 1927	
	Base price	100	Price	Percentage to base
1 busbel wheat.....	\$ 1.50	100	\$ 1.85	110
1 ton coal.....	13.00	100	12.75	98
1 ton pig iron.....	21.00	100	18.25	87
1 yard cotton.....	0.23	100	0.25	109
1 pound sugar.....	0.07	100	0.06	86
	5	500	5	490
		100		98

By representing each base price at the first date as 100, then the index number for the group is 500 which, by the second date, becomes 490. By dividing these numbers by the number of commodities in the list the *simple arithmetic average* is found, which is 100 and 98, respectively. An index number computed in this way is known as the *simple arithmetical average*. This is the method most commonly used because it is easy to compute. By comparing the index numbers for the two dates, it appears that the average of prices for the group of commodities examined has fallen 2 per cent.

It would be unsafe, however, to conclude from these figures that the value of money has risen by that amount. In a small group of commodities a large price change in one commodity may cause a variation in the average for the whole group, but if instead of a group of five, fifty, or one hundred, or even a greater number were chosen and the average still showed a general movement, then there would be grounds for greater confidence in the conclusion that the cause for the change is on the side of money. Although, in the table given, two of the commodities have moved upward sharply while the decline in the other three is small, the net result is a small decline in the index number. Usually an examination of such a list will show that the majority of the prices have moved in the direction indicated by the index number. When this is found to be the case, it is reasonably safe to conclude that the cause of the change is on the side of money.

There are a number of different averages used as index numbers, each of which is intended to take account of certain charac-

teristics of the data studied. The simple arithmetical average regards all items in the series of equal importance, and a change of 10 per cent in the price of one commodity is as significant as a like change in another. This average gives undue importance to extreme changes which may occur among unimportant articles in the list studied. For instance, a change of 50 per cent in the price of pepper would by this average be given as much consideration as a similar change in the price of beef. In order to overcome this weakness of the simple arithmetical average, resort has been had to *weighting*, that is the assigning of an importance to the units chosen which is determined in accordance with some definitely selected standard. As a matter of fact, every index number is weighted, since the mere selection of the items composing the series gives a weight to the average. The difference between a "weighted" and an "unweighted" average is that in the former case the weight is determined in accordance with some systematic plan, while in the latter the weights are haphazardly chosen.¹

It will thus appear that there are two ways of weighting index numbers. First, by the mere selection of the items included in the series. If, for instance, more than one item of a given class of commodities is included, that fact gives a weight to the result.² If, in addition to a ton of pig iron there had been added to the series above a ton of steel, this fact would have given to iron and its products an added importance in relation to the other commodities in the series. This method has been called the "implicit" method of weighting. The second method has been called the "explicit" method, *i.e.*, where the weights have been consciously made "by some outward evidences of importance."³ The weights usually given in the calculation of index numbers of retail prices are based upon the quantity of the goods consumed. For purposes of illustration, let us now assign weights to the articles used in the previous table on the assumption that the weights chosen accord with the yearly consumption of these articles. The following result will be obtained:

¹ SECRIST, H., *op. cit.*, p. 505.

² In the Aldrich index number twenty-five varieties of pocket knives were included, thus "giving this trifling article an influence upon the result more than eight times greater than given to wheat, corn, and coal, put together." *Ibid.*, p. 506.

³ *Ibid.*, p. 506.

Commodities	Jan. 1, 1926			Jan. 1, 1927		
	Weights	Base price	Weighted base	Price	Per cent of change	Weighted change price
1 bushel wheat....	5	\$ 1.50	500	\$ 1.65	110	550
1 ton coal.....	3	13.00	300	12.75	98	294
1 ton pig iron....	1	21.00	100	18.25	87	87
1 yard cotton....	2	0.23	200	0.25	109	218
1 pound sugar....	4	0.07	400	0.06	86	344
	15		1,500		...	1,493
			100		...	99.5

The index number in this case shows a fall in the price level, but the percentage of change is not as great as that shown by the simple arithmetical average. If the weights are carefully determined, the weighted average will tend to overcome the influence of marked changes in the price of unimportant commodities.

Other Averages.—The *geometric mean* is an average that is intended to remove the influence of individual commodities. It is computed by multiplying the value of the items in the series and extracting the root corresponding to the number of items. Use of logarithms makes the calculation of the geometric mean comparatively easy.¹ The *median* is another average that will eliminate the effect of extremes. The median is that number that stands half way between the smallest and the largest item in the series of items chosen. The median may be calculated by the following formulæ: If the number of items in the series is odd, then the formula $\frac{N+1}{2}$ equals the median, but if the number is even, the median lies between $\frac{N}{2}$ and $\frac{N}{2} + 1$.² In each case N refers to the number of items or measurements in the series.

Another common average is the *mode*. Sometimes it is desired to find the representative or typical item in a series, i.e., the one which appears most frequently. For instance, if the heights of the members of a class were arranged in a series from

¹ SECHRIST, H., *op. cit.*, pp. 308-309 gives an easy illustration of this method of calculating a geometric mean.

² *Ibid.*, p. 283.

the shortest to the tallest, it might be found that some height, say 5 feet 6 inches, appeared most frequently. This average is known as the mode. Not every series of data will show a mode, even though it may be a very useful method of indicating the representativeness of the series in case there is a piling up of similar characteristics. In the statement that the "average page contains 300 words," or "the average length of a class recitation is 50 minutes,"¹ the mode expresses an accurate summary of the data. There are a number of other averages used in statistical calculations, but the important point to observe in connection with the general topic under discussion is that by means of index numbers it is possible to show general trends in price movements. While an index number does not correspond to any actual fact, it is useful as a means of expressing the general drift of price movements.²

From the previous discussion, we have seen that general price movements are to be explained on the assumption that the value of money has changed, and that the device of an index number is used not only to show the direction of this change but to give some conception of its quantitative magnitude. This method of testing general price movements has become so well recognized that a number of agencies regularly collect data and publish index numbers which are used as the basis for interpreting price movements. One of the most widely used indices is that compiled by the United States Bureau of Labor Statistics. The following table shows the index number of general prices, and of wholesale food prices since 1890. Retail food prices were added to the list of the Bureau's index numbers in 1907. An examination of this table will show that the index numbers of general prices rose rapidly after 1913, or from 100 in 1913 to 226 in 1920. According to the theory of prices that has been previously presented, this movement of the index numbers indicates a fall in the value of money due to an increase in its quantity either absolutely or relatively.

¹ SECRIST, H., *op. cit.*, p. 295.

² TAUSSIG, F. W., "Principles of Economics," 3d Ed., Vol. I, p. 290.

INDEX NUMBER OF PRICES IN THE UNITED STATES, 1890-1926
(Compiled by the U. S. Bureau of Labor Statistics¹)

	General prices	Food prices	
		Wholesale	Retail
1890.....	81	86	
1891.....	80	85	
1892.....	75	79	
1893.....	77	85	
1894.....	69	75	
1895.....	70	74	
1896.....	67	69	
1897.....	67	71	
1898.....	70	74	
1899.....	75	74	
1900.....	81	79	
1901.....	79	79	
1902.....	84	83	
1903.....	86	81	
1904.....	86	84	
1905.....	86	86	
1906.....	89	83	
1907.....	94	89	82
1908.....	90	91	84
1909.....	97	97	89
1910.....	101	101	93
1911.....	93	97	92
1912.....	99	104	98
1913.....	100	100	100
1914.....	98	102	102
1915.....	101	105	101
1916.....	127	121	114
1917.....	177	167	146
1918.....	194	188	168
1919.....	206	207	186
1920.....	228	220	203
1921.....	147	144	153
1922.....	149	138	142
1923.....	154	144	146
1924.....	150	144	146
1925.....	159	157	157
1926.....	151	153	161

¹ General prices and wholesale food prices from Bull. 367, pp. 8-9, U. S. Bureau of Labor Statistics. It should be noted that the number of commodities entering the index of general prices varied from 199 in 1890 to 404 in 1923, while the number of commodities of food for which wholesale prices were collected varied from 36 in 1890 to 65 in 1923.

In order to throw some light on the actual quantity of money that was in use during this interval, the following table has been compiled:

VOLUME OF CREDIT MONEY IN 1913 AND 1920

	1913	1920	Increase + decrease -
Government notes..	\$ 343,000,000	\$ 339,000,000	\$- 4,000,000
Bank notes.....	726,000,000	4,017,000,000	+ 3,291,000,000
Deposits in commercial banks....	12,750,000,000	31,148,000,000	+18,398,000,000
Total.....	\$13,819,000,000	\$35,504,000,000	+\$21,685,000,000

Between 1913 and 1920, the quantity of credit money was increased by \$21,685,000,000, as shown by the figures in the table above. These figures do not include additions to the quantity of gold and other forms of money. According to the report of the Director of the United States Mint, the net addition to the quantity of gold and silver during the seven years may be put at \$2,661,000,000, which added to the increase of credit money makes a total increase of \$24,346,000,000, or an increase in the total volume of money of 2.4 times.¹ Unless the volume of transactions increased at a similar rate during this period, we would conclude that the value of money would fall, and, since the index numbers show that prices generally were higher in 1920 than in 1913, the lines of evidence corroborate each other so that we may conclude that the increase of money, particularly in the form of bank credit, was a potent cause in the fall in its value, or, put the other way, the cause of the rapid rise in prices.

In the preceding paragraphs, what has been said has concerned the movement of prices between different points of time without any attempt to indicate what the index numbers of prices show when they are plotted through an interval of time. When comparisons are made between different dates, we have seen that the general level of prices may have moved upward or downward as a result of the fluctuation in the value of money. If the index

¹ The gold and silver in 1913 was \$2,613,000,000 and in 1920, \$5,273,000,000.

numbers are plotted through a long interval of time, however, a distinct periodicity of price movements can be distinguished. The index numbers of the prices of some types of economic goods show to a greater extent than others this tendency to a flow and ebb of price movements. The prices of stocks and bonds are particularly sensitive and hence show these periodic fluctuations in a marked degree. Closely associated with price movements of this character are the fluctuations in the general industrial and economic conditions.

The Business Cycle.—In recent years, a very large amount of attention has been given to the variation of price movements and of industrial changes. As a result, a new term has come into common usage both in scientific circles and among business men, namely, "the business cycle." Every student of economic history, as well as every business man, is familiar with the fluctuations in business conditions. During some years, business is operating at full capacity, while in others, plants are idle, labor is unemployed, and business generally is depressed. The term, "business cycle," has come to be generally used to characterize the alternating changes from periods of prosperity to those of industrial depression.¹ It has been found that the changing industrial conditions are associated fairly definitely with movements of prices.

An examination of the index of the movement of prices over long periods of time will show two fairly distinct kinds of change which may be illustrated by the following chart:

¹ ELY's "Outlines of Economics," contains an excellent chapter on the Business Cycle to which I am indebted for some of the material presented in this text. Cf. also EDIE, L. D., "Economics, Principles and Problems," pp. 579-611; FAIRCHILD, FURNISS, and BUCK, "Elementary Economics," Vol. I, pp. 502-519.

tary uses in those countries that had abandoned silver as a monetary standard, and by the growth in the volume of business transactions. This was the period of rapid expansion of railroad building in the United States, with its resulting effects upon the area of the markets. During this period, the markets of the country became nationalized and there was a very rapid expansion of large-scale industry. The corporation became generally used during these years as a form of business organization and through the corporation the scale of industry was greatly enlarged. These industrial changes were accompanied with a larger volume of business transactions which, in turn, intensified the demand for money. In addition to these causes, the annual production of gold was slightly below what it had been from 1850-1873. These influences combined to appreciate the value of money which was reflected in gradually falling prices during these years.

Finally, from 1896-1914, prices moved upward, which can be explained mainly by the discovery of new gold coming from the Alaskan and South African mines, and by the application of more economical methods of producing gold. Another factor of great importance that was operating throughout the nineteenth century, and which became increasingly important after 1863, was the growth of the banking system and the use of bank credit in effecting business transactions. If the use of bank credit had not become more general, either the price level would have been lower, or the volume of transactions would have been smaller than they were. These long swings of prices, which have been indicated, are called "secular trends" and they are closely associated with the changes in the supply of standard money in relation to the volume of trade.

In addition to the long swings of prices, it will be observed that there are changes of shorter duration. A more detailed chart would show these shorter changes more clearly. The term "business cycle" is applied to the movement of prices during the shorter intervals of time. Many industries show rather regular swings upward and downward in the volume of business that is transacted, corresponding to the fluctuations in general prices. While not all industries show this periodicity, and, while some authors have even questioned the regularity of the change,¹

¹ FISHER, I., "Our Unstable Dollar and the So-called Business Cycle," *J. of Am. Stat. Assoc.*, Vol. XX, pp. 179-202.

it will be found that the economic literature of recent years has accepted the regularity of movement as a fact and has given more attention to this phase of our economic life than to any other subject. Both the secular and the cyclical movements of prices and of transactions have been examined with great care and skill, the purpose being to find what constants or uniformities exist so that some conscious methods of control over production might be developed.

Before undertaking a detailed analysis of the business cycle there is one further movement of prices that should be examined. Sharp fluctuations in prices are clearly observable in the above chart at several well-defined periods, such as 1812-1814, 1834-1837, 1861-1870, and 1914-1920. These are all periods of monetary inflation, due to credit expansion in one form or another. Wars have always been accompanied by credit expansion and rapidly rising prices, and were responsible for three of the four periods mentioned. The War of 1812 caused the monetary inflation during the first period; "wild cat" banking and the excessive issue of bank notes were responsible for the second; the greenbacks in the 'sixties, and overexpansion of bank credit after 1914, account for the sharp rise of prices at these latter dates. The very rapid rise in prices after 1914 caused grave consequences in the United States, but the situation was much more serious in Germany, Austria, and Russia during these same years. The enormous expansion of paper currency that took place in those countries resulted in extremely rapid changes of their commodity prices. Much of the distress both during and after the war was caused by the depreciation in the value of their paper money. This arbitrary expansion of currency during a short period of time is what is known as "monetary inflation." If money is expanded faster than the volume of money work to be done, it may be said to be inflated. Inflated currency is always accompanied by sharply rising prices, and the periods here mentioned show this effect in a marked degree.

Four Phases of the Cycle.—Returning to the consideration of the cycle, there are four fairly well-recognized phases, or stages, in the cyclical movement of business, *viz.*, the periods of "recovery," "prosperity," "the crisis," and "the depression." In general, in those industries that show marked cyclical movements, there is an interval of from three to four years between the peak of one cycle and that of another, although the interval of time is

by no means fixed. Some cycles run for eight or nine years or even longer, but statistical studies have shown a regularity in their recurrence that may be expressed roughly by the interval of years given. Moreover, while the different phases of the cycle are definitely perceptible and have characteristics that are distinguishable, there are no sharp lines differentiating one phase from another. The feature which marks the different stages is the existence or prospect of business profit. During the period of recovery, profits are low but the prospect is for business to yield higher returns. Prosperity is associated with great business activity and high profits. The crisis is reached when the prospect of profits collapses, and credit is curtailed. Finally, depression is the period of low profits and great curtailment in business.

1. *Recovery*.—A brief description of some of the more important relations may make the cyclical movement more easily understood. We will start our study with the period of recovery. The conditions within a depression set in motion the forces that make for business recovery. Production has been sharply slowed down, thus giving a chance to use up surplus stocks of goods. Costs of labor and of raw materials have likely fallen and credit has become easy, making business loans less expensive. With these conditions developing, the first indication of a revival is an increase in security prices and in speculative activity. Idle funds withdrawn from productive enterprises are used in purchasing high-grade bonds and stocks. This renewed activity in the security markets results in a rise in the security prices. The upward swing of security prices gives some stimulus to business which, together with the low cost of production, the exhaustion of surplus stocks of commodities, and a sharpening of demand, gives courage to producers. The prospect of profits adds confidence and production begins to expand. The period of recovery has gotten under way.

2. *Prosperity*.—The next phase is just an extension of the first. Prices of commodities continue to rise, security prices continue to advance, and demand increases. This renewed activity is passed on from the middlemen to manufacturers, and by them to the producers of equipment and of raw materials. Larger numbers of workmen are employed, and a larger volume of purchasing power in the form of income is in the hands of the consumer. This, in turn, is reflected through the producers of

consumers' goods, and the circle of increased business activity is widened. Profits are now high and business confidence is strong.

3. *The Crisis.*—But the conditions of prosperity tend to breed those which bring about the crisis. In order to reap as large a share of profits as possible, every business man is tempted to expand his production. His costs are likely to be increasing because his plant and equipment are running beyond the point of maximum efficiency, and wages are likely to rise, as laborers are quick to take advantage of such conditions to demand more pay. Bank discount rates will also tend to rise because the expanding demand for credit throughout the entire business structure brings pressure upon the reserves. The expansion of production causes a competition between the manufacturers and the speculators for a control of funds. The banks resort to higher rates as a means of discouraging further borrowing. These costs will increase faster than business income, and selling prices will reach a maximum, partly because of custom and habit, and partly because some industries, such as public utilities, are unable to change the rates charged for their products or services on account of government regulation. Rising costs and the check on the increase of selling prices will cause a fall of business profits. Marked inequalities as between industries will exist and will unsettle the balance between production and consumption.

As soon as these difficulties come to be recognized, the bankers will bring pressure upon creditors to meet their obligations, and the weaker ones may be forced into bankruptcy. The crisis has now been reached. The business man with large inventories, increased costs, falling profits, and pressure to meet his obligations to the bank, begins to cut prices and to reduce expenses. Since the wage bill is almost always a large item, wages will be cut or men thrown out of employment, with the result that a smaller amount of purchasing power is in the hands of the consumer, which in turn will cause a contraction of demand. Security prices have already begun to fall, bank discount rates have been increased, the federal reserve banks have doubtless raised their rediscount rate which has an important psychological influence on the minds of business men generally, and bank reserves are at a low point. These are the downward moving forces that lead into the fourth phase of the period of depression.

a. *Financial Panics.*—However, before examining the characteristics of the fourth phase of the cycle, two topics should

receive attention at this point, namely, *financial panics and the relation of banks to panics*. The crisis is frequently known as a financial panic and the strain which occurs in the money markets of the country at such times can best be understood by a description of the ordinary course of events in the conduct of business. All business firms are conducted on the basis of giving and taking credit and they have bills to pay for goods purchased, and accounts receivable for goods sold. These bills are ordinarily promptly paid and usually by means of checks against bank deposits. So long as the movement of transactions is normal and the obligations can be met from receipts, no difficulties occur; but should anything happen that unsettles this normal movement, such as a poor crop, or a mistake in estimating demand for some essential commodities, or overexpansion of the means of production in particular directions, so that some firms are unable to meet their obligations when due, then credit facilities may become strained.

It may be that the firm affected is small and only the owners suffer, but if a firm of large size becomes embarrassed, particularly if it should occur at a time when the banks have extended their credit to the margin of safety, the effect may be felt throughout the whole country and the whole credit structure may be shaken. During ordinary times, the business firm simply desires to discount its notes at a bank and secure deposit credit from the proceeds. The banks are usually glad to carry such notes as they constitute an important source of their earnings. But as business expands the number of discounted notes will increase and bank credit becomes greatly expanded. Should anything happen at such times, a financial strain will result and immediately there will be pressure brought to bear on all debtors to meet their obligations in order that the banks may protect their reserves. Those unable to meet their payments promptly try desperately to secure additional extensions of credit. At such a stage there will occur a financial panic, or an acute state in the use of money and credit in the operation of the business of the country.

b. *The Relation of Banks to Financial Panics*.—From what has been said, it is apparent, that the banks play a very important part in every financial panic. So long as the banks can extend their credit, the merchants and business men can continue to operate, but when the limit to the extension of bank credit has been reached, then care must be used to prevent the

strain on bank credit from becoming so acute as to cause a panic with its attendant evils. Prior to the adoption of the federal reserve system, the banks were recurrently under pressure. The lack of elasticity in the extension of bank credit and the impossibility of effective cooperation among the banks under the National Bank Act made financial strains frequent and serious. With the centralizing of the reserves into the reserve banks and the provision for rediscount, the banks under the federal reserve system are in a much stronger position to render effective aid to business firms than before the adoption of this act. With the exercise of sound judgment by member banks in the extension of loans, banks may now come to the relief of the business community and avert the ordinary financial strains. But the system is not free from the possibility of a financial panic if such care is not exercised. In the fall and winter of 1920, bank loans had been expanded to about the full legal limit that could be carried by the combined resources of the reserve banks. There was considerable alarm expressed lest the whole system would break down, and, although the situation was acute, the banks succeeded in coming through it but not without so much pressure on business firms as to cause a decided fall in general prices.

The acute stage of a panic does not last long. The demand for money in the form of deposit credit soon passes and there follows a period during which reserves accumulate in the vaults of the banks. Within a short time the discount rate falls and the banks are willing to extend loans, but the lack of confidence prevents a revival of business activity for a longer or shorter period. During this time depression ensues, and not until available stocks of goods are used up and the caution created in the minds of the business men gives way to a spirit of confidence does business revive.

4. *The Industrial Depression.*—The fourth phase of the cycle is that of depression. The crisis may be of short duration, affecting mainly banks and mercantile classes, or, if it is of sufficient gravity, it may shake business confidence so deeply as to develop into a widespread industrial depression. An industrial depression means a general stoppage of industry, or the operation of industry on part time. It means unemployment for many workers and idle plant and equipment for stockholders. For business generally to be prosperous, there should be a

delicate balance between production and consumption and continuous operation of plant and equipment. It is the stoppage or slowing down of production that constitutes what is known as a "depression." During such periods profits are low, prices of securities are at a minimum, unemployment is widespread, commodity prices fall, and business and trade are generally at a low ebb.

But, within the period of depression, the forces that make for recovery begin to operate. The most energetic business men strive to enforce economies in production and wastes of all kinds tend to be eliminated. New methods are likely to be tried as a means of reducing expenses so that operation may continue on the lower price level. Bank reserves accumulate and bank credit becomes easier. Operating costs are usually low as the prices of raw materials have declined; labor output per man has increased and other expenses have fallen off as compared with those that previously prevailed. These are the conditions that are favorable for recovery. After the depression has run its course, confidence begins to revive, security prices begin to rise, and a new cycle is on its way to repeat the general aspects that have been briefly portrayed above. It may be that the period of recovery will take place slowly because the conditions generally are not favorable for the development of prosperity. A period of comparative equilibrium may last several years, as in the case of the recovery of 1908.

The cyclical movement shows itself not alone in price changes but in fluctuations in production, in trade, in employment, in the consequent variations in the earnings of capital and labor, and in changing rates for the use of bank credit. In fact, it seems that all economic phenomena show the effect of these recurring fluctuations in a greater or less degree.

In the description of the cycle that has been given, a degree of definiteness has been implied that is unreal. For purposes of exposition, the various stages have been set forth boldly but with the full knowledge that they cannot be marked off with the degree of definiteness that has been suggested. In fact, it frequently happens that the different phases of the cyclical movements can be determined only after they have run their full course. A sequence of events quite in accord with the general description here given has been made out by close students of this subject, and the general characteristics listed

above have come to be so well recognized that many business concerns use index numbers of business conditions as an aid in developing their business policies. Through the application of scientific analysis of the various factors that affect business, an element of conscious control is introduced into our complex exchange economy.

Causes of the Business Cycle.—The recurrence of periods of industrial activity and depression has long been recognized and many attempted explanations have been made. Among the theories that have been advanced, the following may be mentioned, although no attempt will be made here to give an exhaustive treatment of the various causes which have been assigned to these movements. The first attempts to explain crises was by means of psychological influences. The crises were held to be the result of a speculative mania. The tulip mania of 1636 in Holland, the South Sea and Mississippi bubbles of 1720 in England and France have been cited as examples of such speculative manias.¹ Psychological factors, no doubt, do have an important bearing on business activity. Much is written and said today concerning "lack of confidence" and the "restoration of confidence" as forces affecting business prosperity. If business men are overoptimistic they may expand production, or the means of production, faster than the existing rate of consumption. Soon, supplies of particular commodities at least, are left unsold. The manufacturers and merchants are hard pressed and unable to pay their bills and, if this condition is general, the forces of regression will soon bring on an industrial depression. In like manner, if business men are too pessimistic and overcautious, their attitude may prevent the readjustment after the depression. Psychological factors do unquestionably have an influence upon business activity, but transactions which are at the basis of all business activity have a more realistic existence than the psychological attitude of the men engaged in business.

Another theory that has frequently been advanced is that of overproduction. It has been argued that the capacity to produce has exceeded the capacity to consume and, hence, there are developed periods of stagnation in business. The earlier economists answered this argument by denying the possibility of a general overproduction. They pointed out that the production of goods fundamentally increased the demand for goods, that

¹ *Elx, R. T., "Outlines" 4th Ed., p. 324.*

people generally produce goods to trade with one another, so that it was impossible to explain a stoppage of industry on the ground of general overproduction. They admitted that there might be overproduction of a particular kind of goods, but not of all goods. As an abstract principle this argument is sound, but the admission of overproduction of particular commodities may, if on a sufficiently large scale, have such a widespread effect as to be an important contributing cause of a financial panic and the recurring industrial depression.

When a country, or a large section of the population of a country, is dependent upon a single crop or industry, it is quite conceivable that the supply produced may be so large as to depress prices below the cost of producing the goods. In that case, the buying or trading power of the people in such regions may be so greatly reduced as to cause a falling off in demand and a maladjustment of production and consumption on such a scale as to result in a general depression. Something akin to this situation occurred in connection with the decline of agricultural prices in this country in 1920. The buying power of the farmers was suddenly reduced. Many of them had obligations which they hoped to meet from the proceeds of the crop of that year. The rapid decline of prices of farm products not only caused a serious credit situation in the agricultural sections of the country, but it also reduced the buying power of the farmers to such an extent that all industries that catered to them were likewise hard hit, such as the manufacturers of agricultural implements and the great mail-order houses.

It may be objected that this decline in price in 1920 was not due to an overproduction of farm products. Absolutely, agricultural crops were not so large as to explain the decline, but relatively the supply was large. World conditions had so changed from the summer and fall of 1919, when the crops were planted, that the amounts harvested in 1920 far exceeded the amounts that could be sold at 1919 prices. The dropping off of European demand and the general credit conditions were significant causes, but relatively at least there was an oversupply of farm crops. The condition of the farmers was a very large factor in the depression which followed. Whenever there occurs a maladjustment of production and consumption of considerable size, whatever may be its cause, the effects will certainly be felt throughout the whole industrial structure. Overproduction of

particular goods may often contribute to an industrial depression. The fault with overproduction as an explanation is not that it is untrue, but, as it has been advanced, it is superficial and too simple to account for the phenomena it attempts to explain.

The socialists gave this argument a special interpretation. They held that the capitalistic wage system, especially with machine production, increased products faster than it increased wages or the consuming power of the laboring classes. If the wage earners were unable to consume at a rate as fast as they produced, the result would be overproduction. Overproduction, or as it was sometimes called, "under consumption," was the primary cause of industrial crises and depressions. There is an element of truth in this argument. There should be a perfect balance between the rate of production and the rate of consumption, and if this balance is maintained, industry will move forward in a perfectly uniform course. The weakness of the argument lay in the stress that was put upon the conflict between capitalistic methods of production and general economic welfare, and the insistence that the inevitable outcome of this system of industry was overproduction, which is checked from time to time by crises. The fact is, that business men are constantly estimating how large a volume of goods the market will absorb and are using these estimates as guides for their production policies. If too much is produced, it is because their estimates are wrong and not because of any imperious force of nature that increases the productive capacity of industry faster than the rate of consumption. Before cyclical movements can be explained, there must be an investigation into the reasons why the estimates of business men are incorrect.

Some writers have tried to explain the business cycle by variation in conditions of the weather, which they have attempted to show has followed some periodic forces of nature. The first of these writers was W. S. Jevons, the English economist who advanced the *sun-spot theory*. Proceeding with the theory that the sun spots had a meteorological effect and that these disturbances occurred about every ten or eleven years, an interval that corresponded closely with the recurrences of crises in England, he argued that the sun spots must affect the rainfall which in turn would affect agricultural crops and cause a maladjustment in the buying power of different groups of people. This explanation has not been proven, although it still receives the

attention of some scholars.¹ There is no agreement among investigators as to the length of the cycle, nor do meteorologists agree as to the existence of periodic cycles in weather conditions. Hence, the attempt to explain the cyclical movements by means of natural forces, such as regular climatic changes, has not been successful to date, yet we cannot ignore the effect of the weather on crops and through them on general business conditions. As has been shown, a large crop may depress the price and actually reduce the buying power of farmers, just as a crop failure may leave them nothing to sell even though the prices of farm products are high. In either case, not only the farmers, but all others who produce to sell to the farmers, are affected by such crop conditions. Generally speaking, large crops enhance industrial activity, while poor crops retard it. If a poor crop should occur at a time when credit conditions are strained, it will almost certainly occasion a crisis, just as a large crop at the end of a period of depression may start business activity on the road to recovery.

Modern explanations of industrial crises are both economic and psychological. The division of labor in modern industry, which makes it necessary to produce in anticipation of demand on an estimate of demand is an important economic cause. We have already shown how the responsibility for making these estimates rests with individual entrepreneurs, each striving, on the one hand, to secure as large a volume of business as possible and to sell at as high a price as he can, while, on the other, he is trying to produce at the lowest unit cost. The length of time that intervenes between the beginning of production and the marketing of the finished good is responsible for a large element of risk and directs attention to the extremely difficult task of estimating correctly the quantity of goods that purchasers will buy. Any mistake in these estimates will result in a maladjustment between supply and demand.

This maladjustment in production is one of the important causes of industrial crises. This cause is particularly important in industries which employ a large fixed plant, such as the railroads, or large manufacturing concerns. If the savings of the country are turned into the construction of plant and equipment at a rate faster than the demand for the products or services therefrom, it will cause a maladjustment of production and consump-

¹ JEYONS, H. S., India, BEVERIDGE, W. H., England, MOORE, H. L., Columbia University.

tion. The rapid construction of railroads after the Civil War into sections of the country that had not yet developed the need for transportation services was an element in the industrial difficulties both in 1873 and 1884. The return from the operation of the roads was not sufficient, so that the owners became involved and were unable to meet their obligations to the banks. When such miscalculations are on a large scale they can affect the whole industrial structure.

The psychological factor enters at this point. Business judgment as to how fast to expand plant and equipment is the important element needed. When business is active, most business men are optimistic. They are inclined to overbuild or over-produce. Since they are engaged in buying and selling to one another, the spirit of optimism is likely to spread. As long as orders are coming in at prices that will yield a profit, the average business man is satisfied. In fact, he is likely to be so absorbed with the internal problems of administration and the continuity of his business that he is unaware of the dangers that may lie ahead of him. Unless he is a man of unusual insight and foresight, or has made a careful study of the past experience of his plant in its relations to the general business conditions, he has a very inadequate basis on which to estimate future production or to plan his policies beyond the day-to-day operations. But so long as he can dispose of his products at favorable prices, why should he worry? It is a frame of mind of this kind that is responsible for the mistaken estimates of production, and when the depression arrives, the spirit of pessimism is as contagious as optimism had previously been. As there are now few orders and heavy fixed expenses, there will be a general lack of confidence that affects the whole business community. These psychological factors are present in every phase of the business cycle and are an important influence in explaining the phenomena in connection with it.

Another factor is bank credit and bank operations. Banks collect the savings of individuals of the community and make these sums available for the use of the business men. In addition, they lend their own credit and thus furnish an elastic supply of purchasing power which expands as the volume of business is increased. But banks cannot continue indefinitely to expand credit and, even if they could, there would be other factors that would react on the extension of business. The mere extension of

credit does not create commodities and services. If bank credit increases faster than transactions, the general price level will rise and eventually, because the prices of different commodities do not rise uniformly, the fundamental forces of production and consumption become maladjusted. The recent experience in European countries with irredeemable paper money has shown again that the credit element in a nation's currency can destroy itself by getting out of harmony with production and consumption.

During the course of a cycle, bank credit behaves somewhat as follows: During the stage of depression, debts to the bank are being paid, loans and discounts decline, deposit liability decreases, and reserves accumulate. Commodity prices in this stage are likely to be lower and the number of transactions much smaller. The volume of money in circulation becomes smaller and the idle money flows into the banks. The net result is a high ratio of reserves to deposit liability. Next, the banks begin to feel the effect of idle reserves and are inclined to lower discount rates, which makes borrowing more attractive. Bonds sell at favorable prices so that all in all it is a good time to build new plants.¹ In addition, low prices foster exports and retard imports. Gold is likely to flow into the country in payment of international balances.

All of these circumstances favor an increase of business activity. The increase of borrowing by business men will show itself in more money paid to laboring men and to those furnishing materials of production. Sales begin to increase and prices begin to rise. Rising prices call for larger bank balances in order to carry on the volume of trade. Extension of loans is followed by an increase in deposits which diminishes reserve ratios. Higher prices require more money in circulation, a condition that helps to draw down reserves. The high prices also tend to discourage exports and foster imports, which, in turn, tend to stimulate the demand for gold in settlement of international balances. In these ways the reserves are drawn down and a limit is reached beyond which bank credit cannot expand. Thus, a positive limit is placed upon business expansion by the extent to which bank credit may be expanded. The crisis stage is reached at the peak of the upward movement of prices.

¹ Some firms regularly take advantage of such periods to make expansion of plant and equipment.

So far as bank credit contributes to higher prices it is an important factor in the business cycle. The banker no more than other business men can estimate the course of production and consumption accurately and, besides, he, like other business men, is endeavoring to secure as large earnings from the lending of bank credit as is consistent with safety. He cannot be expected to look with favor upon idle reserves if those who seek loans can satisfy him of their ability to pay. He will continue to extend loans as long as the prospect of repayment is good, irrespective of the effect of these loans on production and consumption.

While bankers, through their control over credit, do have a tremendous influence in determining the direction of expansion of production, they are interested in the financial success of the business venture and the ability of the borrowers to pay, and, generally speaking, are not concerned, except indirectly, with the general problem of maintaining a balance between production and consumption in the country as a whole. Put in other words, the making of a loan may accord with sound banking principles, but it may at the same time contribute to the maladjustment of production and consumption. It should be said, however, that many enlightened bankers have of late years devoted increasing attention to the problem of general business fluctuations, and a number of banks now publish monthly surveys and reviews upon these general conditions for the benefit of their customers.

However, the estimating of demand is the special function of entrepreneurs, while that of the banker is concerned with the choice of risks. The problem of the cyclical movement of prices and of business activity then comes back to the fundamental problem of estimating demand in anticipation of demand. Crucial mistakes are made at this point and these mistakes run their course through business profits, expansion of plant, extension of bank loans, general business activity, higher prices, higher costs, the crisis, lower prices, lower profits, unemployment, and business depression. The fundamental problem has to do with maintaining a balance between production and consumption.

As has already been stated, the function of estimating demand is the special job of the entrepreneur. When we contemplate the complexity of this problem, and consider the large number of individuals who are performing this function and the factual basis upon which their estimates are made, it is not surprising

that mistakes are frequent. The wonder is that mistakes are not more frequent and more serious than they really are. To state the case in this way is placing no blame on those who are serving in the capacity of business executives and bearing the responsibility of determining the production policy of the firms with which they are connected. The facts are that these men have little to guide them beyond the past experience of their firms together with the judgment with which they themselves have been endowed. They rely very largely upon market price as the indicator of the urgency of the needs of consumers and conclude, when prices begin to rise, that supply in relation to demand is running behind or, in other words, that production is falling below consumption.

From what has been presented concerning the relation of credit and money to prices, it should now be apparent how deceptive price is as an index of demand. If prices are rising as the result of a shortage of supply, the entrepreneurs should hasten or even expand the processes of production in order to meet that shortage. But how is the average business man to know whether the rise in price is due to a relative shortage of goods, or to the falling value of money? His orders may be maintained on the same level or may even be increasing, but how does he know that his customers are not affected by the same conditions as affect him, and if a considerable period of time is involved in producing his goods, what assurance has he that his customers may not cancel their orders because of a false interpretation of the state of the market?

These queries suggest some of the difficulties under which the business executive labors. It has only been within comparatively recent years that economists have undertaken to analyze the causes of industrial crises and to develop principles and agencies by means of which business executives can shape their production policies more intelligently. The study of the business cycle has probably received more attention in recent years at the hands of the economists than any other single question, and it has resulted not only in a better understanding of the problem, but in the formulation of methods of dealing with the phenomena connected with the cyclical movements of industry that make the forecasting of prices and of general economic conditions possible with a degree of accuracy heretofore unknown. The data concerning production, prices, bank clearings, and other

economic phenomena have been gathered and analyzed, and there has been found to be a fairly well-defined sequence between stock exchange speculation, discount rates, and the volume of business activity. On the basis of data of this character the probable course of prices and business activity can often be forecast with a reasonable degree of accuracy. Trade associations, individual firms, and governmental agencies are giving increasing attention to this problem. Many individual firms are studying the problems of their own particular industry in connection with the general conditions, and using the results of these studies to guide the policy of the firm. So far as scientific method makes possible the forecasting of business fluctuations, these fluctuations will tend to disappear because business men will discount more accurately the probable future movements of industry.

The Social and Economic Consequences.—The fluctuations in the value of money and industrial conditions that occur in connection with the business cycle may be regarded as generally injurious. They cause hardships that fall unequally upon different persons or classes in the country. On the upward swing of prices laborers suffer relatively, for commodity prices always rise more rapidly than wages. Offsetting factors are greater regularity of employment and overtime pay. On the downward movement, wages do not fall as fast as prices, so that laborers are relatively better off, except as irregularity of employment may reduce their earnings. As the depression develops and unemployment becomes general, thousands of workingmen *suffer serious curtailment of income and a consequent encroachment upon their standard of living.* All salaried persons and others depending upon fixed income suffer by rising prices and *gain by falling prices.* The debtor class and those engaged in the production of commodities for sale benefit by rising prices. Debtors find that dollars are easier to get and, since their obligations are expressed in dollars, it is therefore easier for them to settle their debts. They are, however, tempted to incur additional obligations on the assumption that prices will remain high and money will continue to be easy to get. Many farmers in 1919 and 1920 were deceived by the existing conditions and were led to incur obligations which they were unable to meet after the prices of farm products fell in 1920. If the debtor classes had sufficient self-control and vision, they might take advantage of cheap money to pay off current obligations and thus benefit

by rising prices. But experience has shown that they do not have sufficient self-restraint, and hence they suffer along with the other classes already mentioned.

Business men quite generally look upon rising prices as a symptom of prosperity. While it is usually easier to operate on a rising than on a falling market, because there is greater assurance that goods contracted for in advance can be disposed of at a price that will show a favorable balance, nevertheless, if the cause of rising prices is due to a depreciating money, it is easy for the business man to misinterpret the results that he has obtained. His profits may be larger in dollars, but the dollars may not be as effective as before, so that the net result may be a loss rather than a gain, when his accounts are figured in real income. While there is a general disposition to emulate the periods of prosperity when everybody is busy and the "wheels of industry are humming," yet if that condition can be secured only at the cost of the subsequent period when thousands of men are unemployed and plants are idle or running on short time, it would be economically and socially better if industry could be organized and operated on a more regular basis. We may conclude, then, that the general level of economic welfare is not promoted by rapid fluctuations in prices, because of the effect such price changes have upon the adjustment of production to consumption. It would seem, therefore, that the attempt at stabilization of industry, or at least the effort to reduce the range of business fluctuations, is a goal worthy of the best efforts of both scholars and business men.

Remedies Proposed.—Because of the consequences of industrial depressions, such as those described above, there have been a great many proposals advanced for eliminating the evils of the cyclical movement of industry. First, we will consider the control of credit. Bank credit, as has been shown, is an important factor in the price changes during the different phases of the business cycle. If bank credit inflation can be eliminated, it will help to remove one of the contributing causes of price changes. The control of bank policies involves what the individual banks can do and what the reserve banks can do.¹ The individual banks could do a great deal to prevent excessive booms, if greater care were given to the quality of the loans made at different phases of the cycle. Bankers could not only raise

¹ EDM, L. D., "Economics, Principles and Problems," p. 606.

rates of discount but, more important, they could insist on a higher ratio between current assets and current liabilities on the part of their customers. If the normal ratio be two to one, and the banks insisted on a three to one ratio, or higher, as bank credit became expanded, it would discourage the business man from making unwise extensions of his business. Such a policy, if adopted generally, would have a powerful effect in checking overexpansion of credit. The difficulty of initiating a policy of this kind is very great because individual bankers are in competition with each other for business, and the fear that some other banker may get the business is a powerful motive in deciding upon a loan. Hope for relief in this direction lies in the more general realization by bankers that their long-run interests are parallel with the interests of society, and that they ought to join hands with each other to use their control over credit as a means of effecting greater regularity of production.

The most potent banking force in this country now rests with the Federal Reserve Board and the federal reserve banks. These institutions have two methods that may be followed, both of which will affect general credit conditions. First, they may change the official rate for rediscounting the paper of member banks. Whenever such a change in the rediscount rate is made, it should be an indication to the bankers of the country that, in the opinion of the officers of the reserve banks, the credit conditions of the districts affected need the attention of all the banks in that district. The psychological effect of this action may be sufficient to check the expansion of bank credit. Unless the member banks hold enough gold in their own tills to make them independent of the reserve banks, an increase in the rediscount rate will be reflected in commercial rates. But at certain times many of the strongest banks may be independent of the reserve banks, because their own gold holdings are large enough to take care of the immediate demand for credit. Through their own resources they may, therefore, so expand credit as to cause prices to rise, and consequently at such times the raising of the rediscount rate may have little effect in checking credit expansion.

At such times, it is argued that the federal reserve banks should exercise their second method of control and engage in open-market operations, *i.e.*, buy or sell bankers' acceptances or government paper in competition with member banks. Such

transactions will tend to "tighten" or "loosen" the money market, depending upon the direction that the open-market operations take. If they sell securities when prices are high, it will tend to take money out of the member banks and thus check credit, whereas, if they buy when prices are low, it will put money back into the banks and thus facilitate the recovery. It is held that in either case the open-market transactions would have a compensating effect upon price movements and credit conditions. The difficult problem is the determination of the right time to begin the open-market operations. This question is too technical for the purpose of this text and students who are interested are referred to more advanced courses in banking or to the literature of this field.

The Stabilization of the Dollar.—Closely associated with the control of credit as a means of affecting greater regularity of business is a second proposal, usually referred to as the stabilization of the dollar.¹ It is well known that the dollar as a standard of value is an unstable unit. If we had to buy and sell cloth or grain by a yardstick or a bushel that fluctuated as much as the dollar, we would have a very difficult time. Imagine a yardstick that fluctuated 100 per cent in the course of seven years! Yet this was what occurred in the purchasing power of the dollar between 1913 and 1920.

At present the dollar remains constant in weight, but fluctuates in value. It consists, as has been indicated above, of 23.22 grains of pure gold, or 25.8 grains nine-tenths fine, but the purchasing power of this weight of gold fluctuates very widely. A proposal has been made that instead of a dollar fixed in weight and fluctuating in value, we adopt one that will fluctuate in weight and remain stable in value. The proposal would necessitate the measurement of the fluctuations in the value of money by means of index numbers, and it would involve the definite withdrawal of gold from circulation and the substitution of gold certificates—a condition, be it observed, not so different from what now prevails so far as the general circulation of money is concerned.

The proposal further provides that when the value of the dollar, as shown by the index number has changed, say by 1 per cent, some government agency, like the Comptroller of the

¹ Professor Irving Fisher, of Yale University, for a number of years has been an ardent advocate of a stabilized dollar.

Currency, would then be authorized to announce the rate at which the treasury department would exchange gold for gold certificates. If gold had depreciated more than 1 per cent in value, then a correspondingly larger amount of gold would be given, or taken, for a gold certificate of a given denomination. In this way, it is argued, that the fluctuations in the value of money could be kept within the percentage range above or below par, at which a change in the ratio of gold to gold certificates was made. The limit of this range would fix the limit of price fluctuations. In other words, prices could neither rise nor fall more than by the percentage of variation from par that was permitted before a change was announced in the ratio of gold to gold certificates. This change in the ratio of gold to gold certificates would have the same effect as varying the weight of gold in the dollar to keep pace with the fluctuations in the value of gold. If this were done, it is held that the general level of commodity prices would remain relatively stable, and the only changes in price thereafter would be those that registered variations in the supply and demand for goods.

The success of such a plan rests first, upon the scientific accuracy of the index numbers used to measure price fluctuations. Granting that the scientific basis can be assured, then the success of the plan would depend upon the protection against political or other manipulation by whatever governmental agency that had the responsibility of announcing the exchange rate of gold for gold certificates. Assuming that a scientific method of measuring price changes could be established, and that adequate protection against manipulation could be devised, it would then probably take a period of ten years to educate the people of the country to the point of accepting a dollar that fluctuated in weight instead of the one with which they are now familiar that is constant in weight. The proposal has merits and, if it could be adopted, would tend to correct the fluctuations in prices due to changes in the value of standard money. Since the volume of bank credit does not follow an absolute ratio with the volume of standard money, the stabilized dollar would not correct all price fluctuations due to monetary causes. Unless steps could be taken at the same time to stabilize credit in terms of money, price fluctuations due to the changes in the volume of bank credit would continue as at the present time. Only the longer swings in price movements that are due to changes in the quan-

tity of standard money and the effect that this larger volume of gold has on credit would be affected by the establishment of a stabilized dollar. Not much can be expected from this proposal in the immediate future so that the hope of relief from the evils of the business cycle seems to lie along other lines of endeavor.

Forecasting and Regularization of Industry.—A third proposal of increasing significance has been suggested under the general term of regularization of industry. This proposal has been discussed both as a private and a public responsibility. The elimination of business cycles involves a matching of production and consumption. The responsibility for private production is vested in the business executive who determines the production policies. The development of statistical knowledge and statistical methods has put into the hands of the executive tools for estimating future demand and for the determination of business policy that formerly did not exist. Information of this character, together with more accurate knowledge concerning unit costs obtained from modern cost accounting methods, equips the executive to predict the course of business more accurately than at any previous time in modern industry. It is claimed that, by giving attention to the cyclical movement of prices and more thought to production policies in relation to market possibilities, a considerable part of the fluctuations in business can be eliminated. Some modern concerns have given this subject a large amount of thought with the result that the extreme fluctuations have been smoothed out and greater regularity of production effected.¹

Closely associated with the proposals for regularizing private industry have been those for the use of public works as an agency for the smoothing out of periods of business prosperity and depression. The expenditures of the various governmental units in this country for public improvements of one kind or another, amount to hundreds of millions of dollars annually. It has long been argued that governments should use this spending power as a method of relieving unemployment. It is at least possible that, by means of long-range planning, public expenditures could be used effectively as a means of relieving industrial depressions. Such planning would involve the setting aside of a definite portion

¹ The Dennison Manufacturing Co. at Framingham, Mass., is a noteworthy example.

of annual appropriations for public improvements to be held as a reserve, and to be freed at the time private industry becomes paralyzed. The volume of public expenditures launched during a period of sluggish industrial conditions would be sufficient, it is held, to revive business and to set it on the road to recovery. Any public improvements would involve the purchase of raw materials of many kinds, so that this new demand would be felt throughout the whole industrial structure and would, therefore, breed a spirit of business confidence and optimism.

It is easy to show how significant the creation of a reserve of this kind would be if all governmental agencies would set aside for this purpose, say, 10 per cent of their annual appropriations for public works. There can be little doubt that, if governments *could* be persuaded to set aside such reserves, and *would* spend them at the right time, it would have a tonic effect on industry along the lines of the claims for such a policy. The difficulty here, as in the case of so many other worthy proposals, is one of practical expediency. What political officers would have the self-restraint to set aside a reserve that their opponents might have an opportunity to spend! Until more effective methods of safeguarding public expenditures than now obtain and a higher type of statesmanship than is generally found are present and guiding government officials, we cannot expect much relief from this proposal, even though it may hold great possibilities for the future.

Another proposed remedy for unemployment is that of unemployment compensation. Arguing from the experience of workmen's compensation as an aid in reducing industrial accidents, some writers have suggested that unemployment compensation would enlist the economic motives of the employer to reduce unemployment in the same way that similar motives have aided the industrial safety movement. It is held that an employer would be slower to make an unwise expansion of his business if he knew that, in addition to the ordinary risks, he had to assume the risk of continuing a portion of his wage bill, even though his plant were idle. Furthermore, the possibility of this charge being continued without a corresponding income would be an additional stimulus for him to find continuous work for his own group of workmen. The desire to avoid the payments of such compensation would have the same effect on a business executive as in the case of workmen's compensation for industrial accidents,

namely, it would put a premium on removing the cause for such payments.¹

This is the main argument, but it should be recognized that there is a marked difference between industrial accidents and industrial depressions so far as the business executive is concerned. Accidents are to a very large extent controllable by the individual firm, whereas industrial depressions are due to causes that lie mainly outside of its control. Hence, the appeal to the economic motive would work differently in the two cases, and this would be true even if the appeal were made to an association instead of to an individual firm, so that the argument from the experience of workmen's compensation is not especially apt when applied to unemployment due to fluctuations in business of the character described in connection with the business cycle. This proposal can be effective only in so far as it is possible to regularize industry through private control. Speaking for the immediate present, there is little to be expected from this proposal as a means of correcting the fluctuations of the business cycle.

The most hopeful line of relief lies in the direction of the study of the cyclical movements and of devising index numbers of prices and production in such a form that business firms may individually use them as an aid in determining their production policy. The furnishing of such information to business concerns has become an important business in itself. As our knowledge of the nature of the cyclical movements increases and the methods of dealing with the varying forces involved are perfected, such services will become even more useful to business men than now. To use such a service as a basis for business judgments requires a familiarity with economic principles and statistical methods that business executives in the past have not possessed. Perhaps the discussion which this subject is receiving, both in scientific circles and among practical business men, will result in the development of a type of business executive who will be able to use such services effectively in the determination of production policies. The application of scientific methods in the conduct of business ought to yield as large a return as the application of engineering principles has on the technical side of production. Society as a whole will benefit by any improvements in the methods of adjusting production and consumption.

¹ Professor John R. Commons, of the University of Wisconsin, is one of the chief champions of this cause in this country.

Suggested References

- EDIE, LIONEL D., "Economics, Principles and Problems." Chaps. XXVI and XXVIII.
- ELY, RICHARD T., "Outlines of Economics," 4th Ed., Chap. XVII.
- FAIRCHILD, FURNISS, and BUCK, "Elementary Economics," Vol. I, Chaps. XXIV and XXV.
- JONES, E. D., "Economic Crises."
- LAWRENCE, JOSEPH S., "Stabilization of Prices."
- MITCHELL, WESLEY C., "Business Cycles."
- , "Business Cycles and Unemployment," a report of the National Bureau of Economic Research.
- MOORE, H. L., "Economic Cycles—Their Law and Cause."
- PERSONS, WARREN M., "Indices of Business Conditions," in "The Harvard Review of Economic Statistics," Preliminary Vol. I, Nos. 1 and 2, January and April, 1919.
- RUFENER, L. A., "Principles of Economics," Chaps. XXIX and XXXI.
- SECRIST, HORACE, "An Introduction to Statistical Methods," Rev. Ed., Chap. IX.
- VANDERBLUE, HOMER, "Problems in Business Economics," pp. 3-62. This entire volume is composed of material that illustrates some phase of the fluctuations in the movement of business.

CHAPTER XV

INTERNATIONAL TRADE AND SETTLEMENT

In the treatment of the mechanism of exchange the discussion, up to this point, has proceeded on the assumption that all trade is domestic and that there are no international complicating influences. For the purposes at hand, this assumption had merit, for international trade does not differ essentially from domestic trade. Because, however, there is so much confusion of thought and national prejudice concerning foreign trade, it seems desirable to treat this subject separately.

The advantages of specialization are readily understood and accepted so long as attention is directed to what has been called occupational division of labor. Thus it is generally recognized that both A and B will gain if one devotes all of his time and energy to the growing of agricultural crops and the other to the manufacture of shoes, and they then exchange their products. Both will have more wealth than if each had produced shoes and grain. But when the same principles are applied to territorial specialization, confusion of thought and some conflict between individual and social interest arise. This is especially true if the specialization takes place within different national boundaries and involves international trade.

A few examples of the way in which popular thinking runs on this topic will serve to acquaint the student with its pervasiveness and to act as a challenge to his understanding of the essential principles of all trade. Every community is likely to have a slogan of this character: "Patronize your neighborhood stores," "Buy it in Evanston," or some other phrase of similar import. These slogans are supported by some such argument as follows: "Money spent in local stores stays at home and circulates in the community. You buy a big beefsteak with your dollar and the butcher buys from the grocer, the grocer buys from the dry goods store, and so on around the circuit."¹ Such arguments are usually accompanied by reasoning of this kind. "Every-

¹ HAMILTON, W. H., "Current Economic Problems," p. 286.

thing bought in the city takes just so much money out of the town." "Down with the parcels post. No more diabolical device was ever perfected by the big cities for stripping the small towns and country districts of all their surplus cash. Let the rich mail-order houses wax fat with the dollars that are the property of the local merchants."¹ "If I were mayor, and had my way, I would place a fine of \$100 on every man who ordered goods from a mail-order house." "The earthquake at Santa Barbara, California, was a great boon to the laboring men in the building trades and stimulated business in that community because it created a demand for labor and for building materials of all kinds."

These illustrations will be sufficient to recall to the observant reader many arguments of similar character which he has heard in his own community. The starting points for clarity in the analysis of this subject are the advantages that accrue from simple division of labor. The fallacy which underlies such slogans and arguments as those given above is the acceptance of an erroneous theory of what constitutes demand, together with a failure to distinguish between the total amount of wealth available for the community and the pecuniary advantage of certain individuals. It is not difficult to see the sequence of events that flows from a disaster such as that which befell Santa Barbara. If the town is rebuilt, then there will be a demand for labor and building materials, and the influence of this demand will be felt all along the line to the producers of the raw materials. What is not so apparent, but just as real, is that the destruction of these buildings which was the occasion of this demand (not its cause), had shut off an alternative line of demand, which would have called for labor and materials of some kind. This loss is equal to the value devoted to the reconstruction of this city. The destruction created no power to produce nor to exchange, so that reconstruction had to be carried on by means of purchasing power that would have otherwise been available for other avenues of expenditure.

The same line of reasoning applies to such slogans as "Buy it in Evanston," which means, of course, that you should make your purchases from local merchants. What is apparent in such cases is that, if you do patronize local merchants, they, as indi-

¹ *Ibid.*, p. 284.

viduals, will have more purchasing power at their disposal. If they follow their slogans in the disposal of their own income, there will be a demand for the products of some other local producers. What is not so apparent is the fact that local products may have greater purchasing power when traded in some other market, if the producers for the other market have lower unit costs than those for the local market. Under such circumstances, buying outside of Evanston would give the community as a whole a larger volume of wealth than if trading had been confined within the area of the local group. But it should be observed that all local persons engaged in the production and sale of goods that compete with those purchased outside of the local area, will find that outside purchases lessen their pecuniary income, hence, they undertake by means of advertising and slogans similar to those given above, to divert this purchasing power to themselves even though they are unable to furnish the products or services as cheaply as they can be secured from some other source. If the logic of such slogans were strictly followed, it would eliminate all territorial or interregional specialization and trade.

It should be noted at this point that demand is fundamentally coincident with production and can be no greater than the value of the goods produced and offered for exchange. This fact is sometimes called "Say's law."¹ This principle works somewhat as follows: One group of workers produces goods which they sell for money and with this money they buy the goods which another group of workers produces. In this way, demand tends to coincide with the goods produced and all trade is essentially of the nature of barter, just a swapping of goods in the market through the instrumentality of money and credit. If a more effective method of production, or a more economical way of marketing goods is discovered, it is to the advantage of the members of a community generally to make use of these new methods, for by so doing the productive energy of society will yield a larger flow of wealth. This statement is true when we speak of the welfare of the whole group, although there are always some individuals who suffer by the introduction of the more effective methods. Those who do not, or cannot, take immediate advantage of the improvements will lose, while consumers who get the goods at a lower unit cost will be in a

¹ TAYLOR, F. M., "Principles," pp. 196-203.

position to gratify their desires more completely and thus enhance their economic welfare.

The Differences in Productive Power the Basis of Trade.—From the foregoing discussion a general statement may be made that all permanent trade, whether local, interregional, or international, rests primarily upon differences in productive power. This principle is basic and underlies all forms of specialization. However, we must recognize that those engaged in trade are following pecuniary motives, and any differences in production that are shown by differences in pecuniary costs will result in trade. For instance, if the general level of remuneration of labor is sufficiently different in two countries, this fact might so affect costs as to overcome natural advantages enjoyed by one of the countries and enable the producers in the other country to gain a profit by the sale of goods based on the lower costs due to this cause. While this is possible and must be recognized as affecting the trade between countries, its continuance rests upon the continued willingness of the laborers to accept this remuneration. If there were a marked difference between the value of their services and their wages, this condition would certainly tend to disappear as soon as this fact became known to the laborers. They would soon demand the full value of their productive effort, which in turn would throw the conditions of trade back upon the differences in the productive powers of the trading countries. While recognizing temporary departures from the general principle, it still remains true that differences in productive power are the more fundamental cause of trade whether viewed interregionally or internationally.

It would follow from this that each territorial region would benefit by specialization in the production of some one or more commodities for which it is best fitted, and engage in trade with some other section that had similarly specialized. The truth of this principle is generally accepted so long as its application is confined within national boundaries. Most persons are willing to admit that it is advantageous for California to specialize in the growing of citrous fruits and to trade them for grain grown in Illinois and Iowa. Both sections will have more grain and fruit by interregional specialization of this character.

The attempt, however, to apply the principle to regions that lie within different national boundaries immediately raises the most vigorous skepticism. For this reason, it will be well to

inquire how territorial division of labor works between different countries. For the sake of simplicity let us suppose that a shoemaker and a carpenter live on opposite sides of the Canadian boundary. Will the imaginary line that establishes the political boundary between Canada and the United States have anything to do with the economic advantages that flow from this occupational division of labor? Or, put the case in another way and assume that grain growing can be profitably prosecuted on one side of the international boundary while the production of automobiles is more effective on the other. Under these circumstances, will the existence of the imaginary line separating those two industries destroy the economic basis for trade and thus eliminate the mutual advantages of specialization? Put in this way, it is apparent that economic principles are not confined by political boundaries, but wherever there are permanent differences in productive power between different territorial regions there will be mutual advantages in specialization and trade.

Viewed from the standpoint of the whole group, there can be no gainsaying the truth of this general statement. The total volume of wealth produced would clearly be greater by following such an organization of industry than if each region endeavored to be economically independent. But business is operated for a monetary return and while in general the receipt of money profits coincides with social interests, this relation does not invariably exist. The greater the degree of competition, the more nearly will pursuit of profits conform to general interests. But the fact, that through some form of restraint, such as a tariff, larger business profits may be obtained by some persons, may obscure the fundamental advantage of specialization and trade. This fact can be illustrated by the general attitude toward exports and imports.

Most persons can readily see the advantages of selling goods abroad but may question the gain of imports, especially if the imports should enter into competition with domestically produced goods. Those engaged in the various steps of exportation sell the rights to goods for a monetary price and receive a monetary profit, but in the settlements, as we shall see, credit instruments that arise from the sale of exports by some other country (imports into our own country) are generally the means of making the payments. In other words, goods are exchanged for goods. When these imports come to be sold in competition with those

made at home, the effect is to lower the price and, thus, the profit on the domestically produced goods. At this point a definite cleavage may arise between the individual and the social points of view, the recognition of which helps to explain the average business man's faith in the beneficent effects of protective tariffs. The interference with his profits is positive and evident, while the gains to the community in the form of a larger volume of wealth are not so apparent. It looks as if the foreigner is underselling the domestic producer as a means of capturing the market, so that it becomes an easy matter to appeal to national prejudice and thus secure restraint on imports. Whereas, as a matter of fact, except as qualified below, the imports are the direct result of the exports and constitute the most important method of paying for them.

International Trade.—The trade between two countries consists of two different classes of items, namely, exports and imports, commonly known as the *visible items*, and credits and services in one form or another, commonly known as the *invisible items*. Before further exposition of the general principles it will be well to get clearly in mind what is included in these two important classes of international trade.

The Visible Items.—The most important phase of trade between countries is the export and import of commodities. The United States exports food stuffs, many kinds of raw materials, and an increasing volume of manufactured goods. It imports teas, coffees, rubber, silks, woollens, many other manufactured commodities, and some raw materials. These are some of the items that appear in the statistics of imports and exports. They are called the visible items because it is difficult to conceal their movement. Commodities of this character are usually brought into a country by vessels or railroads and, since most countries levy duties or imports on foreign-made goods, the quantities must be declared for this purpose.

The Invisible Items.—Transactions between countries are not confined to a purchase and sale of commodities. In normal times, there is a large volume of financial transactions of one kind or another, which constitutes a very significant proportion of the total volume of international trading. Both public and private borrowing take place on a very large scale. National governments may borrow for the conduct of a war, as in the case of the Allies during the World War, or for the establishment of

the gold standard monetary system, or for any other purpose. Municipalities may borrow to make needed improvements, and private industries may find their funds to make extensions or improvements in their productive capacity. These few examples will serve to illustrate the great variety of purposes for which credit is extended between countries. It is immaterial for the purposes of this discussion, whether the credit has been extended by one government to another or whether the lending is by large banking institutions to public or private borrowers abroad. The evidence of the loans will take the form of bonds or other types of securities. It may be that private individuals may buy the stocks of industries in another country, which constitutes a lending of capital funds to the borrowing country. It was stated before the World War that foreigners held \$6,000,000,000 of the securities of American industries. All such investments are in the nature of an extension of credit between countries.

The extension of credit by one country to another, whatever form it may take, will create the right in the borrowing country to buy goods in the lending country. The fact that the borrowing country may make its purchases in a third country changes the procedure but not the essential fact stated, for in such a case the borrowing country has simply transferred its right to purchase in the lending country to citizens in the third country. The important point to note here is that, while the commodities purchased with the proceeds of a loan will appear as visible items, the securities issued as evidence of the obligations do not so appear. They constitute, at the time the loan is made, a debit against the lending country and are among the items that are classed as invisible. This illustration is sufficiently typical of a variety of financial transactions to convey an idea of what is meant by the invisible items in international trade.

A brief listing of the major items that come into this class of trade will be helpful in understanding its significance. The negotiation of international loans has already been mentioned. After the loan has been made, the lending country will receive interest payments on the principal account. These payments create the right to purchase commodities in the borrowing country and will likely swell the importation of goods into the lending country. The imports into the lending country will appear as visible items, but the coupons or other evidences of claims for this interest will not appear as a visible export, although they

have the same effect as an export from the lending country when an attempt is made to establish a balance of trade for that country. By means of these credit claims, the visible imports may exceed the visible exports by an amount that is equal to the interest payments received. These interest payments, then, have to be included among the invisible items of trade. The same is true when the principal account is paid, as it too will swell the imports of the lending country without a corresponding sum appearing in the visible export account.

Another item that has the same general effect is that of immigrant remittances. The situation in the United States is peculiar in this particular because of the large number of immigrants who send money to their friends and relatives in their home countries. These remittances are usually made by means of some form of a credit instrument. They create a right to buy American goods and tend to swell our visible exports to the sum of these remittances. Our visible imports will not show evidence of these claims, hence these sums have to be classed as invisible items.

A similar claim is created by tourist expenditures. When Americans travel in foreign countries they have to secure a letter of credit, or some other credit instrument, which will enable them to trade American money for the money of the country in which they travel. Thus, credit instruments arising from this source create claims against this country and tend to increase our exports of commodities. There is no corresponding item among the visible imports so that such expenditures come within the classification of invisible items of trade.

Claims of a different character, but having a like effect, are those that grow out of services performed by the citizens of one country for those of another. For instance, ocean vessels are owned largely by Europeans, so that freight charges for carrying American goods to other countries have to be paid and these become a charge against our exports. The charge for the transportation service is in the nature of an import but it is an invisible item.

Insurance charges have the same effect. In the case of a concern like Lloyds in England, which accepts risks in all important countries, the premium payments received by this company for insurance written in the United States create claims against American goods without a corresponding item among imports.

These payments are made for a service rendered but that service cannot affect the import figures. All forms of bankers' and brokers' commissions and charges have the same effect as do all other forms of services performed by the citizens of one country for those of another.

These are some of the most important of the invisible items that make up the total transactions between countries. The following table adapted from information compiled by the Bureau

THE BALANCE OF INTERNATIONAL PAYMENTS OF THE UNITED STATES IN 1926

Credits	Mil- lions	Debits	Mil- lions
Visible items:		Visible items:	
Exports of merchandise ¹ ...	\$5,038	Imports of merchandise ¹ ...	\$4,590
Exports of gold.....	116	Imports of gold.....	214
Invisible items:		Imports of U. S. paper cur- rency.....	40
American securities sold abroad.....	636	Invisible items:	
Foreign stocks and bonds sold abroad.....	286	New American investments abroad.....	1,332
Resale and new direct in- vestments in United States by foreigners.....	232	Securities bought abroad..	649
Previous investments paid.	270	Freight payments paid....	197
Freight payments received	129	Tourist expenditures paid..	761
Tourist expenditures re- ceived.....	178	Interest paid to foreigners..	228
Interest payments received	735	Immigrant remittances paid	322
Immigrant remittances re- ceived.....	35	U. S. Government expendi- tures.....	68
War-debt receipts of U. S. Treasury, principal and interest.....	195	Charitable contributions...	46
Other U. S. government receipts.....	17	Insurance paid.....	70
Motion-picture royalties...	75	Miscellaneous expenses....	25
Insurance received.....	80		
Net change in international banking accounts.....	359		
Miscellaneous receipts ²	161		
	\$8,542		\$8,542

¹ This item includes "silver" as a merchandise.

² This item includes \$150,000,000 which the Bureau of Foreign and Domestic Commerce allows for errors and omissions in the estimates.

of Foreign and Domestic Commerce conveys an idea of the significance of the various items in the international trade of the United States with the rest of the world.¹

This table shows very clearly that when we are considering the total transactions between countries, we have to include more than the export and import of commodities and of gold in order to get a true picture of the influences that affect international trade and the methods of making international settlements. These facts will receive further consideration when we examine in greater detail the methods of settlement. We will now revert to a further analysis of the principles of trade between countries.

Absolute and Comparative Advantages.—Differences in productive power have been given as the fundamental basis for trade between two territorial sections. This trade will take place between two sections of a country, or between two different countries, if one of them enjoys either an absolute or a comparative advantage over the other in the production of one or more commodities. Absolute advantages are most pronounced when countries lie in different zones. Tea and coffee are products of tropical regions, while cereals are well adapted to the temperate zones. Many illustrations of trading of this character might be cited which most persons would admit were mutually advantageous to every one in the trading countries.

The case is by no means so generally accepted, when both countries can produce all of the commodities that enter into exchange, and specialization depends upon comparative advantages. For purposes of this argument, let us assume that both steel rails and cutlery can be produced in the United States more cheaply than in England. Does it follow, therefore, that it is advantageous to specialize on one of these products and import the other? If the United States has a comparative advantage in the production of steel rails and England has a comparative advantage in cutlery, it will be mutually advantageous for the two countries to specialize and trade with each other. This is what is known as the *principle of comparative costs*, which may be stated as follows: Where two or more countries can each produce all of the goods that enter into exchange, it is advantageous for each country to specialize on those commodities in the

¹ This table is adapted from Trade Information Bulletin 503 of the U. S. Bureau of Foreign and Domestic Commerce.

production of which that country enjoys a comparative advantage in costs.

Let us assume, in order to make this principle clear, that in the United States it requires nine days' labor (expressing all costs in terms of labor) to produce a ton of steel rails, while in England the production of the same amount would require sixteen days' labor; further, let us assume that it requires three days' labor in the United States to produce a definite quantity of cutlery while the same amount in England would cost four days' labor. It is apparent that in each case the United States has an absolute advantage in production, but it has a greater advantage in producing steel rails than in the production of cutlery. A ton of steel rails would exchange in the United States for three units of cutlery, but would command as much as four units in England, so that it would be advantageous to the United States to specialize on the production of steel rails and trade them for cutlery so long as a ton of steel rails would exchange for more than three units of cutlery.

On the other hand, it would be advantageous in England to specialize on cutlery and buy steel rails so long as a ton of steel rails could be bought with less than four units of cutlery. The reader should understand that the principle here stated does not depend upon the figures given in this illustration, but whenever a country has an absolute advantage in the production of two commodities, but has a greater advantage in the production of one than in the production of the other, it is advantageous to specialize on that commodity in the production of which it has the larger comparative advantage. In the above illustration, the greater advantage in the United States was in steel rails, while in England it was in cutlery. This principle is of very wide application and is based on the one previously stated, namely, that all permanent trade is the result of differences in productive power.

The existence of a comparative advantage is affected by the principle of diminishing productivity, which may limit its application in providing a basis for trade. For instance, we have a decided advantage in the growing of wheat, but if we undertook to supply the world with that commodity, it would require a greater acreage than is now devoted to it and a much higher degree of intensity of cultivation. Consequently, the costs would doubtless rise above those that would obtain in some other

countries. Whenever this occurred, such countries would no longer import wheat from us but would find that it was better to grow their own supply unless, of course, they had a larger comparative advantage in producing some other crop. In this way, the principle of diminishing productivity will tend to set limits on the operation of the principle of comparative advantage in connection with the production of any one commodity.¹ The principle is fundamental, however, as the basis for our reasoning concerning the trading relations between countries. In its general application, there are many complications that need not be discussed in a text of this character.

From the discussion of the fundamental nature of trade given above, a few general maxims may be stated concerning international trade. First, it is economically desirable to import commodities when the exporting country enjoys an absolute advantage in producing the goods which it exports. In the second place, a country may advantageously import a commodity even if it has an absolute advantage in producing this good, provided it enjoys a greater advantage in producing some other commodity. Generally speaking, it is advantageous to import a good whenever the capital and labor used in its production would yield a larger return, if devoted to the production of some other good.

Some Implications of Comparative Advantage.—The acceptance of the foregoing principle will throw light on some popular conceptions concerning international trade. First, it is generally thought that a low-wage country has an advantage over a high-wage country. This is not necessarily true unless, as indicated above, the low wages actually result in low unit costs. The decisive factor in international trade, as in all trade, is unit costs. High wages may be the result of high productive efficiency obtained by superior technical methods of production, or from the use of superior natural resources. Many American products sell in European markets in competition with domestically produced goods, notwithstanding the higher level of wage rates in this country.

Second, it is often argued that a country with a low standard of living can undersell one with a high standard. This argument is closely associated with the first because the standard of

¹ For a more detailed exposition of this principle see BLACK, J. "Production Economics," pp. 129-153.

living is the result of the income received. Here again it is the differences in unit costs that constitute the basis of exchange rather than the standard of living. India and China, with much lower levels of living, have not been able to compete with England and the United States, where machine methods have been extensively used and have given lower unit costs. In fact, there is within India now an attempt being made on the part of laborers with a very low standard of living to compete with machine-made goods. The machine processes are gradually being introduced in that country and the low standard of living of the hand laborers is not preventing machine-made goods from underselling those made by the old, primitive processes. In both of the above arguments it is tacitly assumed that lower wages and lower living costs are accompanied by equal productive efficiency—a condition that seldom obtains. Trade advantages always rest with a country that has lower unit costs.

An inference is frequently drawn that a country benefits when the percentage of manufactured goods in its exports increases. It is thought that additional profit is made through the processes of manufacture. Here again the case turns on whether labor and capital used in one way are more productive than when they are used in some other way. If the productive energy devoted to the extractive industries gives greater comparative advantage in trade, a country will not benefit by the growth of manufactures. It is only when, through the operation of the principle of diminishing productivity, labor and capital become more productive in manufacturing than in the extractive industries that a country gains by the development of manufacturing. The guide post here, as in the previous instance, is unit cost of production.¹

International Dependence.—Specialization and territorial division of labor have developed an increasing dependence of one country upon another. The economic advantages of specialization and trade are not confined to national boundaries and, therefore, there has been an enormous growth in the volume of international trading during the past century. The result is that an increasing number of people in the trading countries are dependent upon the sale of their products abroad for a continuity of their livelihood. Any interruption of industry in one country is, therefore, likely to be felt in the other countries. Measured in terms of goods, all countries benefit by specializa-

¹ Edm. L. D., "Economics, Principles and Problems," pp. 678-684.

tion, but conflict arises when nations endeavor to get control over the unexploited sources of raw materials with the view of restricting their use on nationalistic lines. The gains from such policies are more apparent than real.

Suppose, for purposes of illustration, some new resources have been discovered and are developed. The effect will be to provide a larger supply of the commodities made from such resources, which will lower the trading power per unit as compared with that of other commodities exchanged for it. So long as trade is uninterrupted, as by war or restrictive legislation, all trading countries will benefit by the opening of new resources, irrespective of the control under which they are developed. This will follow because with a given expenditure of productive energy a larger volume of wealth is obtained and will be shared to some degree by all trading countries.

To illustrate this point further, let us suppose that two countries A and B were engaged in trade prior to the discovery of the new source of supply. For sake of simplicity, let us assume that country A had been devoting $2x$ days of labor and capital to the production of wheat and secured 40 bushels, and that B had been devoting $2y$ days of labor and capital in mining 35 tons of coal. Let us assume that country A gave 20 bushels of wheat for 10 tons of coal. Now suppose that B discovers new resources of coal so that the $2y$ days of labor and capital will produce 50 tons of coal with no change in the amount of wheat produced. If 20 bushels of wheat exchanged for 10 tons of coal before the discovery of the new supply, it will now exchange for more than that amount, let us say for 17.5 tons.¹ Before the new coal was discovered, A had 20 bushels of wheat and 10 tons of coal, while B had 20 bushels of wheat and 25 tons of coal. Now A has 20 bushels of wheat and 17.5 tons of coal, while B has 20 bushels of wheat and 32.5 tons of coal. Both countries have gained even though country B controls the new supplies.

The reader must remember that the figures chosen are not the significant facts to be gained from this argument. All that is claimed is that, under unrestricted conditions of trade, the

¹ The new ratio of exchange cannot be expressed with precision because the rate of diminishing utility is not known. We do know that coal in relation to wheat will be less valuable than before the new coal supplies were discovered. This fact gives ample basis for concluding that the rate of exchange has shifted in the direction indicated by the figures given.

benefits from the discovery of new resources of raw materials are mutually beneficial to all trading countries. It is not claimed that the trading countries will share equally in the benefits from these new supplies. The question as to which country gains most in international trade is an extremely complicated one and is not brought into this discussion, as it is not essential to the principles that are being set forth. The conclusion reached is in accord with those set forth in the discussion of the advantages that flow from territorial specialization.

The conflict between nations over new materials is based partly on a misunderstanding of the effect of unimpeded economic forces on the economic welfare of all trading peoples, but more particularly on the fear of war and the belief that additional national strength would result from the control of such resources. This attitude of mind is the direct product of nationalism under which the peoples of all countries are living. It will continue unless at some time an effective method of control is developed that will insure freer access by all nations to the benefits flowing from the sources of raw materials. In the mean time, the operation of the principle of mutual advantage from territorial specialization will frequently be interrupted by restrictive legislation and nationalistic conflicts.

International Settlements.—The basis for trade and the advantages flowing therefrom have been set forth above. The next question is, How are settlements between two countries made? In domestic trade, settlements are by means of money and credit instruments in general use within the country, but in international trade a new problem arises when payments are made between countries. Since monetary systems are confined to national boundaries, it becomes necessary to find a method of effecting international settlements that is mutually acceptable. Generally speaking, it may be said that gold has become the money of international trade, which means that obligations between countries are quoted in terms of gold, and when a medium of exchange is required to make international payments, gold is generally used for that purpose. This does not mean that two countries using silver as their money do not trade with each other in terms of silver, or that a gold-using and a silver-using country could not exchange goods, even though the making of settlements between such countries would be complicated by monetary units of varying values. What is meant is, that most countries buy

and sell on the basis of gold money prices, and gold is generally used in making international settlements.

Par of Exchange.—Importers and exporters buy and sell, as all traders do, in terms of prices, and prices are always quoted in money units of some country. One of the first problems in connection with exchange between countries is the equating of the monetary units, which consists in comparing the gold in the monetary unit of one country with that of the unit of another country. For instance, when we compare the amount of gold in the dollar with that in the pound sterling in England, we find that the pound sterling is composed of 4.8665 times as much pure gold as the gold dollar. When the exchanges between England and the United States can be made on this basis, we say exchange is *at par*. Par value of exchange between two gold-using countries is always found by comparing the amount of gold in the standard unit of money in one country with that in the other. The problem becomes somewhat more complicated when silver is used by one of the trading countries, although the principle is approximately the same. The only difference is that the value of the two metals does not always move in the same direction at the same time, so that par of exchange has to be continually corrected for the differences in the value of the two money materials. The same statement holds if one country is on a paper-money basis.

Bill of Exchange.—In international payments, as in domestic trade, methods have been devised for economising in the use of gold in making settlements. The instrument used for this purpose is known as a "bill of exchange," which consists of an order drawn by one party, the drawer, against another party, the drawee, directing payment to a third, the payee. The drawer must pay the bill if the drawee does not, but when the drawee "accepts" a bill, then the obligation to pay rests on him. A check drawn against a bank in favor of another person is in the nature of a bill of exchange and is a kind of "inland" bill. The bills used in foreign exchange have the peculiarity of the differences in the monetary systems already noted, *i.e.*, they must be drawn in some one monetary unit, as in pounds sterling, in dollars, in francs, etc. There are a number of kinds of bills, such as bankers' bills, commercial bills, cable transfers, time bills, etc. Since all of these bills, in one way or another, grow out of the commercial transactions between two countries, a commercial bill will be used for purposes of illustration.

A Commercial Bill.—Commercial bills arise from the export and import trade between two countries. Let us suppose that an American exporter A ships a cargo of wheat to an English importer B, the value of which is £1,000. The question is, How can A collect for his shipment from his customer who resides in England? Under such circumstance, it is customary for the exporter to draw on his English customer a draft for the amount of his sale. This draft may be payable on demand, or at the expiration of a definitely stipulated period of time. A draft of this character constitutes what is known as a bill of exchange. The American exporter could send his bill to England, collect the payment in gold, bring the gold to this country, take it to the mint and have it coined into American money. Such a process would be expensive and require a considerable period of time. There would be an expense involved in boxing, expressing, and insuring the gold during transit and, in addition, the exporter would lose interest on the money during the time necessary to transfer and coin the gold.

Now let us suppose that an English exporter C sells to an American importer D a cargo of woolen cloth to the value of £1,000. He would have a similar problem of collection as that described in connection with the American exporter A. The English exporter C could also draw on his American customer a bill of exchange for the amount owed him. If two bills were drawn, one would represent a debt owed by an American D to an Englishman C, while the other would represent an obligation of an Englishman B due an American A. Since, according to our assumption, the sums involved are equal, if the two bills could be brought together they would cancel each other and the American debtor could pay the American creditor, and the English debtor could pay the English creditor, and the whole transaction could be completed without the transfer of money between the trading countries. Suppose, for instance, that D, the American importer, should buy from A, the American exporter, his bill of exchange, paying him its face value in terms of American money. This would complete A's claim against his English debtor. Then suppose that D should send this bill, accompanied, as it would have to be, by a bill of lading from the transportation company, by insurance, and by such other documents as are required to guarantee its authenticity, to C his English creditor. This bill would convey to C the right to collect its face value from B, the

English importer of wheat. As soon as B pays C the two transactions have been settled. Thus the shipment of wheat *from* America became the basis of paying for the importation of woollens *into* America. Both transactions were conducted in terms of money, but the bill of exchange which was, in fact, a credit instrument was used in effecting the settlements between the two countries.

The method commonly used is not quite so simple as this illustration implies, but this example does set forth in bold form the essential features of such transactions. The illustration departs from actual practice in that it implies that both in England and America the exporter and importer are in contact with each other, and actually make payments as described. In fact, they are not, and the process of settlement is conducted by means of banks or other dealers that engage in the purchase and sale of bills of exchange. Instead of selling this bill of exchange to D, the American exporter A will sell it to a bank and D will buy it from this bank, or more likely the bank will use A's bill to build up its balance with a correspondent abroad and sell to D a bill against that balance. The foreign bank in this instance, or the foreign branch of an American bank, would then collect from B, the English importer. The English exporter may receive a bill from the American importer and deposit it with his bank. The English bank may then forward this bill to its New York correspondent, using it to build up its balance in New York, the collection being made from the bank from which D bought his bill; or it may collect direct from the English bank against which the bill is drawn. The entrance of banks that specialize in foreign exchange has changed the process of making international settlements, but their presence has not affected the essential fact that these exchanges can be carried on without the movement of gold. From this discussion and illustration it should be apparent that international trading is primarily barter and that exports really pay for imports through the use of bills of exchange.

Difference in Bills.—The illustration above is a simplified, documentary commercial bill. Such a bill may be either a *sight* or a *time* bill, the only difference being that the former is payable on demand and the latter runs for 30, 60, or 90 days, and requires an interest calculation to determine its present price. A banker's bill is one sold by bankers against their foreign

balances. It also may be either sight or time. For purposes of understanding the simple principles, it is not necessary to go into a detailed description of the various kinds of bills that are actually used in international exchange. It is sufficient to know that by means of credit instruments of the character described the movement of gold bullion is economized.

Gold Points.—From what has been said, it is clear that bills of exchange are bought and sold and, by implication at least, it may be inferred that they have prices and are subject to the influences of demand and supply as other marketable commodities or goods. It so happens that exporting and importing of commodities do not proceed on an even basis. During some parts of the year the flow of goods out is likely to be greater than during others, and the same statement can be made concerning imports. The result is that the supply of bills, that is, the amounts receivable by one country is greater than the demand or the volume of indebtedness. Under such circumstances, the price of bills of exchange will fall below par. Let it be understood that the purchasers of bills are those who have settlements to make in a foreign country, while the supply arises from those who have sold goods and therefore have collections to make. If the supply of bills exceeds the demand, the price will fall, and the question now is, How much will the price of bills depart from par and what fixes these amounts of fluctuation?

A bill of exchange sells at par when it brings its face value, as the £1,000 in the illustration above. If the exports and imports just equal each other, we might conclude that exchange would be at par, but this conclusion must be qualified by the statement that exchange will be at par when all payments due a country, from whatever source, equal those owed by that country. Whenever the supply of bills exceeds the demand, the bills will fall below par and the limit of this deviation is normally fixed by the cost or expense involved in the actual movement of the bullion. These expenses consist of boxing, expressage, insurance, loss of interest while gold is in transit, and any other expense that may arise in connection with the actual movement of the gold. In normal times, these expenses amount to about 2 cents per pound sterling, hence, bills may deviate up or down from par by that amount, or from \$4.84 to \$4.88. These points are known as the "gold" or "specie" points, for when the price of bills reaches one of these points, gold bullion begins to move in settlement of

international obligations. If the seller of a bill of exchange could get only \$4.84 per pound sterling, he would realize as much from assuming the expense of collecting the gold and having it shipped. In like manner, when bills are scarce and sell for more than \$4.88, importers could afford to buy gold bullion and assume the expense of shipment.¹

International Balances.—Fundamentally, bills of exchange depend upon the volume of exports and imports, but there are other items and transactions that affect the price of bills that need to be taken into account. In addition to bills growing out of export and import of commodities, the following items must be considered: First, investments by one country in another. When investments are being made they will show themselves ordinarily as an excess of imports into the borrowing country. English citizens invested heavily in the construction of our canals in the 'twenties and the 'thirties, as they did also in our railroad securities in the 'seventies and 'eighties. When these investments were being made, they gave rise to bills that enabled American importers to buy commodities abroad without a corresponding flow of goods out of the country in settlement of these purchases. Second, the making of interest payments. Once the investments have been made, interest payments begin on the principal account and these payments show themselves in an excess of exports from the borrowing country. We would, therefore, expect a debtor country to show a larger volume of exports than of imports, and a similar situation would be found when payments were made on the principal account. Third, expenditures of travelers. When an American travels in England he must buy a letter of credit or some form of credit instrument which enables him to use English money in paying his traveling expenses while abroad. If there are more Americans than Englishmen traveling, then the balance of claims from this source is likely to show itself in

¹ The actual transfer of gold between countries is handled by banks rather than by importers and exporters of merchandise, as is here implied. There is nothing but the inconvenience, however, to prevent an importer or an exporter from contracting for gold in the manner described in the text. For purposes of simplicity in the exposition of this topic, the discussion has proceeded on the assumption that such contracts are made. Even if this assumption does do some violence to the actual practice in the movement of gold, it makes for greater simplicity in explaining the causes for the limits on the fluctuations in the price of bills of exchange and, thus, in establishing the gold points.

an excess of American exports. Fourth, remittances to friends. In the United States there are a large number of immigrants who send money "back home." What they normally do is to send some form of a credit instrument which gives the home country a right to demand goods in payment of these claims. Where such remittances show a balance against a country this balance is likely to be settled by an excess of exports. Fifth, freight charges and payments for other types of services. If a larger volume of freight is carried by vessels owned by the citizens of one country than by those of another, the differences in these charges will appear in the trade balances between the countries. All of these factors affect the total supply of bills and, hence, affect their price. The rate of exchange may move upward or downward because of a change in the supply of bills that have originated from other causes than a shipment of commodities.

Three-cornered Exchange.—In balancing bills of exchange between two countries, all of the above items and any additional ones that may give rise to international payments besides exports and imports, would have to be taken into consideration. When bills from all sources are balanced, there may be a credit or a debit balance in favor of one of the countries. Such balances may be settled by a shipment of gold. However, there are still other possibilities of adjustment. The foreign trade of a country is not confined to trade between two countries, as the discussion above seems to imply. In fact, in making international settlements, the total export trade of a country is balanced against the total import trade. It may happen that the United States sells more agricultural products to European countries than it buys of European goods. In that case, American bankers would have balances in Europe. Now, if purchases from South America exceed our sales to those countries, American merchants could buy bills on the European balances and transmit them to their South American creditors. These bills would enable South American merchants to buy more European goods and pay for them by their excess sales in the United States. By this means, the total trade of a country can be balanced without the movement of gold bullion.

In addition to the use of credit instruments of the kind described as a means of making payments, bankers often sell bills in anticipation of subsequent trade. For instance, the shipment of grains and agricultural products is likely to be heavier

in the fall of the year, while imports are likely to be heavier in the spring. If in late spring foreign balances run low, bankers may sell bills in anticipation of those that will arise from an excess of exports a few months later. By this device, the inequality of transactions between different seasons may be balanced and the shipment of gold avoided.

Arbitrage.—Banks and brokerage houses deal in bills of exchange. They buy and sell exchange to those who have international payments to make. The specialized dealers tend to become concentrated in the financial centers of the respective countries such as London, New York, Paris, Berlin, and so on. The rates of exchange may be such that the brokers can make a payment in London to their own advantage by selling exchange in Paris and by buying bills there on London, or even on Berlin and then on London. This process of buying and selling exchange in the most favorable market is known as "arbitrage." By this process there is an equalization of exchange rates among the principal financial centers of the world. London has long been the clearing house, as it were, of the financial and trading operations of international trade. By means of arbitrage, rates in the different countries are kept on approximately the same basis.

Gold Settlements.—When all of the methods of economizing in the use of gold in making international settlements have been exhausted, gold bullion will then be shipped. The effect of gold shipments should be carefully noted. In the country from which the gold is exported the bank reserves will be drawn down, thus limiting the possibility of extending credit. Discount rates are likely to rise, and money prices to fall. It will make this country a more favorable place in which to buy and less favorable in which to sell. In the country to which the payment is made, bank reserves will be increased, discount rates will tend to fall, and money prices will tend to rise. The higher prices will stimulate sales in the latter country and discourage buying. Thus, by an interaction of forces, a balance of international payments that results in a movement of gold starts trading along lines that will eventually cause the gold bullion to flow back to the original country until the gold holdings tend to establish a balance in the price levels of the trading countries. This statement assumes a free movement of gold which may not be permitted by governmental or other conscious action, in which case, if the gold enters into the monetary system it will cause a permanently higher price

level in that country. Banks may discourage the export of gold by raising the discount rate. In a center like London, where the commercial banks follow in the main the lead of the Bank of England, an increase of the discount rate not only discourages borrowing, but makes the London market an attractive place in which to lend. In this way, the Bank of England exercises a powerful influence on the flow of gold in and out of the London market. Governments may impound the gold and prevent it from leaving the country, as was done in the United States during the World War.

Favorable Balance of Trade.—We are now in a position to examine a topic that always receives comment in *financial sheets*, and concerning which there is widespread misunderstanding. It is very generally believed that a favorable balance of trade is desirable and that an unfavorable balance discourages and depresses industry. The balance of trade is said to be favorable when a country exports more commodities than it imports. It is generally assumed that the country which sells more than it buys is prosperous, a line of reasoning that is carried over from the observation of the economy of an individual and is influenced to some extent by an economic philosophy of the sixteenth to eighteenth centuries, known as Mercantilism. It is a well-recognized fact that the individual accumulates and gains in economic power by selling more than he buys. Reasoning from this experience, it is an easy step to conclude that a similar policy is economically advantageous for a nation.

The economic welfare of a people does not depend upon the amount of money in their country but on the volume of goods which is available for the satisfaction of their desires. We have now seen that a favorable balance of trade may mean such a variety of different things that it is unwise to conclude that the people of a country are more prosperous or economically better off on account of a favorable balance.¹ Among various things, a balance of trade may mean that the country is paying interest on its indebtedness, or is making payments on the principal of this indebtedness, or is paying to foreign citizens freight charges on goods sold abroad and carried in foreign-owned vessels, or that its citizens are traveling abroad, or its immigrant population is making remittances to friends at home.

¹ The previous table on international payments should be consulted at this point.

A favorable balance of trade may arise from any one or a combination of these causes, and in each case the country would be shipping out more goods than it was importing. It is certainly not self-evident that an excess of exports coincides with a high degree of satisfied wants in the exporting country. If a favorable balance of trade were always coincident with the economic prosperity of a people, it would be hard to understand why the Germans should resist so strenuously the payment of indemnities which certainly would show a favorable balance in Germany. It is not likely that the German indemnities can be paid without the German people shipping more goods out of the country than they import. In fact, the economic welfare of a people depends upon the goods they have to consume rather than on those they export. It would follow from this, therefore, that a people is more concerned with the nature of the imports than in the volume of exports. If by trading goods a people can secure those which will yield greater satisfactions, it may be said that they are promoting their economic welfare.

The idea that a favorable balance of trade promotes the welfare of a nation has been handed down from the sixteenth century or even earlier. It is associated with that period in the history of European nations when trading between countries was in its infancy. To trade successfully it was necessary to have an acceptable medium of exchange. Gold and silver were in great demand for this purpose. Those nations that had an ample supply became strong politically because they could trade effectively. Other nations, that were just as ambitious but less well equipped with the precious metals, tried to adjust their trading relations with other countries so that a flow of gold and silver into their boundaries would be assured. To sell more under these circumstances than was bought meant a balance paid in money. The money thus obtained fostered trade and increased the volume of goods available for consumption within the country. Nations generally at this time encouraged exports and discouraged imports in order to obtain a supply of the precious metals.

This doctrine, that an excess of exports contributes to national prosperity, has continued in popular thinking through the intervening years, and many people still believe that a favorable balance of trade is economically advantageous. Aside from the different conditions that may give rise to a favorable balance of

trade in our day, there is no longer the need for a flow of gold and silver that formerly existed, and today the payments in gold tend toward higher prices which fosters selling and discourages buying. The welfare of a people is dependent upon the amount of wealth available for the gratification of their desires rather than on the volume of money in existence. The acceptance of this fundamental fact will help students to clarify their reasoning on the whole subject of international trading.

Suggested References

- BOGGS, T. H., "The International Trade Balance in Theory and Practice."
 BLACK, JOHN D., "Production Economics," Chap. V.
 BASTABLE, CHARLES F., "Theory of International Trade."
 BROWN, HARRY G., "International Trade and Exchange."
 BYE, RAYMOND T., "Principles of Economics," Chap. XIV.
 CLARE and CRUMP, "The A B C of the Foreign Exchanges."
 CROSS, IRA B., "Domestic and Foreign Exchange."
 EDGEWORTH, F. Y., "Papers Relating to Political Economy," Vol. II, pp. 3-60.
 ELY, RICHARD T., "Outlines of Economics," 4th Ed., Chap. XVIII.
 ESCHER, FRANKLIN, "Foreign Exchange Explained."
 FAIRCHILD, FURNISS, and BUCK, "Elementary Economics," Vol. I, Chaps. XXVI-XXVIII.
 FLUX, A. W., "The Foreign Exchange."
 GIDE, CHARLES, "Principles of Political Economy," Rev., pp. 249-250.
 GOSCHEN, GEORGE J., "The Theory of the Foreign Exchange."
 GRIFFIN, C. E., "Principles of Foreign Trade."
 JOHNSON, A. S., "Introduction to Economics," Chap. XIX.
 MILL, JOHN S., "Principles of Political Economy," Book III, Chaps. XVII-XX.
 RUFENER, L. A., "Principles of Economics," Chap. XXXIV.
 TAUSSIG, FRANK W., "Principles of Economics," 3d Ed., Vol. I, Chaps. XXXII-XXXV.
 ———, "International Trade."
 TAYLOR, FREDERICK M., "Principles of Economics," Chaps. XVI and XVII.
 VANDERBLUE, HOMER B., "Economic Principles," Chaps. XVII-XIX.
 WHITTAKER, ALBERT, "Foreign Exchange."

CHAPTER XVI

PROTECTION AND FREE TRADE

In the previous discussion of international trade, it has been assumed that the trading has been free from restrictive regulations. Under such circumstances the trade between countries tends to follow the principle of comparative costs. Trade will arise and continue between two territorial regions providing one of them has a comparative advantage over the other in the production of one or more commodities, and providing, further, that the people residing within these regions desire the goods produced. This principle is the direct result of the advantages accruing from territorial specialization. The economic gain from specialization of this kind is generally recognized when the differences in costs are absolute, as in the case of coffee and corn, one a product of the tropical zone, and the other a product of the temperate zone. But as we have seen, trade is mutually advantageous to two or more countries, provided one of them has a greater advantage in the production of some commodities than it has in the production of others. If the United States, for instance, can produce both steel rails and cutlery at lower unit costs than England, but has a greater advantage in the production of steel rails, it would be economically advantageous for the people of the United States to specialize in the production of steel rails and to trade with England for cutlery.

This principle of comparative advantage is quite generally recognized when it applies to interregional trade between different sections lying within a single political jurisdiction. Few persons would seriously question the advantages enjoyed by the residents of Illinois, Iowa, and the Dakotas by specializing in grain growing and trading with California and the southern states for citrous fruits and cotton. The advantages here are obvious, even though the basis for such trade is the comparative differences in the costs of producing the goods in question. However, the acceptance of this principle is by no means so general when the trading is between territories lying within

different political jurisdictions. The recognition of these doubts and the knowledge that most modern countries have resorted to restrictive legislation is sufficient justification for considering at this point the arguments for and against free trade and protection.

The Meaning of Free Trade.—Free trade means unrestricted trade. The main argument in favor of freedom of trade is that which has been developed in connection with the advantages gained from territorial division of labor. So long as there are differences in the productive power of two territorial regions, whether these differences are absolute or comparative, there will be advantages in specialization and trade. This is the fundamental principle underlying all division of labor and specialization. Any departure from this principle should be made with a full realization that the diversion of labor and capital from channels dictated by this principle will result in a smaller volume of wealth produced. The validity of the principle is not affected by political boundary lines. If the farmers of Minnesota are economically better off by specializing in grain and buying fruit from the farmers of California, they would benefit by specialization and trade with the farmers of Ontario, providing there were similar advantages in the costs of producing the goods to be traded. The existence of the political boundary line does not affect the operation of the principle. It follows from this that a strict application of the principle of comparative costs would lead one to the acceptance of freedom of trade between nations as the condition most conducive to the economic welfare of all the people residing in the trading countries. The tariff policies adopted by the different nations, however, have not been controlled by these considerations but have usually been dictated by the special interests of the favored group together with the fiscal interests of the government itself.

The Meaning of Protection.—Protection is the advantage a domestic industry enjoys when a tax is levied on the importation of foreign-produced goods of the same kind. Suppose that woolen cloth of a recognized grade is produced both in the United States and in England, and that Congress levies a tax on the importation of woollens of this character. This tax may be either *specific* or *ad valorem*, the first being a definite amount per yard, the other a percentage of the value of the goods, the valuation being determined either at the port of shipment or the port

of entry, depending upon the provisions of the law. For purposes of this argument let us assume a tax of 50 cents per yard. It would follow, then, that if the manufacturing costs were the same in the two countries, the English producers could not compete, for in addition to the transportation charges a duty of 50 cents per yard would have to be met. The American manufacturer would not have to bear this expense so that, unless the market price rose by 50 cents or the cost of the imported goods were reduced by 50 cents per yard, the foreign manufacturer could not afford to export his goods. Whichever of these alternatives existed, the American manufacturer would enjoy an advantage equalling the amount of the tax. It is this advantage that is known as protection.

It should be observed, first, that domestic producers always enjoy what may be called *natural protection*. The nearness to the market saves transportation charges which are usually a considerable item and give the domestic producer an advantage over his foreign competitor. This advantage will always exist unless the trading countries are contiguous, in which case the costs of transporting foreign-produced goods may be no barrier to their importation. Then, there are advantages from the use of a common medium of exchange which avoids the inconvenience of quoting prices and in collecting sums due from the sale of the goods in a foreign monetary standard. In addition, domestic trade avoids the risk of loss from fluctuations in the rates of exchange. These and other similar advantages are enjoyed by the domestic producer over his foreign competitor. All such advantages may be regarded as natural protection. Improvements during the past century in the methods of transportation and communication and the development of international banking have lessened these advantages, but they are still significant influences affecting the trade between countries. The problem with which we are concerned here is not that of natural protection but the advantages which some producers enjoy as the result of the levying of duties upon the importation of foreign goods.

Incidental Protection.—Protection may be said to be *incidental* when the primary purpose for levying the import duty is the raising of revenue for defraying the expenses of the government. All governments require a revenue for the conduct of their various functions, and import duties have been found to be a

convenient means of raising such funds. If the duties are skilfully placed, they will yield large sums without much opposition on the part of the users of the goods, because the tax is made a part of the price of the commodity and is, therefore, largely disguised. A tariff duty levied primarily for fiscal purposes may afford incidental protection to some domestic producers, because the tax will give them an advantage over their foreign competitors by an amount equal to the duty. In the levying of custom duties there have been mixed motives, but an examination of the arguments of the proponents of protection clearly reveals that the fiscal motive has in a majority of the tariffs in this country been secondary to the desire to secure by means of the duties an advantage for some particular industries. If the fiscal purposes were controlling, that fact would be revealed by the character of commodities chosen on which to levy the duties.

Tariff for Revenue.—There are two main principles that would govern the selection of commodities if the main purpose of the tax were the raising of revenue. First, commodities that are generally consumed and are produced only in a foreign country would be selected. Tea, coffee, sugar, etc., usually make good revenue-yielding commodities and may, therefore, be taken as illustrative of this type of goods. The list chosen would vary from country to country because of differences in the national habits of the people. For instance, tea is a better revenue article in England than coffee, whereas in Germany coffee is a better article than tea, because of the differences in the national habits in respect to the use of the two commodities. Articles of the kind mentioned would give a regular revenue to the state, which is one of the requirements for fiscal purposes and, furthermore, the whole volume of goods consumed would contribute toward that revenue. Changes in rates on such articles would yield corresponding fluctuations in revenue because the demand for them is general and largely inelastic.

A second principle of selection would be goods widely used but produced both in the home and foreign countries. Emphasis in both cases is on wide and general consumption of the goods chosen, which means that the demand for the commodities should be inelastic. Sugar is a good example of such a good. In cases of this character, there would be incidental protection which could be nullified by an excise tax equal to the protective duty,

if the controlling motive in levying the duty were the raising of revenue. Suppose that an import duty of 50 cents per yard were levied on woolen cloth of a recognized grade and that an excise tax of the same amount were levied on the domestic production of similar woolen cloth. In this way, the protection afforded the domestic producers could be nullified by the excise tax and the treasury department would secure revenue from all of the goods consumed rather than from that portion of the supply that is imported. The protection afforded domestic producers by custom duties could be offset in this manner, if the raising of revenue were the sole purpose for which such duties were levied. It is interesting to know that this principle was adopted during the Civil War as a means of offsetting the protection afforded by the high tariff rates which the government found necessary to levy at that time. At the close of the war the excise tax was removed without a corresponding reduction in the tariff rates which had the effect of greatly enhancing the amount of protection given the American manufacturers.

The question of protective duties has been a hotly debated public issue throughout our whole national history. It was raised in the debates on the first revenue measure that was introduced into the Congress of the United States in 1789, but it received much greater attention in connection with the Tariff Act of 1816. At this time, the new industries that had been established during the War of 1812 were suffering from the importation of goods from England and an insistent demand was begun for protective relief. From this time onward, protection has been continually a political issue and has occupied a prominent place as a campaign issue in each successive general election. The different sections of the country have taken sides on the tariff question very largely in accordance with the dominant economic interests of the respective sections. The agricultural sections, and particularly the South, have favored free trade or low tariff rates, while the eastern sections, where manufacturing and commercial interests were concentrated, have strongly favored protective duties. As a result of this situation, the tariff question has been a political issue from the beginning of our country's history and, broadly speaking, the contest has been between the agricultural and extractive industries on the one side, and the manufacturing and commercial interests on the other. The public discussions of the tariff and of protection, therefore, have

been much more an attempt to influence the popular vote in behalf of some advantage for an industry than a searching inquiry into the economic effects of protective duties. The arguments advanced by the two leading political parties cannot be said to show the results of a sincere attempt to inform the electorate on the economic consequences of the legislative measures proposed.

A recognition of this fact makes it clear that the arguments of the politicians are not always founded on a careful consideration of the economic consequences of the proposed tariff changes, nor are they always reliable. Democratic politicians have claimed for years that their party has championed a tariff for revenue only, but at no time has the party platform declared in favor of an excise tax that would offset the incidental protection afforded domestic producers by an import duty. On the other hand, the accusations of the leaders of the Republicans, that the Democrats are a free-trade party, are equally untrustworthy. An examination of the tariff schedules passed during the first administration of President Wilson, when the Democratic Party was in power, will show a degree of protection that is higher than the revenue measures enacted during the Civil War. The only difference between the two leading political parties on the tariff question, if a judgment is based upon the acts passed while the party was in control of the legislative branches of the government, is that one party favors higher protective duties than the other. Most of the popular arguments, either for or against the protective tariff, are colored by the political bias of the proponent. Thoughtful students of this question will recognize this fact and will endeavor to discriminate between the political and the economic aspect of the problem. As an aid in this task, a few of the arguments most frequently urged in behalf of protection will now be reviewed.

Infant Industries.—The question of protective duties became actively debated immediately following the War of 1812. Several arguments were advanced at this time that have persisted until the present day, one of the first being the *infant industries argument*. During the War of 1812, importation of European goods was shut off. A large number of establishments were launched in order to satisfy the domestic demand for goods that had previously been imported from Europe. So long as the war continued these new industries enjoyed a high degree of

natural protection, but with the close of the war English manufacturers began to unload their surplus stocks in the American markets. Domestic producers were unable to meet this competition, partly because the English factories were able to produce the goods at lower unit costs, and partly because the English goods were a surplus supply that had been accumulating during the war and were now being sacrificed by the owners. There is evidence to show that English manufacturers did not look with favor on the development of competing factories in this country and that the goods were sold at a price that made it difficult for the American producers to compete.

There immediately arose a demand for a duty on imported goods as a means of protecting the infant industries that had been started during the war. This argument was advanced partly on patriotic grounds and partly as a matter of economic policy. It was held that, since the manufacturers had risked their capital in these new industries during a national crisis, it was only fair that the government should come to their relief in this struggle with their English competitors. It was argued, also, that the duties would not need to be permanent, but were necessary only during the early years while the new industries were being established. These arguments met with a cordial public response, as England was not in good repute in the country at this time. The tariff of 1816 was adopted and the law provided for a gradual reduction of the rates which indicates that the protection afforded by the duties was not regarded as permanently necessary. Thus began one of the most persistent arguments in favor of protection, which has been repeatedly urged by those interested in the subsequent establishment of new industries.

Economists generally have recognized that an industry may properly be helped during the initial stages in which it is being established. Industrial specialization may rest merely on the momentum of an early start, but once this advantage is obtained, it requires a considerable force to overcome it, even though the industrial conditions may be more favorable for a different localization of the industry. This was largely the case in the United States in 1816. Cotton and wool were grown here, shipped to England, and the cloth manufactured from them was imported into this country. English manufacturers had the advantage in the types of machines used, in a supply of skilled

labor, and in the subsidiary industries that had been developed and from which necessary equipment could be easily obtained. The American manufacturers had the advantages of nearness to raw materials and to the market, but at this time the transportation charges from the South to New York and to New England were not so much less than to England as to make this advantage very great. If by means of a protective duty the new industry were able to outgrow these handicaps, as it was urged at this time, and could develop independent means of competition, then the duty would be justified. In other words, whenever the general conditions of production are favorable to territorial specialization, a duty is economically sound if it will aid in overcoming the artificial advantages of an early start.

The effect of such a duty will be to transfer the industry to the more favorable territorial location where the unit costs will be permanently less. It may be urged that, if the advantages of the new location are real, this transfer will take place without the aid of an artificial stimulus, but the answer may be correctly made that the advantage of an early start will enable an industry to compete successfully for a long period with those more favorably located. If such an industry continues to enjoy the advantage of nearness to subsidiary industries, of availability of a supply of skilled labor, and of market connections which develop along with the growth of the industry itself, it can continue to prosper for a long time. A protective duty will doubtless hasten the transfer of an industry under such circumstances.

As a matter of policy, the difficulty in levying a tariff for this purpose is not with the reasoning that underlies it, but in the ability to reduce the rates once they have been established. Those who have enjoyed the benefits of a protective duty can always show that they will suffer losses if the rates are removed or reduced. These effects are direct and immediate. The benefits to those who gain by the removal of the tariff are less perceptible because they are indirect and are very widely spread among the users of the good. Experience in the United States has shown that once a tariff has been levied as a means of stimulating an infant industry, it is likely to be continued long after the initial stages have been passed and, in many instances, the rates have been increased rather than reduced. When this condition is reached, the tariff rates constitute a kind of public favoritism.

It is clearly contrary to a sound economic policy to foster an industry by means of protective duties, if the industry cannot exist independently after the initial stages have been passed. We may conclude that the infant industry argument is economically defensible, when the duty serves as an aid in effecting a more economical use of the factors of production. The danger in accepting this argument as the basis for a national policy is the difficulty in determining when the duty is no longer needed, and in marshaling sufficient political strength to reduce the rates in face of the opposition of those who have enjoyed the benefits of the protection.

Home Market Argument.—A second argument, which also occupied a large place in the debates over the Tariff Act of 1816, was that the tariff creates a *home market* by checking importation of foreign goods. In the early years of our country, agriculture, except for a few other extractive industries operated on a minor scale, was the principal industry. It was argued that if a tariff were levied, it would stimulate manufacturing and thus create a market for the home-grown crops. This argument in 1816 appealed to the South for it was not yet clear that England would be able to absorb the cotton grown in that section of the country. It was held that by checking the importation of English goods manufacturing would be developed here and that the domestic demand for cotton would be more regular and dependable than the foreign demand. This argument convinced the political leaders of the South and they supported the Tariff Act of 1816.

If we assume the existence of an urgent demand, there can be little doubt that a market can be created by a tariff, but the created market is not an additional market, but a substitute market. To see the effect of this one influence, it is necessary to assume that the supply of labor and capital in a country is constant in volume. With this assumption in mind, suppose a prohibitive tariff is levied on the importation of some good, say, cotton cloth, and suppose a demand for this cloth that is so urgent as to enable the goods to be produced and sold locally at a profit. If the price the buyers are willing to pay is high enough, the returns will attract private enterprise and investment, and, hence, the goods will be produced. But to produce the goods locally, capital and labor must be withdrawn from other industries. For purpose of simplicity, let us assume that prior to the

levying of the duty, this capital and labor were employed in agriculture. Those who are withdrawn from agriculture to become employed in the manufacture of the cotton cloth have to be fed, and their demand creates the local market for agricultural products which did not previously exist.

This fact is obvious, but what is not so clearly seen is that a market has been lost. Prior to the levying of the tariff, the cotton cloth was imported from a foreign country and paid for by means of exports, presumably of agricultural products. We have already seen that imports are generally paid for by means of exports, which is as true of one country as of another. The stopping of the importation of foreign-produced goods will make it impossible for that country to pay for the goods that it has been accustomed to buy here. In this case, it will be unable to buy as large an amount of agricultural products as before. Hence, the created local demand for agricultural products is not a net addition, but is merely a shift in demand from the foreigners engaged in manufacturing the cotton cloth to those now employed in making it within this country.

The question of importance is not whether a market has been created, but whether the labor and capital employed in one way yields more wealth than when it is employed in some other way. In this case, the withdrawal of labor and capital from agriculture will reduce the supply of agricultural products. The real question of significance, concretely put, is whether the people of a country have more cotton cloth and agricultural products before a duty is levied, when they specialize in agriculture and trade their surplus supply with a foreign producer of cotton cloth, than after the duty is levied, when both types of goods are produced within the country. If the total supply of wealth (which means, in this case, both agricultural products and cotton cloth) is greater after the change has taken place, then the people of the country have benefited by the levying of the duty. It should be recognized, however, that the issue turns on the relative productivity of different methods of using the labor and capital of the country and not on whether a market has been created as a result of the tariff duty. This issue should not be confused with the question of a profitable return to private enterprise for, as has been indicated above, if the demand is urgent enough, buyers will pay a price that will yield a return sufficient to attract private investment. As an economic ques-

tion, the issue rests upon the differences in productivity in the use of the productive energies of the nation.

Diversified Industry.—Closely associated with the home-market idea was the *diversified industry argument*. In fact, the two were inseparable, except that the latter had certain angles that differentiated it from the former. By levying a tariff, it was argued, the industries of the country would become diversified. Instead of relying on agriculture and commerce, manufacturing would be developed and the country would be greatly strengthened by such diversification. It would no longer be dependent on foreign countries, especially England, in times of war. This argument was particularly appealing in 1816, when the inconveniences and hardships experienced at the opening of the War of 1812 were still fresh in mind. Two things should be said concerning this argument. First, diversification of industry is not an end, and its desirability depends upon the same principle as was stated above. If the capital, labor, and natural resources are more productively used in one direction than in another, the country will benefit economically by diversification. The second point is that the motive to diversify industry does not depend solely upon the levying of a tariff, but is found in the advantages gained from territorial and occupational division of labor. If a tariff is needed permanently to insure diversification, the presumption is that the productive energies could be used more effectively in some other way. As a means of developing national independence, especially for war purposes, this argument has some merit, but even that phase can be pushed much beyond its social significance.

Labor and Wage Arguments.—Labor arguments for protection have been among the most persistent of those used by protectionists. By 1840, the political advocates had formulated an argument aimed to show to the workingmen of the country, who were just coming to be recognized as an important political factor, what the tariff meant to them. This argument first took the form that the tariff was necessary as a means of protecting the wages of American workmen against the pauper or low-paid labor of Europe. It was held that the American manufacturers could not meet foreign competition because of the low wages paid in England and Europe, and the only way to protect the American wage earner was to levy a tariff on imported goods, which would enable the manufacturers to pay the pre-

vailing wages and also to meet the European competition. Later this argument took the form of asserting that protection was the cause of the high wages paid.

Three things should be said concerning the labor argument. First, it has never been shown that the levying of a tariff on behalf of an industry has been followed by a voluntary increase of wages on the part of the employers who have benefited by the tariff. Second, there have been no reliable figures presented to prove that wages in protected industries are higher than the wages of similar labor in unprotected industries. Third, as will appear in a later chapter, wages depend fundamentally on the productivity of labor, and unless it can be shown that a protective tariff results in enhancing the productivity of labor, this argument cannot be accepted. Unless the tariff actually results in a more effective use of labor and the other factors of production, the claim that the tariff is the cause of high wages cannot be upheld. It should be stated also that the *pauper labor argument* is not wholly acceptable, for the ability to compete with foreign countries does not depend solely on the wages paid but, rather, on the costs per unit of product. It has been repeatedly demonstrated that high wage rates may exist alongside of low unit costs. The relative efficiency of labor and capital instruments, the richness of the natural resources on which labor and capital are used, all contribute to low unit costs.

Of course the money wages paid cannot be disregarded as an important element in determining unit costs, but the proper approach to the principle involved is through the productivity of the labor employed. If labor is more productive in one country than in another, wages may be correspondingly higher without destroying the ability of the manufacturers to compete on an equal basis. The proof of this conclusion is found in the volume of our exports. If the low-wage country always had the advantage in competition, it would be difficult for the United States to find a market for its products, yet our exports about balance our imports, which means that, notwithstanding the high money wages in exporting industries, the unit costs enable American producers to sell in foreign markets. However, when an industry has been created by means of a tariff, the wages it is able to pay are dependent, to a large extent, upon the protection received. This question should be considered also in the light of the principle of comparative advantage and the fundamental fact should

not be forgotten that trading will be mutually advantageous where countries enjoy a comparative advantage in producing any product.

Equalizing Costs.—One of the most recent arguments is that the tariff should equalize the costs of production at home and abroad. The extreme to which this argument leads one, can be seen from the following extract from a speech of former Senator Aldrich of Rhode Island. In debating the Tariff Act of 1909 he said:

Assuming that the price fixed by the report is the correct one, if it costs 10 cents to produce a razor in Germany and 20 cents in the United States, it will require 100 per cent duty to equalize the conditions in the two countries . . . And so far as I am concerned, I shall have no hesitancy in voting for a duty which will equalize the conditions. If it were necessary to equalize the conditions . . . I would vote for 300 per cent as cheerfully as for 50 per cent.

If a country should adopt this reasoning as a guide for its industrial policy, there is no limit to which it could be carried. This argument would enable one to justify a tariff that would make profitable the growing of coffee in Illinois and lemons in Maine. If such a policy were strictly applied, it would destroy all of the advantages from territorial division of labor and, therefore, banish all international trade.

Nevertheless, this argument has been widely accepted in political circles and has been championed as a scientific basis for the tariff, but the determination of the costs that will be taken as the basis for equalizing the conditions of production is by no means an easy task. As every informed person knows, costs vary widely from establishment to establishment within a country. What costs, then, should be taken as the basis for equalizing conditions, the high-cost firm, the low-cost, or an average of costs? If the high-cost firm is chosen, the amount of protection given those with low costs is excessive; if the low-cost firm is taken, then those with higher costs receive no protection. This phase of the tariff discussion has not added much to the economic justification of protective duties. About all that can be said for this argument is that it does constitute a rough measure for maximum duties. The duties should never be higher than is sufficient to offset the differences in costs at home and abroad.

Political Argument for Protection.—The above discussion by no means exhausts the arguments for and against the protective tariff, but gives the student a point of departure from which all arguments should be analyzed. As stated above, most of the arguments in favor of protection rest on political rather than economic considerations. So long as each nation must maintain its own existence, there are undoubted limits beyond which international trade contributes to national strength. At this point, political considerations should be controlling and the unrestricted use of labor and capital should give way to conscious direction of these factors. England, which has been regarded as the champion of free trade, has followed a policy which, though not protective in the strict sense of this issue, has nevertheless a protective element in it. Situated as England is, freedom of trade has been an essential condition for the development of her economic welfare. In order that free trade be maintained, England has striven for naval supremacy. A strong navy has been necessary, not solely to protect her political power and her dominions, but to keep a regular flow of wealth into the island itself. The English have paid few protective duties, in the ordinary sense, but they have taxed themselves heavily to maintain a strong navy which has been a powerful influence in keeping trade free. There is some similarity in the taxes paid by the English people for the support of their navy and those paid by other nations for the development of industries essential to national defense, such as the encouragement of aircraft, of chemicals used in warfare, of explosives, and of other industries similar in character. Tariffs for these purposes are based upon political rather than economic considerations but, as was stated above, these arguments can be pushed much beyond their social significance.

Difficulties in Removing a Tariff.—Another important consideration in the tariff discussion is the fact that industries become adjusted to protective duties. Labor and capital tend to become specialized and, when tariffs have fostered the development of an industry, the sudden removal of these duties would work a hardship that would be felt beyond the protected industries themselves. This fact explains why it is extremely difficult to reduce or remove a tariff once it has been levied. The managers of the favored industries can show very definitely what results will follow the removal of the protective duties. The injury to these

industries is positive and immediate; private investments may be jeopardized, a condition that makes a strong personal appeal to the legislators. The impersonal appeal of the effect on the volume of wealth produced, if the labor and capital were readjusted, has no such impelling force. Men can see and understand private business losses, but the readjustment cannot be made without some such loss, even though subsequently the total volume of wealth may be greater. This fact explains why tariffs persist, even though at the time of enactment it is definitely provided that the protective duties are to be of limited duration. In addition, sudden reductions in the tariff are harmful, unless the amounts are small, because of the effect on the protected industries. Depression in these industries might be on a scale large enough to affect the general industrial conditions. Once a country has adopted protection as a policy, it not only finds it extremely difficult to change that policy, but in making a change it must recognize that the favored industries should be given time to make their readjustments. It would follow from this that, if a country decides to lower its tariff, wisdom suggests that the rates be reduced slowly and that the protected industries be given an opportunity to adjust themselves to the new conditions.

Incidence of the Tariff.—A much debated question has been, Who pays the tariff? This issue has been pretty thoroughly analyzed and the consensus of opinion is that, in most cases the tariff is borne by the domestic consumers. The proof of this conclusion is to be found in the effect that the tariff has upon the price of a protected article. In the absence of an import duty, the market price will tend to coincide with the costs involved in producing the marginal supply of the good. If the foreign producer can manufacture the good at lower unit costs than can the domestic producers, the foreign cost will tend to control the market price, providing the foreign producer can supply the full amount demanded. The levying of the duty acts as an additional item in costs of furnishing this supply. When the domestic demand is active, the domestic producer can charge a price that equals the full costs to his foreign competitor, which consist of his manufacturing and transportation costs, plus the tariff rate; hence, domestic prices of protected articles are likely to be equal to the foreign price, including transportation charges, plus the tariff. We may conclude that, generally speaking, the consumer bears the burden of protective duties by

paying higher prices for his goods than would prevail if trade were unrestricted by import duties.

There are some qualifications of this principle that may be stated as follows: Suppose the customary price steps in merchandising a good are in 25-cent units, as \$1.25 for one grade, \$1.50 for the next higher grade, and so on. If a 40 per cent tariff were levied on an article that ordinarily sold for \$1.50, it would cause the following price problem: The tariff would add 60 cents to the selling price and the merchant would have the question of determining the effect on his sales of offering the article at \$2.10, or at \$2. If the customary 25-cent steps were adhered to vigorously, he might find it difficult to overcome custom, and decide to quote the new price at \$2. In such a case, the middleman absorbs a part of the tax.¹ But if the price of the article were ordinarily \$1, the tariff would add 40 cents and it would probably be quoted at \$1.50, in which case the consumer would pay more than the tax. It should be observed that in this case the Treasury department does not receive the full amount of the additional price paid by the consumer, as a portion of this sum would be added to the profits of the merchant. There are a great many different ways in which the incidence of a tariff may be distributed, but the reader in his thinking on this question will do well to proceed from the effect of the tax on price, making such qualifications of the general principle as the individual conditions demand. Generally speaking, such a tax will enhance the price and the burden will fall on the consumers of the taxed goods.

There is one possibility that the foreign producer may absorb the tariff. Suppose he is producing an article under conditions of decreasing unit costs, and that his domestic market will not absorb all of his product at a price that will yield him a profit. Under such circumstances, he is confronted by the alternative of lessening his total output and selling it solely in the home market, thereby increasing his unit costs; or, of enlarging his production and selling a portion of his output abroad for less than the domestic price. If he can maintain the domestic price by limiting the amount sold in the domestic market, he can afford to sell the surplus abroad at less than the domestic price and gain by

¹ The figures assumed are not the significant part of this argument. They are intended to illustrate the fact that in practice the burden of the tariff may be split. Cf. EMERY, H. C., *Am. Econ. Rev.*, Vol. V, pp. 534-553, for an analysis of this question.

the transaction. If the economies from large-scale production are sufficient, the differences in unit costs from producing a large volume, as compared with a smaller one, will more than offset the difference in the domestic and foreign prices. The foreign producer, confronted by conditions like those here described, might accept a price for the surplus sold abroad that would absorb the tariff. The practice of selling goods abroad at a lower price than they command at home is known as "dumping." While, theoretically, it is possible for the incidence of a tariff duty to be absorbed as described, it should be regarded as a minor qualification of the general principle that domestic consumers pay the tariff.

By-products of Protective Duties.—There are several by-products of the tariff that should now be considered. Tariff legislation in the United States has always given rise to a large amount of "log rolling." Each representative or senator has constituents who desire recognition for some article. The tariff schedule becomes exceedingly long and complicated, and the ease of showing political favoritism is palpable. The student of this question will find it instructive to examine the last tariff act and, after noting the long list of protected articles, ask himself whether he thinks his own representative could have voted intelligently on the schedule of rates and commodities contained therein. What happens is, that with comparative ease a large amount of protection can be slipped into such schedules without attracting public attention. The ease with which this can be done fosters "log rolling" among the legislators in satisfying the demands of their local constituents.

Another by-product is irregularity of income. Custom duties are easy to levy but, unless they are placed upon goods of general use, they yield a fluctuating income. The Treasury of the United States sometimes has had enormous surpluses, due mainly to the revenues from the protective duties, while at other times, when importation was not so large, revenues have run low. Either situation leads to an unsound fiscal policy. When large surpluses exist they foster wasteful expenditures. During the 'eighties, there were many appropriations made for rivers and harbors, and for other similar purposes which were not a real contribution to public welfare, but were made because large sums of money were piling up in the United States Treasury and were thus an influence in causing financial stringencies. Instead of lowering the tariff rates, thereby reducing the income, the legislators

enlarged the expenditures. When these expenditures are extended over a number of years, they come to be an embarrassment from a fiscal point of view. A sound fiscal policy demands regular income and a balanced budget. The history of the tariff in this country shows that protective duties have not conformed to this standard.

When the effects of restrictive legislation and protective duties are carefully analyzed, the presumption in favor of freedom of trade is strengthened. The basis of this strength lies in the gains from a division of labor. The arguments that have been presented, showing the advantages of specialization, hold true in territorial division of labor between countries. This is the fundamental principle from which the student should shape his reasoning on the tariff issue and, when practical or political considerations demand a qualification, the point of departure for this change should be the principle developed in connection with territorial specialization.

Suggested References

- CARVER, THOMAS N., "Principles of National Economy," Chaps. XXXI and XXXII.
- EDIE, LIONEL D., "Economics, Principles and Problems," Chap. XXXIII.
- ELY, RICHARD T., "Outlines of Economics," 4th Ed., Chap. XIX.
- FAIRCHILD, FURNISS, and BUCK, "Elementary Economics," Vol. II, Chap. XLVIII.
- FAWCETT, HENRY, "Free Trade and Protection."
- GIDE, CHARLES, "Principles of Political Economy," Rev., pp. 259-272.
- HAYES, H. GORDON, "Our Economic System," Vol. II, Chaps. XXXIV and XXXV.
- HIGGINSON, J. L., "Tariffs at Work."
- PAGE, THOMAS W., "Making the Tariff in the United States."
- RUFENER, L. A., "Principles of Economics," Chap. XXXV.
- SEAGER, HENRY R., "Principles of Economics," 3rd Ed. Chap. XXII.
- SELIGMAN, EDWIN R. A., "Principles of Economics," pp. 597-612.
- STANWOOD, EDWARD, "American Tariff Controversies in the Nineteenth Century," Vols. I-II.
- TAUSSIG, FRANK W., "Principles of Economics," Vol. II, Chap. XLVII.
- , "Some Aspects of the Tariff Question."
- , "Free Trade, Tariff, and Reciprocity."
- VINEY, JACOB, "Dumping, a Problem in International Trade."

CHAPTER XVII

GENERAL THEORY OF DISTRIBUTION

In the preceding chapters, we have treated four main groups of topics, namely, the economic structure of modern society and its chief characteristics; production, its nature and organization; division of labor, the market and exchange; and finally, the mechanism of exchange—money, banking, and credit, with their attendant problems of international trade and exchange. We now come to the next main division of the subject which is commonly treated under the title of *distribution*. Heretofore, we have been concerned with the production of wealth and the principles that tend to direct the operation of industry. Now the task will be to find how the wealth produced is divided and to examine the principles that govern its disposal.

Even at the expense of some repetition, a restatement of what has gone before will help to make clear the problem of dividing the wealth produced. It will be remembered that production has been made to include all human activity that results in some means of gratifying human desires. Conceived in this way, production may be considered as a great social process, which consists of the cooperation of a vast number of individual, producing units that are engaged in rendering personal services and in making the stock of wealth which we find on the shelves of the merchants when our desires impel us to buy; or it may be considered as a private and acquisitive matter as, for example, the ownership of a store like that of Marshall Field & Company, or the private business of any individual or group of persons. Economically speaking, those portions of the human family who are engaged in producing and exchanging goods among themselves, whether this production and exchange takes place within national boundaries or between nations, constitute a complex, social structure into which the individual endeavors to fit himself and from which he strives to secure his living. Viewed in its broad, general aspects, then, production is a great social, cooperative process for the purpose of furnishing the human family with

the means of gratifying the desires of its individual members. But considered from the point of view of the owner of a producing unit, production means a private business from which he expects a surplus from the sale of the goods he produces. From either point of view, production is not an end in itself, but is carried on for the purpose of securing the means of gratifying the desires of mankind. Therefore, the productive processes are not complete until goods are in the hands of the consumer.

Distribution Defined.—The methods and principles of sharing the products of industry are the province of distribution. Like many other terms used in economics, distribution has a specialized meaning with which the reader should become familiar at once. As here used, distribution is that division of economics which treats of the sharing of the wealth that results from the various lines of production. This meaning should be distinguished from two other common usages of the term. Sometimes distribution is used in connection with the ownership of wealth, as when reference is made to the amount of wealth which an individual possesses. Some may own much and others little. The emphasis in this usage is on the holdings which individuals possess. Many of the proposals for social reform refer to the distribution of wealth in this sense and are likely to stress the inequality in the ownership of wealth. While this usage is significant, especially as a matter of general social policy, and while the whole institution of property has a very important bearing upon the operation of our economic system and greatly affects the amount of wealth which individuals enjoy, yet this is not the sense that will be stressed at this point. We will be concerned here with the broad, general principles that tend to govern the process of sharing the wealth produced. Distribution, then, as a division of economics, is concerned with the formulation and application of these principles.

Another use of the term with which one should be familiar is that employed in marketing and transportation. Here, the term refers primarily to the location or situs of the goods. A manufacturer may sell his product over a wide market area, in which case it is customary to say that a product is having a wide distribution. The emphasis is upon the process of disposing of the goods, or on their exact location throughout a marketing area. While each of these usages has a legitimate place, we are concerned at this point with the processes and

principles that affect the sharing of the products that result from the combination of the various agents or factors of production in modern industry.

It will be recalled that production involves the combination of land, labor, and capital. This process requires an interval of time, during which there is a constant flow of ripening wealth. At any given instant of time, goods will be found in all stages of development from raw materials to finished, consumers' goods available for immediate consumption. This volume of consumers' goods is commonly referred to as the *social income*, or sometimes as the *national dividend*. It will be evident that these terms refer to the total stock of goods available for the satisfaction of human wants. The economic welfare of mankind is dependent upon the size of this stock. If the social income is large, mankind will be economically prosperous and have a high standard of living; but if it be small, mankind will be poor and have a low standard of living. It is from this stock available in the market, that we as individuals get our *real* income as distinguished from *money* income. As will be shown presently, the money income of individuals may arise from several sources, but in the satisfaction of desires, all persons alike must exchange their money incomes in the market for goods, and these goods make up the social income that results from production. It will now be our task to examine more specifically how goods are shared.

The sharing of the products or income from industry may be studied from two points of view. First, there is what is known as *functional distribution*, or the assigning of shares to the factors of production on the basis of a function performed. Land, labor, and capital, as the principal factors of production, are assigned shares of the income in accordance with the function which each performs. The shares assigned to these factors are known as *rent*, *wages*, and *interest*. In addition, the share known as *profits* is usually included with this group, although, as will be shown later, it differs from the other three shares in several important respects. In other words, functional distribution is concerned with the allocation of social income in accordance with the several different functions that have been performed in bringing that income into existence.

In the second place, we have *personal distribution*, or the determination of the shares of individual persons. A little study will reveal that while these two phases of distribution are closely

related, they are not identical. For instance, to explain why city lots command high rents is one problem, but to explain why large amounts of rent fall to one man, the owner, is another. In the former case, we are concerned with an explanation of why a definite amount of the social income is received in the form of rent, while in the latter we are dealing with the source of personal income of an individual. We will soon find that personal income may arise from a variety of sources. One may inherit wealth and be able to enjoy it without performing any social functions, which is made possible by the institution of private property. Then, one may own land and receive all of his income from the services which land renders. Or, he may perform a labor service and be compensated for the benefit he has created. In the case of labor, functional and personal distribution approximate each other.¹ We may illustrate the possibility of different sources of income by reference to a farmer, who may receive part of his income in the form of rent from the ownership of land, a part as interest from the ownership of the capital with which he tills the soil, and a part as wages which he receives as a payment for the services that he performs on the farm. From an individual point of view, personal distribution is the matter of chief concern but, since the income of an individual may arise from more than one source, it is important to understand what forces are at work to determine these amounts. This is the chief problem in the consideration of functional distribution, hence, this topic will be treated first.

Evaluation and Distribution.—One of the first points to be noted in the study of this topic is the fact that the determination of either a functional or a personal share in our present economic system is a phase of the general problem of value and evaluation. Under communism or socialism, the distribution of the shares of income would likely be determined by entirely different principles from those which prevail in the present individual exchange economy. Under either communism or socialism, personal income might be quite arbitrarily determined by standards previously chosen by the properly constituted authorities. But in the economic order in which we live, both functional and personal shares are the result of the operation of market forces. A few simple illustrations will help to make clear how personal income is so determined, though it will require more detailed

¹ Ely, R. T., "Outlines," 4th Ed., pp. 382-401.

analysis to show how the same forces are at work in fixing the functional shares.

It is a matter of common observation that men in all kinds of business are engaged in producing some commodity or in rendering some form of service, which is sold in the market. For instance, we may think of a truck gardener who grows a variety of vegetables and brings these to the market and exchanges them for money. The amount of money that he gets for his products will depend on how much importance is attached to the individual units of his goods and upon the volume he has for sale. If his goods are in great demand, he will receive a relatively large money income. The amount of his money income will determine how many goods he is able to purchase for his own personal use. In each case, the process of evaluation and the phenomenon of value appear. It is immaterial whether one is disposing of a tangible commodity or an intangible service, the amount of wealth he is able to command for his own purpose depends upon this evaluation process. The same principle holds true in the determination of the functional shares, but to make this matter clear will require further elaboration which will follow below. At this point it is important to realize that in dealing with the subject of distribution, we are still concerned with the principles of value previously set forth.

In applying these principles to the determination of the functional shares, it must be recalled that, in the explanation of value, perfect competition and perfect mobility of the factors of production were assumed as the starting point of our analysis. This assumption must be repeated at this point. Under these conditions each factor of production will be assigned a share of the social income in accordance with its contribution to that product. If, for instance, land is relatively scarce, its services will carry a high valuation and, hence, will exchange for a larger relative share of the total social income than the services of capital or labor. The same conclusion will likewise be true if the services of either of the other factors are relatively scarce. Whenever either capital or labor is scarce, either will command a high price for its services. The truth of this principle can be illustrated concretely in a great variety of ways. First, let us look at the case of land and take for our purposes land situated in the highly congested business districts of a city. The value of such land is often extremely high in price because of its limited amount,

which makes the services rendered by that area extremely valuable. The same situation obtains with all land in a country that is gradually increasing in population, for the services of such land become increasingly important and, hence, its value tends to rise.

When capital is scarce its services are also greatly appreciated and people will pay a high price in order to command it. The payments for the use of capital are always higher in those countries where it is scarce than in those where it is relatively plentiful. The principle applies in like manner to labor. If there are many persons who can perform a given kind of service, the value of this service will be relatively low but, on the contrary, if only a few persons can render the service, then it will be highly esteemed and command a high price. In 1919, just following the World War with business unusually active and the supply of immigration definitely limited, the rates of wages offered for unskilled labor were high. Attempts to lower money wages for this type of labor were unsuccessful because the services were relatively scarce and, hence, valuable. Whenever any one of the factors of production is performing a valuable service, it will command a relatively high price and will, therefore, exchange for a large volume of products. This statement should make it clear that distribution is a phase of the value problem. In other words, to explain the prices paid for labor services is to explain wages, and to explain the prices paid for the services of land and capital is to explain rent and interest.

The important consideration is to discover how these market forces operate to determine the various shares. It is evident from what has just been stated that the share of any factor of production is dependent upon its relative significance in the productive process. This significance is not dependent solely upon the technical efficiency of the factor but upon the social significance of the product or service rendered by that factor. For instance, we may think of a very efficient machine or a very highly skilled man. In either case, the degree of technical efficiency will not insure that the product will have a high social significance unless the machine or the man makes a product that people want.

Diminishing Productivity and Distribution.—The recognition of this fact will bring to mind the principle of diminishing productivity and the law of variable proportions that was dis-

cussed in the chapter on Production.¹ The first of these principles made it clear that by increasing any one of the factors of production more rapidly than the others would probably be followed by an absolute increase in the product, but this increase would be at a diminishing rate. When labor and capital were applied upon a definite area of land, after a certain combination had been reached, the absolute amount might increase but the rate of increase would not be in the same ratio as the increase of labor and capital. The law of variable proportions took account of the fact that the same result might be obtained by various combinations of the factors of production and stressed the entrepreneur's problem as to how to secure a given result with a minimum of expense.

The entrepreneur has to buy the services of these factors, and the price which he pays for them appears as an expense to him. The fact that he can get the same volume of product by different combinations will, generally speaking, lead him to seek the least cost combination. This principle of least-cost combination holds, except for a minor qualification that has been previously explained. The important consideration is that these production principles have a significance here as an aid in determining the functional shares, and also a bearing on the methods of measuring the function performed by the separate factors.

In applying the principle of diminishing productivity to the problem at hand, we may think of a fixed area of land being worked by a varying number of laborers, with a definite amount of capital instruments. The laborers, we will assume, are equally efficient. As their number increases, the social significance of the service of each laborer tends to decline. Under the conditions assumed and at a given instant of time, the significance of the service of any individual laborer is as great as that of any other worker employed, and is measured by the significance of the last, or marginal, man. Since the laborers are equally efficient, it is immaterial which worker is regarded as marginal. By reducing the number of workmen the significance of the service of the marginal laborer and, hence, the significance of each of the remaining laborers, is enhanced. We conclude from this that the functional contribution of labor in this illustration is determined at the margin of use, and is equal to the contribution of the marginal man times the number of men

¹ Chapter VI, pp. 124-140.

employed. The assumptions that have been made as the basis for this conclusion should be carefully observed. The only variable here considered was the number of men employed. If there were variations in the productive capacities of the different workmen, the general principle would have to be qualified to take account of these differences. But all such qualification would be made from the general conclusion that the significance of the labor services of a group of workmen of like capacities tends to equal the significance of the service of the marginal man employed.

This principle has a very important bearing upon the explanation of the shares which are assigned to the factors of production. In accordance with this principle each factor of production is assigned a claim upon the income from industry that is equal to the significance or the value of the service of the marginal unit of that factor employed in industry. In other words, in so far as competition is operating and there is mobility of the factors, each factor will get what its services are worth in production. To make the case concrete, we may ask, What is the maximum amount that can go to a laborer in the form of wages in any industrial enterprise? Take the laborer off the job and measure how much the product is reduced. Put him back at work and see how much the product is increased. The value of this added product would be the maximum claim that could be made not only by this laborer, but by all others who are rendering the same kind of service.

The Marginal Productivity Theory.—This method of determining the economic shares of wealth is known as the *marginal productivity theory of distribution*, and is applicable to all of the factors of production. The principle may be formulated as follows: Where competition and mobility of the factors of production are perfect, each factor will tend to get a share, the value of which is equal to the value of the services of that factor as determined by its marginal productivity.

Two implications of this principle should receive attention. First, the marginal productivity of any factor is affected by its relative supply. By reducing the supply of any factor relatively, its marginal productivity will be enhanced or, in other words, greater significance will be attached to the services of the last unit employed of that factor. Under these circumstances, we would assign or impute to the factor a larger claim on the income from industry. The second implication is that the value of the

services of any factor is derived from the value of the products that result from its use. For instance, the services of a laborer would have no value unless they were employed in producing some good which people want and are willing to buy. In this way, the value of a good in the market sets the limits to the value that can be imputed or assigned to the factors which are required to bring the good into existence.

Economic versus Contractual Shares.—The conclusions just stated were predicated upon the assumption of perfect competition and mobility of the factors of production. The two qualifications must be repeated here that were emphasized when we were discussing value, *viz.*, the assumed conditions do not coincide perfectly with the actual conditions. First, there is always a large amount of economic friction which prevents the perfect operation of the market forces; and, second, there is a large amount of conscious interference, or monopoly control, over these forces. The actual shares may, therefore, depart from the theoretical ones by the extent to which the actual market forces deviate from those assumed, whether this deviation is due to economic friction or to conscious interference. In our economic order, personal distribution or the share of individuals is the result of a contract or a bargain. The landowner bargains with a tenant for the use of land and the amount he gets is the result of the terms agreed to in the contract entered into by the two parties. The same is true between the employer and the employee, and the borrower and the lender. In each case, the actual amount received is the result of a contract or a bargain. This fact necessitates a distinction between the economic or theoretical share and the contractual or actual share. The economic share is the amount imputed to the factor and is determined by the marginal productivity of that factor in whatever use to which it is devoted. The contractual or actual share is the amount fixed by bargaining between the parties to the contract. The contractual share may be more, but it is likely to be less, than the economic share by an amount that depends upon the general conditions of bargaining. The more nearly the conditions conform to those originally assumed, namely, competition and mobility, the more closely will the contractual and economic shares approximate each other.

Distribution as Affected by Legal Institutions.—The bargaining referred to in the previous section is for the purpose of

securing control over one of the economic or imputed shares. When the employer hires a laborer, the object of the employer is to get control over the value created by the services of the laborer, or in other words, his product. The laborer is likewise endeavoring to secure control over the value of his product, but his services are worth little unless combined with capital and land, which he does not control. If he gets the value of his labor, he must secure it through his contract with the employer. The same situation as that just described exists between the lender and the borrower, and the landowner and the tenant. In all of these instances, one or more of these functions may be combined in one person, as when a farmer, owning his own land, buys the tools he uses from his own savings and performs all the work on the farm. In such an instance, there would be no contract over the shares and all bargaining would be between the farmer as a producer and the consumers of his product. But even here the share of wealth which this farmer enjoys depends upon the contracts which he makes with the users of his products.

From the above discussion, it will be evident that all actual shares are received as the result of bargaining. The process of bargaining is protected by the legal rights of property and contract. The right of two or more persons to enter voluntarily into an agreement is, as we have seen, one of the fundamental characteristics of our economic order. When such an agreement has been made, it establishes definite property rights. Suppose, for instance, a man wishes to secure a lot on which to build a home. He finds a realtor who has a lot for sale. After negotiations, an agreement is reached and a contract is signed. The realtor has a property right in the sum of money which the man has agreed to pay for the lot, while the man has a property right in the lot. It is evident that the transaction has consisted in an exchange of property rights in wealth. Distribution or the sharing of wealth is, therefore, to a very large extent, a question of ownership or property rights in wealth. The wealth of an individual consists in the property rights to valuable things or services. He may wish to enjoy this wealth personally, or to transfer his rights to others for some other form of wealth which he prefers. Business may be regarded as the process of exchanging rights to valuable things or services.

Every extension of property rights by society extends the field of private gain and thus affects the distribution of wealth.

Not only is this true, but the theory of property also greatly affects the distribution. Property rights in land were originally held to apply not only to the surface area, but they extended from the center of the earth to the dome of the sky. As a result the owner of the surface area owns any valuable deposits lying beneath the surface, such as minerals, oils, coal, etc., and in addition he has property rights to the space above this area. The question of air rights is becoming increasingly important in these days of skyscrapers and airplane transportation.¹ These changes are enlarging the scope of property and consequently have an important bearing upon the distribution of wealth.

Influence of Property and Contract.—It is evident that personal distribution is very greatly affected by this theory of property rights. The discovery of oil or gas underlying a given area has often made men wealthy over night, because the theory of property rights gave the owners of the surface control over the valuable elements that lie beneath it. Society may by legislative enactment restrict property rights and thus change the distribution of wealth. Take the legislation in the field of public utilities and the work of administrative commissions, such as the Interstate Commerce Commission, or the public service commissions of the various states. The actions of these bodies greatly affect the actual distribution of wealth through the effect of their decisions in changing the property rights involved.

The theory of rights which bulk so large in our legal relations was inherited from the eighteenth century. It was during this century that the idea developed and gained general acceptance that man was born into the world endowed with a stock of inalienable rights, and that the state existed for the purpose of maintaining these rights. Most fundamental among these were that of property, including the right to the free disposal of property, and the free pursuit by the individual of his own economic self-interest. A property right was regarded as an inalienable

¹ A very large building, to be used as an office building, is to be erected soon in Chicago over the Illinois Central tracks. The prevailing theory of property rights gives to the Illinois Central Railroad Company control over the air rights above its tracks and any use of this space by others must be secured as the result of a contract with this company. Another interesting aspect of this general theory of property rights is being brought vigorously to public attention by the rapid development of aviation. If the ancient theory of property were vigorously applied, every driver of an airplane would be guilty of trespass as he flies from city to city.

right of the individual, but it was assumed that the exercise of this right would result in a service to society. This service, however, was a by-product. Rights existed anterior to the performance of a service and were not deducible from the discharge of a function, but were inalienable privileges of the individual. These philosophical notions were written large into the laws of our own country and those of western Europe, and they still exist, though greatly modified at many points.¹

It will be evident that the acceptance of these notions of rights has greatly affected the distribution of wealth. If function were accepted as the sole basis for a claim upon the income from industry, rights to an income would be derivative from the performance of a service. The amount which an individual could claim would depend upon the service which he renders. The economic shares, described above, are based upon the performance of a function. In a free market, with perfect competition and perfect mobility of the factors of production, each factor would get a share of the income from industry equal to its contribution to that income. Among the influences that prevent the attainment of this method of distribution is the effect of legal rights, especially property rights, described above. To the extent that property rights establish a claim upon social income without a service having been performed, distribution on the basis of function is interfered with. As a matter of social policy it would be better to have rights to a share in the social income rest upon a service rendered, than to recognize rights as anterior to the performance of a function. Should this principle be accepted and our legal institutions become adjusted to it, then individuals would tend to be compensated more nearly in accordance with the imputed or economic share. The only cause for departure from this share would be the obstacles to the perfect operation of competition. These would be natural obstacles and not subject to social control.

The conclusion from this general treatment of distribution is that, except as affected by legal institutions, the share which any factor gets tends to equal the value of the services of the marginal unit of that factor. If the services of a unit of land in a given usage is worth \$100, the owner of that land will tend to get that sum for it as rent. In like manner, if the services of a laborer, say a farm hand, are worth at the margin \$60 per month, then

¹ TAWNEY, R. H., "The Acquisitive Society," especially, Chap. II.

farm hands will tend to get this sum for their labor. Likewise, if the services of a marginal unit of capital in a definite usage are worth \$10 per year, the lender will tend to get that sum for its use. In each of these cases the expression "tend to get" has been used as an indication that the actual share at any one time may depart from the theoretical share due to the fact that the bargaining takes place prior to the sale of the finished product so that, at best, the parties to the bargains can only estimate the value of the services employed in the production of the good. Past experience is the only guide for this estimate and this is by no means accurate, as no one can know definitely what change may come in the estimate which marginal users may place on the good during the interval in which the good is being completed and brought to the market. The principle then expresses the long-run tendency which holds so long as, and to the extent that, competition is an operating force in the use of the various factors of production and in the disposal of the goods produced. In other words, the marginal productivity theory of distribution expresses the long-run influences in sharing the income from industry.

There are some important implications attaching to the acceptance of this theory.¹ Any attempt to fix arbitrarily the share of any factor of production, without changing the demand for, or supply of, that factor, will most likely fail. If Congress attempted to fix the price of wheat at \$3 per bushel, without attempting to affect the demand for wheat, the project would fail because the marginal significance of wheat would not remain fixed at that point. Those farmers who could grow wheat for less than this sum would be stimulated to increased acreage. The added supply would depress the marginal significance of wheat unless additional uses of wheat could be found. The recognition of this principle has a very important bearing on the proposals so actively discussed at the present time as an aid to the farming population, and helps to explain why it is necessary to include an exporting agency as a part of the policy of maintaining a definite price for wheat in the United States, as is being proposed. An examination of the various arguments on the proposals for the relief of the farmers will reveal the fact that any policy which overlooks the fundamental principles of market valuation will surely set into operation other forces that tend to nullify arbitrary attempts to fix the shares of any group.

¹ TAYLOR, F. M., "Principles of Economics," pp. 447-455.

A second implication is that the share received by any individual tends to vary inversely as the numbers in his group. For instance, taking the farmers again, the most effective way of enhancing the price of farm products and thus affecting the share of the farmers is to reduce farm production. Very soon the prices of farm products would be increased, and likewise the income per farmer. The same principle holds for all classes of labor. Increase in numbers lessens the productivity of marginal units and thus depresses the individual shares. As a corollary of this conclusion it may be stated that the share of any group will tend to increase as the result of the increase in the size of other groups. If farm population increases, or if it becomes relatively more productive, then those working in other industries will enjoy a larger share of wealth in the form of a greater purchasing power in exchange for farm products. A similar statement can be made in respect to the effect of an increase in population which tends to lower all shares, and especially wages, except rent, which will tend to increase.

A third implication is that, in general, any improvement in the methods of production, or any discovery that lessens the cost of production will, in the first instance at least, tend to enhance interest and wages. Such opportunities enhance the marginal significance of saving and will also make the services of labor more valuable. Without doubt, the student will recognize from his own experience and observation many other instances of the operation of the principle of marginal productivity playing an important part in determining the share which individuals and groups of individuals get.

Suggested References

- BYE, RAYMOND T., "Principles of Economics," Chap. XVII.
CARVER, THOMAS N., "The Distribution of Wealth," Chap. III.
———, "Principles of National Economy," Chap. XXXIV.
CLARK, J. B., "The Distribution of Wealth," Chap. IV.
DAVENPORT, H. J., "Economics of Enterprise," Chaps. X, XXIII, and XXIV.
ELY, RICHARD T., "Outlines of Economics," 4th Ed., Chap. XX.
FAIRCHILD, FURNESS, and BUCK, "Elementary Economics," Vol. II, Chap. XLI.
MARSHALL, ALFRED, "Principles of Economics," Book VI, Chaps. I-II.
TAYLOR, FREDERICK M., "Principles of Economics," Chaps. XXXVI and XXXVII.

CHAPTER XVIII

RENT

In the previous chapter, the general principles of distribution were given consideration. We will now turn to a more detailed study of the application of these principles to the determination of the four principal shares: rent, wages, interest, and profits. We will direct our attention first to that of rent. As generally understood, the term "rent" means the price paid for the use of any form of wealth as, for instance, the rent paid for the use of a house, an apartment, an automobile, a typewriter, a dress suit, etc. In this commercial sense, the meaning of rent is almost synonymous with that of "hire," although there is at least one type of hire that is not ordinarily included under the term of rent, *viz.*, the case where a company owns a patented machine and lets its use out to other firms, not for a definite rental, but for a price per unit of product made by the machine. This form of hire is commonly known as "a royalty." In other instances the commercial usage of rent means the payment for the use of any form of wealth as distinguished from the payment for the full property rights in that wealth.

Rent Defined.—In economics, the term "rent" is generally used in a much more restricted sense, as a payment for the services of land. Before this definition can be understood, it is necessary to know how the economist uses the term "land." By land we mean not only the superficial area, but all the natural resources that aid man in securing a living. In this sense, land would include mineral deposits, forests, fisheries, water power, electrical power, etc., as well as the surface of the earth. On the basis of use, Ely and Morehouse have classified land into four groups:¹

- A. Subsurface appropriation—oil, gas, minerals, stones, salt, etc.
- B. Surface appropriation.
 - 1. Site purposes.
 - 2. Land for agricultural purposes.

¹ ELY and MOREHOUSE, "Elements of Land Economics," p. 30.

3. Land for transportation and communication.

4. Recreational land.

C. Water and land connected with it.

D. Supersurface appropriation for the aeroplane and radio.

Rent, as a payment for the use of land, was at first defined by Ricardo as a payment for "the original and indestructible powers of the soil."¹ While this conception of rent is not tenable in the form in which it was advanced by Ricardo, the essence of the doctrine still holds true. There are, perhaps, no qualities of the soil that are indestructible, but there are qualities of land, as defined, that are permanent and indestructible. For instance, land located in the temperate zone has certain climatic conditions that are permanent so far as human life is concerned. Whatever cosmic changes may go on to alter the character of this land through the geological ages of the future, these can have no bearing upon the productivity of land during the lifetime of man in any comprehensible period of time. Rainfall, temperature, and other natural forces of similar character are indestructible qualities of land and materially affect its productive power.

In addition, the action of natural forces in past ages that has resulted in the deposit of minerals, coal, oil, and other useful elements, is beyond the control of human beings. By this action, these useful things are differently distributed. In some places, the veins of coal are thick and pure; in others, they are thin and mixed with shale or other non-combustible elements. Iron ore deposits, as well as the deposits of other minerals, are richer in some localities than in others. These facts, so well known as the result of human experience, bear out the conclusion that there are some permanent qualities of land which greatly affect its productivity and, hence, its services. With this enlargement of the concept, we may say that rent is a payment for the services of land, meaning by land all of the forces of nature that aid man in production and are limited in supply relative to his wants.

Rent and Capital Expenditures.—The use of rent in this sense excludes the payment for the services of improvements made on land, such as houses, mills, factories, fences, etc. When a family rents a house for \$100 per month, a part of that payment, say, \$60, is made for the use of the house itself, and \$40 for the

¹ RICARDO, D., "Political Economy," Chap. II, Sec. 24.

land.¹ The latter is pure rent and is sometimes referred to as ground rent. The other payment is made for the services of capital and labor, and appears in the form of interest and wages. While this statement is true generally of improvements, there are some forms of improvements to land that once they are made thereafter become permanent characteristics of the land itself. For instance, in those regions of the country which were formerly covered by forests and have now been turned into farm lands, a large amount of labor and capital had to be expended in removing the stumps and roots of trees before this land was tillable. Or, take an instance of a marshy or swampy region that has been drained. All such improvements to the land, while being made, are of the nature of capital expenditures. They have a definite cost in wages and interest. But once a section has been cleared of timber, or a marsh drained, thereafter these improvements behave as permanent characteristics of the land. It is appropriate and customary to apply the term rent to the payment for the total services of the land, including those of the improvements of this character.

Three Uses of the Term "Rent."—Starting then with this definition of rent as a payment for the services of land, we must distinguish three uses of the term, namely, *economic rent*, *contractual rent*, and *commercial rent*. By economic rent we mean the full share of income that is imputable, or assignable, to land by virtue of its productivity. By contractual rent we mean the sum paid by the tenant to the landlord for the control of the economic rent, which may be less or more than the economic share, depending upon the bargain entered into by the landlord and the tenant. If the tenant pays more than the land produces, his capital and labor employed will receive less than similar capital and labor used in other pursuits and he will presumably demand a new adjustment of the contractual rent at his first opportunity. Likewise, if the landowner discovers that the tenant has paid him less than his land produces, he will demand a larger payment at the date for the renewal of their contract.

¹ The student should not accept these figures as an accurate measure of the difference between the return on land and that on improvements, as these relations will vary from community to community and from tract to tract. All that these figures are intended to illustrate is the fact that the total payment made should be divided in accordance with the importance of the separate services.

Contractual rent, therefore, tends in the long run to approximate the economic rent.

Commercial rent is the amount that is provided by a contract for the use of any form of wealth, and includes a return in excess of economic rent, as defined above. This excess return is a payment for the services of the other factors of production. The illustration concerning the "renting" of a house, given above, is an instance of commercial rent where the payment includes not only a price for the services of the land, but also for the improvements on the land, which in this case consist of the dwelling house. It is this form of rent with which the business man is most likely to be concerned. Some authorities have used rent in this sense, but there are considerations which are convincing to the author for confining the meaning of the term to the definition already given. As will be shown later, the behavior of the non-reproducible qualities of land is different from that of qualities which may be reproduced and have a definite cost of reproduction.

The Nature of the Services of Land.—Our next problem is to discover what constitutes the services of land for which rent is a payment. How is it possible to measure the contribution of land to the social income, or the value of its services? Is it not true that the whole social income should be attributed to land, for how could there be any production without the use of land? Looked at in this way, we could say that land is indispensable for every kind of production, but it is equally true that no production could take place without labor and capital and, hence, we could say with equal force that the total income should be assigned either to labor or to capital. But in all forms of production, we are not concerned with land in general, or capital and labor in general, but with definite and specific amounts of land, labor, and capital. Our problem is to explain the share assigned to a particular area of land, say an acre, or a lot containing a definite number of square feet.

In dealing with this subject there are two qualities of land that have an important bearing upon the problem, *viz.*, fertility or productiveness, and location. In agricultural land, we are concerned mainly with the fertility of the soil, while with urban land, we are concerned with location which shows itself in greater returns to the business done on a given site. In either case, there are differences in the yield to a given application of

labor and capital employed. In agriculture, fertility is not uniform and different acres, though cultivated in precisely the same way, will yield different amounts. While differences of this character occur in the use of both labor and capital, they are not as pronounced as in the case of land. There are probably no two machines nor two workmen whose output is precisely equal, but the variations in production of these factors are much less than in the case of land. We may illustrate this characteristic of land by assuming a number of separate acres cultivated in precisely the same manner. The best acre will yield, let us say, 40 bushels; the second, 35; the third, 30; the fourth, 20, and so on. These different returns are significant in determining the services attributable to land.

Location and Rent.—The differences just described are easy to observe in the case of agricultural land where they appear in the form of concrete products of unequal amounts. Similar differences are as positive, though not as obvious, in the case of urban land where the product is an intangible service. In urban land one business site with the investment of a definite amount of labor and capital will yield a certain volume of trade, while another site with the same investment yields less. In this sense we may say that one site or location is more productive than another. These differences in yield, whether due to differences in fertility or in location, are the important considerations in determining the services and, therefore, the rent of land.

Location has been discussed here in its application to urban land where the importance of site is most evident and pronounced. But even in agriculture the returns from land are affected by location. Acres that yield the same number of bushels when cultivated with a definite amount of labor and capital may have, because of differences in location, different transportation costs in the marketing of the products. A farm two miles from a shipping point will have a lower cost per unit of output than one equally fertile that is situated ten miles from transportation.

As a matter of fact, it is immaterial whether the differences between two pieces of land are due to fertility or to location, the fact of these differences will be reflected in the returns to the employment of capital and labor on the land. There will be a larger yield per unit of labor and capital in the one case than in the other and it is this difference of yield that constitutes rent. The difference in productiveness is imputed, or rather attributed,

to land and this surplus of production is thought of as the distinctive services of the land to which it is assigned.

Methods of Measuring the Services of Land.—Our next question is, How can we measure the services of land? In answering this question, two things must be kept in mind. First, demand and supply always tend to be equal. Goods are produced to sell and if a surplus of any kind develops, the rate of production is slowed down, or even stopped, until the consumers have taken the existing stock off the market. While there can never be a perfect balance between production and consumption (supply and demand), there is a constant pressure to equalize these two forces. As a result of this tendency to equalization, we can conclude that demand for its services will determine how much land will be cultivated and how intensively it will be utilized.

In this connection it will be well to note a peculiarity of land that affects the problem of determining and measuring its services, namely, the distinction between its physical and its economic supply.¹ While the physical supply is spatially limited, it far exceeds the amount that is now utilized by man. Land differs in this respect from capital instruments, the physical supply of which does not greatly exceed the instruments in actual use. Capital instruments wear out and are withdrawn from use when they are no longer capable of rendering service, so that their value tends to coincide with the costs of reproducing them. While some qualities of the land, such as fertility, show this same tendency, there are other qualities as location that do not ordinarily depreciate with use. Hence, the economic supply of land depends upon the value of its services and, as these services become more valuable, that supply can be greatly increased because the physical supply is so much greater than that in use.

There is a similarity in this respect between land and labor, as the number of persons capable of performing labor services is greater than the number actually at work. The physical supply of labor, therefore, exceeds the economic supply, but not as much as in the case of land. The economic supply of any of the factors depends upon the evaluation placed upon the services of the factor, but because the physical supply of land so greatly

¹ Cf. DORAU and HINMAN, "Urban Land Economics," p. 469, for further treatment of this topic.

exceeds the amount in use, we may say that the economic supply is more elastic than that of the other factors.¹ At all times, it is the demand for the services of land that determines the margin of its use.

In the second place, in the use of land the principle of increasing costs applies. Any attempt to increase the supply of products from land will be followed by increasing unit costs, unless the character of the capital instruments or the labor employed becomes more productive. More efficient labor or improved capital may offset the tendency for costs to increase, but this fact should not obscure the general statement that in the absence of these improved methods, an increase in supply of the services from land can be secured only at higher unit costs. To understand this statement will require the application of the principle of diminishing productivity to land.

Diminishing Productivity and Rent.—We may lay down the general proposition that, if it were not for the principle of diminishing productivity, there would be no rent. It will be necessary now to see how this principle applies to land and to show its significance in the explanation of rent. The principle may be briefly restated as follows: After a certain point has been reached in the utilization of a definite area of land, additional applications of labor and capital to that area may result in an absolute increase in the product, but the rate of increase will not be in proportion to the increase of labor and capital used.² By way of illustration, let us assume that we apply x labor and y capital to a definite area of land and get P product. We might then apply $2x$ with $2y$ and get more than $2P$. This would be a case of increasing productivity, for the product being greater than $2P$, would be in greater proportion to a unit of labor and capital than was previously obtained. But in accordance with the principle of diminishing productivity, beyond a certain combination of the factors of production, the rate of return will

¹ To the extent that there are persons who now render no economic service but who might become employed, we can say that the supply of labor is equally elastic with that of land. If the intensity of work should be increased and a larger total output produced, we could compare this effect with the intensive use of land. But neither the amount of unemployed labor nor the degree of intensity of its use is as great as in the case of land, hence, the supply of land may be regarded as more elastic than the supply of either of the other factors.

² Students should refer to Chan. VI for a fuller explanation of the principle.

decline. If we use $3x$ with $3y$ and get less than $3P$, it will indicate that the productivity has begun to decline per unit of labor and capital applied, *i.e.*, the addition to previous products is less than the earlier units yielded. The extent to which additional units of labor and capital will be used on a single plot of land will depend, as just noted, upon the demand for the goods. In other words, production may be carried beyond the point of diminishing productivity, as demand is the force that locates the margin of cultivation.

This, then, is the way in which this principle applies to land and it is valid on the assumption that we are dealing with a given area of land at a given period of time and with a given stage of technical methods of production. If, for instance, new methods of production were discovered, whether they consisted of new machines, new fertilizers, or a more scientific treatment of soils or plants, the effect of such changes would be to counteract or overcome the tendency to diminishing productivity. Whether the application of such methods would completely offset this tendency would depend upon the significance of the changes themselves. The introduction of new methods explains why, over a long period of time, we may find a given supply of labor and capital yielding a larger product per unit than formerly. The output per man in agriculture in this country has certainly increased in the past fifty years. The reader should distinguish these two forces and recognize that one pulls in one direction, the other in the opposite direction. Diminishing productivity tends to increase unit costs while improvements in the arts tend to lessen these costs. The principle of diminishing productivity is always operating, even though other forces may tend to counteract it during a period of time. Nevertheless, it is a dominant influence and controls the characteristics of production in any given stage of the arts.

With these facts in mind, we now proceed to answer the question of how to measure the services of land. Suppose we take first agricultural land. We have seen that the same treatment of separate acres will give varying results. Let us assume that, with a definite amount of labor and capital applied to each acre, we get 40, 35, 30, and 25 bushels per acre, and that an equilibrium between demand and supply is established with the use of the 25-bushel acre. This acre would then represent the poorest or least efficient parcel of land brought under cultivation and is

called the *marginal acre*. Since demand and supply are, according to our assumption, just equalized at this point, the demand price offered for the product will just balance the cost of producing it. For purposes of exposition, we have here simplified the illustration by considering the relations of only four separate acres. The relations shown, however, represent a condition that is approximated on a general scale in the use of all land. Demand for its services or products will come into equilibrium with the supply at some point and the balance between these forces determines how far the land will be utilized. The least efficient plots are known as *marginal land* and are commonly referred to as the *extensive margin of cultivation*, in distinction from the *intensive margin*, to be discussed presently. The important point to note is that the demand price just balances the unit costs on marginal land.

From this discussion it will be evident that three of the above acres will yield a surplus over the fourth. With the same treatment of labor and capital (x and y) we get 40, 35, 30, and 25 bushels, respectively. On the first acre there will be 15 units of surplus product, 10 on the second, and 5 on the third. These surplus products are regarded as the distinctive services of the respective plots of land. With a given outlay of labor and capital the cultivator of the first plot will have fifteen units more product than the man operating the fourth plot, but if the product from the fourth plot is required to meet the demand, then the

demand price will cover the unit costs on this acre, which consists, in accordance with the assumption up to this point, of the labor and capital costs alone.

Relation to Marginal Land.—

These relations can be represented by our familiar graph.

By representing different areas

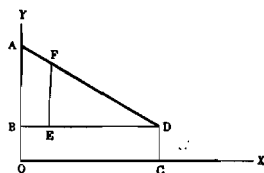


FIG. 27.

of land along the line OX in Fig. 27 and the productivity along the line OY , and by assuming that the variation in productiveness is infinitesimal, we have our familiar curve AD , which represents in this case the diminishing productivity of different pieces of land. The line DC represents the product at the margin of cultivation and the area ABD is the surplus product which the more efficient

acres yield. The distance from BD to AD , as EF , measures the distinctive service of any plot or acre, which, according to our previous discussion, constitutes the economic rent of that land. The area ABD , then, is the total economic rent arising from all of the land OC in use. From this analysis we may conclude that the distinctive service of any piece of land is the superiority of that land as shown by a comparison of its products with the products of the poorest or marginal land utilized.

Rent in Terms of Value.—Up to this point, we have discussed rent in terms of the concrete products obtained from the utilization of the land. Since most economic relations are conducted in terms of value, it will now be necessary to translate the physical productivity aspect of the problem into terms of price and value, remembering throughout the subsequent discussion that the existence of differences in productivity is the basic factor in explaining rent. The transfer to value terms is comparatively easy. The products or services yielded by land command a money price as determined by their market valuation. Likewise, the labor and capital used can be expressed in terms of value as costs, or outlay, in securing the product. According to the assumption in the previous discussion, if the same outlay in labor and capital is applied to all of the varying grades of land, the return per unit of outlay would vary. The area ABD in Fig. 27 can be thought of in terms of value which may be arrived at by multiplying the units of product by the market price per unit.

A Rent Formula.—But as the outlay for the labor and capital, as well as the receipts from the sale of the products, is in terms of value, we may, therefore, state a simple formula for expressing the rent of any piece of land. Let T equal the value of the products or services from the land, X the payment for the labor, and Y the expenditure for the capital used. Then the rent, R , will equal $T - (X + Y)$. If R equals zero, the land bears no rent and it is marginal, but if R is a positive quantity, the land bears rent and R will measure the superiority of any particular plot over marginal land in use and will constitute the rent of that plot.¹

¹ This method of expressing rent was first brought to my attention by George J. Cady, an instructor at Northwestern University. The method is particularly significant as a means of determining rent on any given plot of land.

Thus far in our consideration of rent we have made our comparisons between separate plots of land, and measured the rent from the extensive margin, or the least efficient plot in use. We have also assumed that the same amount of labor and capital of equal degrees of efficiency is employed on the various plots. We need now to ask, To what extent do our previous assumptions depart from actual practice? Two modifications of these assumptions need to be made in order to approximate the actual conditions in the use of land. We must recognize that neither the amounts nor the kinds of capital and labor applied to the different areas of land are equal. Since we are looking for the variations in the services of the land, we will continue the assumption that the labor and capital in use remain unchanged except as to the amounts employed. If the variations in the amount of the product were due to a difference in the kind of labor and capital used it would be incorrect to attribute this change to the land, but it should be thought of as a contribution of one or more of the factors whose efficiency had changed. Likewise, if there should be a variation in the amount of product with no variation in the kind of labor and capital used, the cause of the change should be assigned to the influence of land.

The Intensive Margin as a Basis for Measuring Rent.—One may well ask at this point, If some plots of land are more productive than others, why is it necessary to resort to the poorer grades of land? Why not continue to apply additional amounts of labor and capital to the better plots? It is at this point that an understanding of the effect of the principle of diminishing productivity becomes essential to the development of the next steps in the rent theory. The answer to these questions is that the successive applications of labor and capital to a plot of land will not yield a uniform amount of product. We find that as we intensify the cultivation of a definite tract of land, we soon encounter the tendency toward a diminishing rate of production. The absolute amount of the product may continue to increase, but the rate of increase will decline. The effect of variations of this kind can be seen in the table on page 126. It is there shown in column IV that the additions to previous product continued to increase up to the fourth application of the labor and capital used on the one plot of land, and that thereafter the net additions decline. The figures given in this table illustrate a general principle of production. Nature does not respond uniformly as

we apply more labor and capital to land, hence, it becomes advantageous to bring into use less fertile, or less effective, land as we intensify its use.

This principle operates in all uses to which land is put. In the United States we are not so familiar with the intensive utilization of land for agricultural purposes as we are with its extensive use, but in countries like Belgium, Denmark, and Holland, a large amount of labor and capital is applied per acre with the result that the acreage yield in these countries is much above that in America. We do find in this country that the land near cities is cultivated for garden products much more intensively than it is for general farming. Wherever land is intensively used, the pressure of diminishing productivity is encountered and its effect will be shown by variations in the output per units of labor and capital employed in its cultivation.

This principle applies equally to urban uses of land. Ordinarily skyscrapers are not built in small towns and villages, for the reason that land is plentiful and the labor and capital employed on it are more productive when applied extensively. Where land is relatively scarce, builders begin to economize in its use by erecting tall buildings, which is another way of saying that more labor and capital are applied to a given plot of land. But the higher the building, the greater the amount of labor and capital necessary to render a unit of service, because the increase in the height requires stronger foundations, the installation of elevators and ventilating systems, and other expenditures that make high buildings serviceable. While the price paid for space may justify these expenditures, nevertheless, the additional space is secured only as the result of additional amounts of labor and capital, which tend to yield less service per unit as the height of the building increases.

Since the principle of diminishing productivity applies to all land utilization, we come next to the question of the effect of the intensive use of land on its rent. As in the case of the extensive use, it is the demand for the services of the land that determines how much labor and capital will be applied to a given plot. Additional amounts will be applied until the price offered for the last unit of service added will be just balanced by the expenditure for the labor and capital necessary to obtain it. This point is known as the *intensive margin of cultivation*. Theoretically, the return to a unit of labor and capital at the intensive margin is

just equal to the return to a similar unit applied at the extensive margin. Demand for the services of land determines in both cases the location of the margin. When demand justifies the use of additional land, it will be worth while to cultivate the better land more intensively. It follows from this statement that the intensive margin is just as effective as the extensive margin as a basis for measuring rent. The difference between the value of the total yield and the value of the total amount of labor and capital employed on any given plot of land will measure its rent. Our conclusion is that the services of land can be measured from either the extensive or intensive margins of cultivation.

These relations can be shown from the following graph:

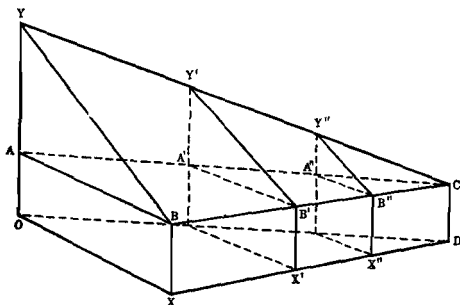


FIG. 28.

Along the line OX is represented the application of additional units of labor and capital to a given plot of land. The line OY represents the return to the most effective unit of labor and capital, the line BX , the return to the marginal unit, diminishing productivity per unit being shown along the line YB . The area $OXBY$ represents the total return from the plot, while the area YAB represents the total surplus of the more effective units of labor and capital over the least effective or marginal one. Along the line OD are represented different plots of land of equal area with the first plot, the line YC showing the varying returns yielded by the same application of labor and capital to the different plots, and CD representing the return to the marginal plot, which is just equal to the labor and capital expended upon it.

The distance of the line XD from the line OD represents the intensity of the cultivation of each plot. From this graph it will be evident that theoretically the return to labor and capital at the intensive margin will just equal that at the extensive margin, or, in other words, BX , $B'X'$, $B''X''$, are each just equal to the return CD . It is further evident that the volume above the plane $AA'A''CB''B'B$ is the total surplus of the more effective units of labor and capital over the least effective ones, both extensively and intensively considered. This total surplus represents the total rent from the whole area, just as the surplus above this plane represents the rent of any single plot, as YAB , or $Y'A'B'$. It is evident that the intensive margin of cultivation can be used just as effectively as the extensive margin in measuring the distinctive service of land and, therefore, its rent.

By applying this conclusion, it is feasible to determine the economic rent of a particular plot of land from the experience of that individual plot. According to the assumption underlying the statement of the rent principle, the market price of the products obtained at the margin may be taken as the true measure of the labor and capital costs involved in securing the marginal units of the supply of the commodity or services under consideration. This assumption is permissible because the prevailing market price just balances the marginal costs of producing the good, and supply and demand come into equilibrium at that point. In an actual case, the market price should be corrected for seasonal fluctuations as well as for cyclical and secular trends, which can be approximated by averaging the prices over a period of time of sufficient length as to give a normal price. In like manner, the yield of the plot and the labor and capital costs should be averaged over a period of time that will even out the fluctuations of these items. Then the average product, times the average market price for the good or service, minus the average total expense of operation, will give the relation of the return of this plot to the marginal use of land.

The Completed Rent Formula.—These relations can be represented by a slight modification of the formula presented above. Let T equal the average yield in dollars from a given tract of land.¹ Let X and Y equal the average unit costs

¹ This means that an average both of the output and of the price must be calculated for a length of time of sufficient duration to give a representative experience from the land being studied.

(corrected as has been suggested) of labor and capital used on the land. Let a equal the number of applications of labor and capital, then the formula would read $R = T - a(X + Y)$. Since $a(X + Y)$ equals the total expenditure for the labor and capital employed, we can substitute this sum in the formula and conclude that the rent equals the total yield in dollars less the total costs of cultivation. If there be a surplus, this surplus will measure the distinctive service of this plot in its existing usage, and will, therefore, be the economic rent of the land. If there be no surplus, the land is marginal for that type of service. If there be a deficit, then the land is submarginal and the labor and capital employed on it are receiving less than the return in other uses.

Caution in the Use of the Formula.—A caution in the use of this formula should be noted at this point. As it is stated, the formula assumes that the same degree of managerial ability is used in the cultivation of the land, an assumption that is permissible when an individual user is endeavoring to measure the rent of the land that he is utilizing. But if the formula were used to compare the rent of different pieces of land with different managers, a correction would be necessary to take account of differences in managerial ability, if any existed. In other words, it would not be quite correct to assume that all of the difference between the value of the total yield and the labor and capital costs was due solely to the services of the land. This difference would include differences that might arise from the various degrees of management ability which the operators possessed. There is no precise way of separating from the total surplus that portion which is due to the superior ability of the operator.¹ If the present operator possesses managing ability that is above the average, that fact will show itself in the surplus when calculated in the above manner, consequently, in the application of this formula, due recognition of this qualification

¹ TAYLOR, H. C., in his "Agricultural Economics," pp. 116-130, makes a distinction between capacity and efficiency and comes to the conclusion that as a guide for the combination of the factors of production that those "factors which possess high efficiency should usually be associated together. The high-efficiency land should be occupied by the high-efficiency farmer operating high-efficiency equipments." Since this is a problem primarily of farm management or a general theory of social economy, it does not belong to or constitute a part of the explanation of the general principles of rent and land income.

should be made whenever one is attempting to compare the rent of different pieces of land, or is endeavoring to establish the rent of an individual piece for purposes of investment.

While it may not be possible to measure the rent with mathematical exactness, nevertheless, the owner of land following this principle can approximate the rent in its present usage. This is particularly true of agricultural land where its utilization results in a concrete product that has a regular market price, but the principle is equally applicable to urban land, even though its use yields an intangible service. The services of urban land command a market price, and the bids in the form of contractual rent for the use of a site are a fair guide in estimating the value of the services and, therefore, the rent of that site. In bargaining to control the services a plot of urban land, the prices offered will not, in the long run, exceed what the services are worth, nor will those services be available very long for less than they are worth. Competition between users tend to keep contractual and economic rent approximately equal.

Rent as a Producer's Surplus.—From the foregoing discussion, it should be apparent that the difference between the total product from a given plot of land and the return to an equal amount of labor and capital applied at the margin of cultivation is the measure of the distinctive contribution of that plot to the social income. This is the theory commonly known as the "Ricardian theory of rent." It is so named because it was worked out and given publicity by David Ricardo, an English banker and economist of the early years of the nineteenth century. From the statement of the theory, it is evident that rent is a producer's surplus. It is the advantage enjoyed by those who occupy the more productive plots of land. The advantage may be caused by greater fertility of the soil, or more favorable location of the site. Whatever the cause of the superiority, it will show itself in a surplus return over marginal costs, and will accrue to the owner of the land. In this sense rent is a producer's surplus.

Relation of Rent to the Value of Land.—If the rent is known, it is then easy to determine the value of the land as an investment, for its value consists merely of the rent, calculated in the above manner, capitalized at the current rate of interest. If the rent were \$7 per acre and the rate of interest 5 per cent, then the land is worth \$140 per acre. This method of finding land value does not take account of the speculative element that may arise

from the future growth of population. In estimating the present value of land, the buyer usually includes in his reckoning the probability of a greater demand for its services in the future. This causes him to offer a price that is in excess of the market yield for investments of equal safety. In dealing thus with the speculative elements, we may say, as is commonly done, that the income from the land is capitalized at a lower than normal rate of return, or we might as well say that the buyer overestimates the rent from the land because of his optimism concerning its future uses. In whichever way the speculative element is put, it results in placing a value on the land that exceeds its present earning capacity when the income is capitalized at existing rates of interest. The formulation given above is intended to constitute a basis for measuring the present value of the rent in the current uses to which the land is put. This, in fact, is the true basis from which any estimate of the future possibilities of the land should be made. Hence, economic rent is the primary basis for estimating the value of land.

The question of the method of determining land value should be carefully distinguished from the question as to whether the present owners of land are securing normal returns on their investment. It is sometimes argued that landowners do not receive a surplus return and in fact are getting less than the yield on other lines of investment, the implication being that there is something faulty in the rent theory. It probably is frequently true that landowners receive a low return on their investment, but this question should not be confused with the problem under discussion. One is concerned with the explanation of the cause of land income and its relation to land value, the other with the acquisitive problem of land as an opportunity for private investment. Investors in land may overestimate its future possibilities and offer a price per acre, or other unit, that will yield relatively a low rate of return. As stated above, the capital value of land for investment purposes is merely the income derived from its use (rent) capitalized at the normal rate of interest. If the purchaser overestimates this income, the result will be a high capital value and a correspondingly low yield on the investment. This may happen even though the actual yield per acre, or other unit, may be increasing. Any influence that will intensify the demand for the services of land will increase its yield per unit, even though prospective pur-

chasers may, because of mistaken notions as to the future possibilities, offer a price that will give them an exceptionally low rate of return on their investment.

This situation may be illustrated by what occurred in many agricultural sections in 1919. The prices of farm products were unusually high at that time and the return from farm land distinctly above its normal level. If a farmer purchased land at that time without discounting properly the exceptional circumstances, as many did, he was sure to be disappointed when the subsequent decline in the prices of farm products caused the income from his investment to fall sharply and in many instances to disappear entirely. He suffered what every investor has to assume when he overestimates the prospective income from one of his investments. An understanding and a strict application of the principle of rent, as set forth above, would have led the prospective purchasers of land to inquire into the possible continuance of the existing demand for its services. They would have known that a decline in this demand would have been followed by a decline in land income, which would have necessitated a revision of the prices offered for it. The source of land income and the causes of its fluctuations should be distinctly separated from the issue as to whether landowners are receiving attractive returns on their investments.

Rent of Mines.—In the previous discussion, rent has been considered primarily from the point of view of agricultural and urban land, but the principle applies to all forms of land, although it requires some slight modification in respect to some types, such as mineral, forest, and any other types in which usage results in an actual depletion of the useful elements of the land itself. The qualifications may be discussed in connection with coal mines. Coal deposits vary in thickness, in purity of the coal, in the depth of the deposits, in the character of the formation surrounding them, and in other conditions that affect the costs of producing a standard unit of coal. These differences in the character of the deposits will result in varying returns to the application of the same amounts of labor and capital in different mines. In this respect the rent principle, as applied to mines, is identical with that already discussed.

But in the case of coal mines, as well as other similar types of land, the valuable element is constantly being removed. Hence, in dealing with land of this character as a private and acquisitive

problem, account must be taken of the depletion of the valuable element itself. This is an accounting problem rather than a general economic issue. In determining the returns to the owners, in addition to the general operating costs found in other industries, there must be added a depletion cost to cover the gradual loss of the valuable properties of this land. The only question of general economic consequence that flows from a recognition of this fact is concerned with the social problem of conservation of national resources. What has been said concerning coal land applies generally, at least, in connection with all mineral lands and forests. The differences in productiveness of different plots of land of this character account for the difference of returns to the owners. The fact that some coal mines yield a large return to the owners while others do not pay the expenses of operation is proof of the soundness of the conclusion concerning the application of the rent theory to this type of land. Aside from any peculiarities of the other types of land discussed in this connection, the same conclusions would likewise hold for them also.

Suggested References

References to the subject of Rent will be found at the close of the next chapter.

CHAPTER XIX

RENT (*Continued*)

Growth of Population and Rent.—There are several corollaries that flow from the acceptance of the rent theory as set forth in the preceding chapter. First, every increase in population, other things remaining unchanged, will tend to increase rent. This consequence follows because an increase in the population will increase the demand for the products and services of land without a corresponding increase in its productive power.¹ It might be asked, Why will not the increased number of people just enhance the labor power in the same proportion and therefore increase production by the same ratio as the increase in population? The answer here is again found in the application of the principle of diminishing productivity. Nature does not respond in the same way to the additional laborers as to those previously employed. Production per unit of labor is diminished. Therefore, to get the additional product required by the increase in the number of people, poorer grades of land must be brought into use. In other words, the margin of cultivation will be lowered and, hence, the difference in return to labor and capital at the margin and to the most effective uses of land will be increased. This relation can be illustrated by the following graph on page 436.

The graph used is identical with Fig. 28 that was used to illustrate the relation of the extensive and intensive margins of cultivation. Intensive use of the land is measured along the plane OXD , the extensive use along the plane $OYCD$. Suppose that the demand for the products of land could be satisfied by the cultivation of an area represented by the line OO' on OD . The extensive marginal product would then be represented by the line $O'Y'$. From previous discussion, we know that the better grades of land would be cultivated more intensively as the

¹ It is possible for an increase in population to result in a lower standard of living per individual but the total demand for products would likely increase even in this case

extensive margin was extended. This fact is shown by OH along the line OX . The intensive marginal product is represented by the line HF . The solid figure, represented by $IOHFY'O'$ constitutes the product that would just compensate the labor and capital employed in producing the entire supply, represented by $YOHFY'O'$. The volume above the plane IFY' represents the rent on the better grades of land.

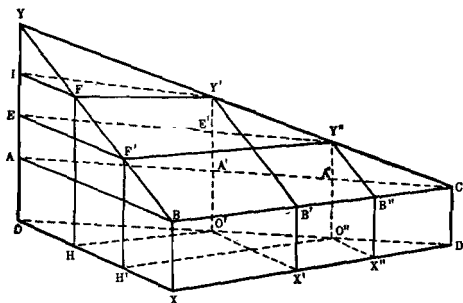


FIG. 29.

Now suppose that a growth in population takes place and, in order to secure the additional supply of goods, it is necessary to extend the cultivation extensively from O' to O'' on the line OD , and intensively from H to H' on the line OX . The marginal product has now been reduced from $O'Y'$ to $O''Y''$ at the extensive margin and from HF to $H'F'$ at the intensive margin. The marginal cost will have increased, because, according to our assumption, the same amount of labor and capital has been used per acre, or other area of land, while the product has declined. Under the new conditions, it would require the solid represented by $EOH'F'Y''O''$ to compensate the labor and capital. That portion of the figure lying above the plane $EF'Y''$ now represents the rent on the superior land. It will be noted that the land at O' on which there was no rent before the increase of population, now yields a rent represented by $Y'E'$. This statement of the rent theory accords with common experience. Everyone knows that as a region grows in population, the value of its land is increased. The increase in land value arises because the economic rent has

been increased and the capitalized value of this additional amount enhances the value of the land.

Improvements and Rent.—The converse effect on rent will follow from any form of invention, discovery, or any other means by which the labor and capital employed become more productive. Suppose the invention of a labor-saving machine that results in a greater product per unit of labor and capital used. The effect will be that a given product can be secured on less land. Under these circumstances, some of the poorer or less effective land will be thrown out of use, the margin of cultivation will be raised and hence the rent will be lowered. Often it is thought that the labor-saving device, or other form of improvement, will enable poorer lands to be brought under cultivation. What is overlooked in this thought is the limit to the assumption. We are assuming that there is no change in the demand for the products and that the only change is in the method of producing them. While it is true that the improvements in production would make it advantageous to cultivate poorer grades of land if the demand for the goods increased sufficiently to absorb the additional supply, yet it is equally true that if only a specific supply is to be produced, it can be secured from less land, and, therefore, the rent will tend to decline. This conclusion will hold true for any kind of an improvement, whether it be the application of a labor-saving device, or the development of a new type of product, or system of culture, so long as the new method enables the production of the supply with a smaller expenditure of labor and capital. The effect of any such improvement will be to counteract the influence of diminishing productivity and increasing rent.

Rent and Private Property in Land.—In the development of the rent theory, we have attempted to distinguish sharply between economic and contractual rent, the former being the full measure of the superiority of any piece of land over the products obtained at the margin of cultivation, while the latter is the amount agreed to between landlord and tenant for the control of that surplus. With these distinctions in mind, we are now in a position to consider the problem of private property in land. The discussion above has made it evident that the existence of a surplus on some pieces of land is quite independent of the ownership of, or property in, that land. We may now ask, Why do people wish to buy land? What does one buy when he purchases or

invests in land? What does one buy when he leases land? These questions bring out at once the problem of property in land. Clearly, when one buys land what he gets is the legal right to receive the economic rent of that land. If he buys as an investment, he cannot afford to pay more than the capitalized value of the economic rent, capitalized at a rate of interest that is approximately equal to that obtainable in other lines of investment. The amount of risk and uncertainty may affect the rate and, hence, the capital value, or in other words, the purchaser may regard land as a safer investment than any other he can find and, therefore, he may be willing to accept a lower rate of return. Under such circumstances, he will use a low rate of capitalization which will give a high capital value for the land.

Also, an investor in land may, as has been indicated, take into account the speculative possibility of an increase in the land value, which is only another way of saying that he is expecting an increase in the economic rent from the land. Hence, he may be willing to pay a capital sum for the land which represents a very low present interest return. Such an investor has a high estimate of the future uses of the land. But for our purposes at this point in the discussion, it is clear that what the investor buys is the right to receive all future increments of economic rent. Likewise, in the case of a lease, the tenant buys from the landlord, for a limited period, the right to enjoy the economic rent from the land. If he pays more than this rent, his capital and labor will receive less than normal returns. If he pays less, he will enjoy a profit, or a surplus item. It is apparent, then, that in both the case of ownership and of leasing, the basis of the contract is the control over what has been called in this discussion, economic rent, the control being perpetual in one case and definitely limited in the other.

As was shown in a previous chapter, private property is a legal institution and has developed as the result of human experience in an organized social group. It may be added, therefore, that private property in land is an institution that rests solely upon social expediency. It is generally accepted that property rights in land act as a stimulus to its more effective use and, in general, to the conservation of the natural resources. It is commonly believed that the private owner will keep his land in better condition than will a tenant, a belief that is borne out by observation in a great many instances. But to conclude

that private property in land always results in a conservation of natural resources would be erroneous, as can be shown in the case of the mineral and forest lands of our country. In many instances private property in these lands has resulted in exploitation and wasteful usage.

One of the most striking illustrations of this fact can be seen in the oil industry. Property rights in the oil follow the ownership of the superficial area. If an oil well is driven, it immediately becomes incumbent on the owners of all the adjacent land to drill wells also, or otherwise the oil underlying their property will be pumped out. There is no legal nor technological method to prevent this action and the result is that, as soon as oil is discovered, there immediately begins a competitive scramble to get the oil to the surface. Private property in oil deposits, therefore, results in a very great waste of an important natural resource and is one of the causes of the severe fluctuations in this industry. As a result of conditions of this kind the present generation may use up a natural resource with little thought for the needs of future generations. Such instances stress the fact that the social group may collectively have more permanent interests in these resources than can be measured during the lifetime of a human being and under such circumstances national welfare may require a limitation on present property rights in behalf of future generations.

Then again, the purpose of property rights should be kept clearly in mind. Property is a social institution developed as a means of protecting the individual in the enjoyment of the fruits of his own productive efforts. It has been found that the individual may be protected by rights in land that are definitely limited. A long-time lease may be sufficient protection for the individual to call forth all of his productive efficiency. Many plots of land, especially in large cities, are improved by magnificent modern buildings on the basis of a long-time contract. In fact, there is a growing disposition among present-day realtors to hold that revaluation clauses should be eliminated from these contracts, because the owner of the land receives in contractual rent the full value of its services at the time the contract is negotiated. Should the uses of the land increase during the lifetime of the lease, they should accrue to the leaseholder who has improved the land and made its services available. The owners, it is argued, will benefit from the enhanced value at the

expiration of the lease, when a new contract is negotiated. From both of these considerations, it should be apparent that property rights in land rest upon social expediency and may be subject to change in accordance with the policy that is accepted as harmonizing individual and general, or social welfare. This conclusion strengthens the previous one that the existence of rent is quite independent of the question of ownership and enjoyment. Property in land permits rent to accrue to private individuals as a matter of generally accepted, social policy.

Rent as an Unearned Increment.—A great deal has been written concerning rent as an unearned source of income. It has been asserted that land in all its various forms is a gift of nature and any income that may be attributed to it should accrue to society as a whole rather than to individuals and, furthermore, if private property in land income is permitted, the owners of land receive this income without rendering a corresponding service. It is easily perceived that this argument concerns the beneficiaries of rent and is, therefore, clearly a question of social policy. The issue as to whether private property rights to rent shall, or shall not, be permitted must be decided on the basis of its general social effects. This is a problem that need not concern us in a general treatment of the principles of rent. To repeat, the existence of rent is quite distinct from the question of who shall receive it. The former is a problem lying within the field of functional distribution, the latter in the field of personal distribution. Private property in land is solely dependent upon the accepted laws of property that prevail in any country.

The Meaning of an Earned Income.—While the ownership of land is a social issue, there are, however, certain phases of the question which depend upon economic principles. The concept of an earned income presumably implies effort or disutility on the part of the individual who receives it, and the income is regarded as an offset for the disutility involved in the production of a good. When one makes a pair of shoes, a hat, or any other usable good, the income which he receives is likely to be regarded as an earned income. But even here the commonly accepted notion of an earned income implies an ethical element. If, for instance, the maker of one of the above mentioned goods had a monopoly over the supply and could, on this account, control price so that his income was higher than that received by

other capital and labor similarly employed, the general feeling, judged by innumerable instances of this character, would be to condemn the price as extortionate, and the income in excess of normal returns as unearned. It seems evident that the notion of an earned income, then, is derived from the functional share that is imputed to one of the agents of production as it operates in a perfect market. We have seen in Chap. XVII that each factor of production tends to get as compensation, a value that is equal to the value of the services of that factor employed in production. In the case of labor and capital this return tends to offset the disutility of labor and of waiting on the part of the laborer and capitalist respectively. If, therefore, the income of any individual, received from the sale of some useful article, just compensates the cost or disutility on the part of the laborer and capitalist who assist in its creation, the income is regarded as earned.

It is argued in the case of land that there is no disutility involved in securing a supply and, therefore, those who receive rent have contributed nothing by their own efforts for which this income is a compensation. In fact, it is held that much of the utility of this land, instead of resulting from the efforts of the owner, is more likely to result from the efforts of others in their individual or collective capacities. In the case of urban land, an owner may hold his lot vacant while his neighbors improve their lots. He will benefit by their efforts and the value of his land will be enhanced without effort or cost on his part. Or, the city or municipality may improve the streets, put in sidewalks, water, gas, build schools, public parks, etc., the costs for which are spread generally. These general improvements, it is argued, will enhance the value of his lot out of proportion to any tax assessments levied against it. It is concluded, therefore, that his income is not earned.

In the consideration of this issue there are some phases of it concerning which there can be reasonable agreement of opinion. In the case of land that depends for its existence in its present usable form upon the expenditure of labor and capital upon it, the income received will be partly earned. For instance, some land now used for agricultural purposes was formerly covered with timber. Before this land was ready for crops, the timber had to be removed and the land cleared of stumps, roots, and such other obstacles to cultivation. The trees may have been

cut and sold for lumber and the labor and capital employed charged as a cost against that product. But, in addition, much more labor and capital had to be used before the land was tillable. What has been said of timberland applies equally to swampy land or land covered with glacial boulders. In both cases, labor and capital expenditures had to be made before the land could be used for productive purposes.

We may take the case of some urban communities, where tracts of land of considerable area have been reclaimed from the ocean or a lake by filling with refuse materials, or by other artificial means. All such areas have had a definite cost in labor and capital. Such expenditures are in nature identical with those required to make any other usable good, and the income received by those who have produced the land in any of the forms indicated is an earned income in the same sense as that received by any producer of a useful good. But once these expenditures have been made, we find that thereafter the value of this "made land" behaves differently from the value of other goods. It is at this point that the distinction between land and capital becomes significant. Thereafter, the value of "made land" and the improvements follow the rent principle rather than that of interest. Having no reproduction cost to limit its value, land may continue to increase in value indefinitely with the growth of population and increasing use. The increments of value that arise in connection with the improvements in the land, or with "made land", in excess of their costs accrue to the owners of the land without effort on their part, and on this account may be regarded as unearned.

In the case of all other types of land, there is no production or reproduction cost. In his thinking on this problem, the reader should keep clearly in mind the distinction between the social and the individual points of view. From the general, or social, point of view, land, except as indicated, is not made. To have a cost, or involve disutility, a good must be the product of human effort of some kind and, since land generally is not produced, its existence does not involve a cost. If it does not involve a cost, then cost can have no influence on its value, nor on the supply available for use. It follows from this that that portion of land income which results from differences in the natural qualities of the land, together with that portion which is in excess of normal returns on the costs of permanent improvements to the land, is

not earned by the landowner in the same sense as the payments made for the services of an individual who produces a good, or who furnishes the waiting necessary for capital formation. The recognition of this fact is not an argument that any definite social policy should be adopted in respect to land income, but does emphasize that, so far as the individual is concerned, the receipt of income in the form of rent rests on a different basis from the compensation to labor or capital in the form of wages or interest. The question as to whether land income shall accrue to the individual or to the state has nothing to do with the explanation of the cause or the amount of that income, and is one on which men may have honest differences of opinion as a matter of general social policy.

Land as a Private Investment.—So long as the prevailing laws permit private property in land, it will be bought and sold for purposes of private investment. There are important costs, or derivative disutilities, in connection with land ownership, which the prospective purchasers should carefully consider.¹ There is, first, the loss of return from the use of the capital value employed in some other direction. The capital sum invested in the land might be employed in other ways and, hence, yield an income, or it might be loaned for a contractual interest. In either case, an individual purchaser should compare the prospective return from the land with the possibility of income from these other uses of his capital value. Then there will be taxes, special assessments, transfer charges, including resale commissions, etc., all of which become charges against the income from the land. Interest on the above items should be compounded at current rates yielded by investments that are equally as attractive as that of the land. A little figuring will convince one that these carrying charges increase rapidly with time, and, unless the market price of the land continues to increase rapidly, these charges will prevent the return to the land from exceeding the normal interest yield and may very soon turn the investment into a loss. It may be found that the value of the rent just about equals these disutilities, but instead of the disutilities having an influence on the amount of rent, the causal relation runs in the opposite direction, i.e., the rent should be sufficient to offset the disutilities of the kind indicated before land is purchased as an investment. If the purchaser buys the land primarily with the

¹ TAYLOR, F. M., "Principles," pp. 466-467.

hope of an increase in land value, he is merely speculating on the future possibilities of the land. A return of this character which an individual may enjoy should be differentiated carefully from land income discussed above as rent. Purchases for speculative purposes will result in losses, unless the demand for the services of land continue to increase.

Ripening Costs of Land Utilization.—There is one aspect of this problem which has a bearing on what has been said about land income that deserves further consideration. Most land, and especially urban land, does not exist in a form adapted for immediate use, but requires that important costs be incurred in the preparation of the economic supply of such land as distinguished from its physical supply. Before urban land, for instance, is available for residential or other city uses, it must be withdrawn from some other form of utilization, which we will assume has been that of agriculture. The transfer of the land from agriculture will cause a loss of the income which it has previously yielded. In addition, it is necessary that such land be plotted, that streets and alleys be improved, that sewers, water, gas, and other similar improvements be provided. The costs of furnishing such improvements are less per unit (front footage, or lot) when the area prepared is sufficiently large to take advantage of modern mechanical methods used in preparing urban subdivisions. It would be clearly uneconomical, even if it were possible, for urban growth to proceed by the transference of one lot at a time. The result is that land is subdivided much in advance of its complete utilization. The interval of time, during which the demand for urban land is growing up to its supply, has been called a period of ripening use and is compared with the waiting period required while an agricultural crop is ripening into use. The preparation of such land, together with the waiting mentioned, involves a cost both in the form of a direct expenditure and an indirect loss of income which, under our present laws of property, is assumed by the individual landholder. To all such costs, Professor Ely has applied the term "ripening costs of land utilization," and argues that such a waiting or ripening period is present in all forms of land utilization.¹

That such costs exist is evident, and so far as they limit the economic supply of land for any specific use, they may be regarded

¹ Cf., "Economic Essays," contributed in honor of John Bates Clark, pp. 129-130.

as a charge against the income from that land. The recognition of these costs is an aid in distinguishing between an earned and an unearned income from the utilization of land, but they are of greater significance as an influence affecting the desirability of land as a private investment than as an aid in explaining the source of land income and rent.

Rent in Relation to Price.—Those who have followed the previous discussion carefully will have foreseen the conclusion that rent is not a cost that enters into the determination of the market price of a good in the same way as do wages and interest. This conclusion seems to run counter to practical experience. The business man who seeks a site for the operation of a store, or other form of business establishment, knows that one of the first contracts that he has to make is with a landlord, or his agent, for the use of space. The outlay for this space is just as definite and positive an expenditure as the payment of wages and interest, so that to him the statement that rent is not a cost is a violation of everyday experience. The confusion here is due to the difference in the individual and acquisitive point of view as compared with the social point of view. To the tenant, rent is an expense or outlay in the operation of a private business and, from his point of view, it is as necessary for him to reckon it among his expenses as it is to figure wages and interest.

But from the social point of view, the problem is somewhat different. What we are concerned with here is whether the payment of rent really affects the market price of the product or services derived from the land. Put in another way, we may ask whether the price of the product is high because the rent is high, or is rent high because the price of the product is high? After all, what is the causal relation between the price of a commodity or service and the rent paid for the use of the land necessary to produce it? Considered in this way, rent is a result rather than a cause of high commodity prices.

This relation can be shown by our familiar rent graph. We have seen that demand and supply for any good come into equilibrium at the point of highest unit cost of producing that good. Also, the highest unit cost is to be found at the margin of cultivation, which is located in Fig. 30 at *C*. At this point the value of the product *CD* is just equal to the expenditure for the labor and capital required to bring the good into existence. Rent appears only on those portions of the land lying between *O* and *C*, and is

represented by ABD . Since the amount of labor and capital employed on each plot is the same and the yield varies, as is indicated by the distance of the line AD from OX , then the highest unit cost is found on the no-rent land located at C . But as the market price tends just to cover the cost of producing the marginal increment of the supply, we then conclude that rent does not enter into this cost. The area ABD represents a

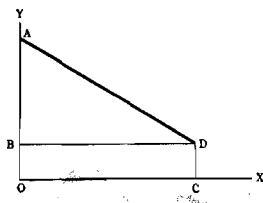


FIG. 30.

surplus above marginal costs. This surplus is due to the differences in the productive power of the better and the marginal uses of the land. Hence, economic rent that arises from a specific use of a given plot of land is not a part of the costs that are price-determining in the production of the commodities grown thereon.

We may illustrate this statement concretely by assuming that the land represented by the line OC is devoted to the growing of wheat. If we assume that the labor and capital expenditures on all plots amount to \$180, and that the yield on the marginal and on the best plots are, respectively, 180 and 250 bushels, then, the best land will yield a rent of \$70, when the market price is \$1 per bushel. The value of the product grown on the marginal land is just enough to pay for the labor and capital expenditures. The better land yields a surplus which is the result of the prevailing price of the wheat rather than a cause of it.

This discussion has assumed that all land is treated alike in the application of labor and capital in its cultivation. We have already seen that more will be employed on the better land as the less effective land is brought into use. However, the recognition of this fact does not affect the conclusion reached concerning the relation of rent to price, for the intensive margin is just as effective as the extensive margin in determining the high cost units of the product and is equally free from rent. Our general conclusion, then, may be restated in the following form; rent that arises from a specific use of land, does not enter as a price-determining item into the costs of the good produced thereon.

Rent as a Marginal Cost.—This conclusion concerning the relation of rent to cost is the one most commonly expressed in

textbooks on economics. In the formulation of the above statement, care has been exercised to confine this conclusion to rent that arises from a specific use of a given plot of land, or what may be called "differential" rent. There is, however, an important qualification of the relation of rent to price which will now be discussed under the topic of "marginal" rent. It is a matter of common observation that land often passes from a lower to a higher use, as, for instance, when grazing land is transferred to farming land, or farm land passes over into urban land to be used for residential or manufacturing purposes of one kind or another. The land in the lower use is likely to have yielded a rent. The question to be discussed here is, How will the rent in this lower use affect, if at all, the marginal cost of the product in the higher use? This relation can be shown by the following graph:

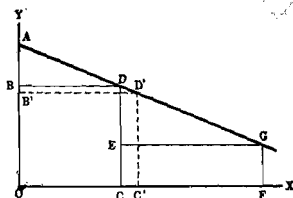


FIG. 31.

Suppose that an area of land indicated by the line OC in Fig. 31 is devoted to one use and another area indicated by the line CF is devoted to a lower use. It will be noted that a differential rent appears on each area, as indicated by ABD and DEG . Under what circumstances will the area devoted to the first use be extended to include a portion of the land devoted to the second, as indicated by the area CC' ? When demand for the product of the land in the first use is sufficient to pay all the costs involved in securing the additional volume of product, the transfer will take place. In addition to the labor and capital costs necessary for the production of the good, there will have to be added a sum sufficient to cover the total yield on the land in the lower use. What will determine the amount of this sum? Theoretically, it is the amount of rent that arises on the land in its less intensive use.

There are only two ways by which this land can be transferred from a lower to a higher use, namely, by purchase or by lease. If a lease is negotiated, a payment at least equal to the existing rent must be made before the present owner will part with the income arising from its present usage. If purchased, the sum paid will at least equal the capitalized value of the existing rent. Practically, the sum paid is likely to exceed the present rent, as the owner will endeavor through bargaining to secure a portion of the future increases of rent. Now, the question is, whether this payment for the control of the additional area of land to be devoted to a more intensive use has any influence upon the marginal cost of rendering the service or of producing the product on this land in its higher use?

It is argued by some authorities that the alternative uses of land can have no influence on the marginal cost of rendering a service, and they cite as proof of their contention that many instances can be found where land used, let us say, for urban purposes, does not yield a return sufficient to pay the capital and labor costs in its present usage and, hence, does not include a marginal rent as a part of the cost of rendering its present services. That such instances may frequently be found in all forms of land utilization cannot be doubted. They appear whenever land is kept too long in a given usage, and are illustrations of the obstacles to the free movement of the factors of production from one use to another. Two things should be kept in mind in thinking of this problem. First, the value of a good is never determined solely by its cost, but depends upon its utility to marginal users. The cost of reproducible goods fixes a limit below which that value cannot long remain. This principle of value applies to land services as well as to any other form of economic goods, whenever a shift in land uses is involved.

The second thing to remember is that cost always operates through its effect on supply. If the cost of producing a good is not completely covered by the price offered for it, the supply will be contracted until this balance is established. In the case cited, if it were practical to transfer urban land back into agricultural uses, when its urban use was less valuable than its agricultural use, then the return to the land in urban uses at the margin would not be less than the labor and capital costs, plus a rental charge equal to the rent that would arise in the agricultural usage of the land. But when land has once been transferred to

urban uses, it tends to become specialized and there are serious obstacles to transferring it back into agricultural uses. Sidewalks, paved streets, and other similar improvements, required for urban purposes, are obstacles that would retard this land from being used again in agriculture. Hence, if the demand for urban land services is not sufficient to bring all land that has been transferred and prepared for urban purposes into use, it will have no value for immediate purposes, whatever costs may have been incurred in securing control of this land and preparing it for these uses. If the future uses are uncertain, such land may have no value. It is a part of the physical, but not of the economic, supply of urban land.

Obstacles of the kind cited may cause individual owners of land to suffer a loss. As a matter of private economy, such owners may "revalue," or "write down" the value of their land assets, which, after all, is only a practical way of recognizing that the value of the services of the land is less than had been estimated. Such instances have little bearing on the theoretical issue being discussed, except to show the kind of practical problems that owners encounter in the utilization of land.

Returning to our main argument, we may repeat that if land used for agricultural purposes yields a rent, this rent will retard its transfer into urban uses until the return in those uses is sufficient to pay the labor and capital costs involved in rendering them, plus the rent yielded by the land in agriculture. If the land yielded no rent in agriculture nor in any other form of utilization, then, there would be no influence retarding the extension of the urban uses of the land except the labor and capital costs necessary to render urban land services. Under such circumstances, land would be a free good. The existence of this rent tends to restrict the economic supply of urban land and acts as any other cost in affecting the supply of services from such land. Marginal rent as defined, then, affects marginal costs and, through marginal costs, has an influence on price that differential rent does not exercise. The explanation of the relation of rent to price that has been presented is predicated on the assumption of perfect mobility of the factors of production and should be qualified to the extent that obstacles, natural or controllable, prevent the free movement of the factors from one use to another. The conclusion drawn from the specific illustration is applicable whenever land is transferred from a

lower to a higher use. The effect of marginal rent on land, as a retarding influence in its transfer, will depend entirely upon the circumstances of a particular case. Generally speaking, the higher the rent in the lower use, the greater the resistance to such a transfer.

In the above discussions of the rent principles, we were concerned with the productive powers of land. While competitive forces tend to equalize the returns to those who lease land, it does not follow that they all get the same returns. Differences in ability among tenants, or other users of land, are as pronounced as among entrepreneurs in other fields of business enterprise. Differences in returns due to superior managing ability should not be confused with the differences in the productive power of land. The rent problem is concerned solely with the services rendered by the land itself. The unusual return that can be ascribed to the superior ability of the user of the land has sometimes been called "rent of ability."

Taxation and Rent.—Another question of significance in connection with rent is that of the incidence of a tax on economic rent. Suppose a tax is levied upon the rent of land as it has

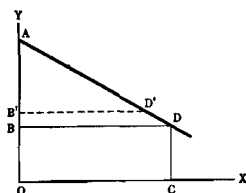


FIG. 32.

been defined, Who will bear the burden of that tax? The answer is that the whole burden of the tax will fall upon the present owners of the land and cannot be shifted to anyone else. In support of this conclusion, we may return to our rent graph as an aid in the exposition of our argument. It is evident from the assumption that the only part of the land that would be taxed would be the area lying between O and C in Fig. 32, for this is the only land on which economic rent appears. At C there is no rent, except possibly a marginal one as just explained, the amount of which is not affected by the tax levy. The cost of any good or service rendered by the use of this land is higher per unit at C than on any land between O and C . Hence, the cost at C tends to be price-determining. Therefore, a tax on the rent cannot be shifted to the consumer in a higher price, for the tax will not affect the marginal costs of rendering a service or of producing a good.

Neither can the landlord shift such a tax to the tenant, for if he tries to do so, the tenant will find that the maximum he can afford to pay is the full economic rent of the land, as this sum just measures the superiority of the land in question. If tenants make a contract assuming the tax, they will find that their capital and labor are receiving less than normal returns and, hence, they would be unwilling to take over the land with this additional cost. If the landowner wishes to sell his land, the prospective purchaser will discount the capitalized value of the rent by an amount equal to the capitalized value of the tax. In the foregoing graph, the rent is represented by the area ABD , the tax by the area $B'BDD'$. The tax cannot affect the marginal cost of a product or service from this land. Therefore, the consumers do not bear the burden. The prospective purchaser discounts the tax and lowers their bid for the land accordingly. They could offer a higher sum, equal to the capitalized value of the tax, if the tax did not exist. As the whole of the tax will be capitalized, the existence of marginal rent will not affect the problem.

The Single Tax.—There are those who argue that, since land is a gift of nature, rent should accrue to society as a whole and not to private individuals. It is held that, if a tax equal to the economic rent were levied on land, it would be a burdenless source of income to the government, because rent is an unearned share of income. It is further claimed that all governmental expenses could be defrayed from the revenue derived from this source, hence, all other forms of taxation could be abandoned. For this reason, this tax is commonly known as the "single tax." The issue raised by this proposal is, as has been shown, a matter of general social policy. If such a tax had always existed, or if it were confined to the increments of value that would accrue after a specified future date, the tax might be referred to as burdenless. But if it were levied in a form to absorb all of the existing rent, it would be far from burdenless, since it would destroy the wealth of all present landowners. If the single-tax notion were carried to this extent, it would result in the confiscation of the property of all present owners of land, and would establish social, in the place of private, ownership, as there would thereafter be no reason why an individual would wish to own land. All that he would wish under such circumstances would be protection in its usage, which could be secured by means of a long-

time lease with adequate provision for compensation for any permanent improvements that he might make on the land.

Since this issue lies in the realm of public policy and does not involve any peculiar economic principles, it need not be discussed in detail in this text. There can be little doubt that economic operations could be adjusted to a condition of universal leasing of land, but the process of making this change would doubtless involve heavy personal losses to those who have invested in land, unless they were compensated for their holdings. To go further into a discussion of the merits of the single tax as a social policy would carry us beyond present purposes. All that needs to be said is that the single tax, as a scheme of social reform, has been built on the rent principles set forth above. Care has been exercised in this discussion to distinguish between an explanation of the existence of rent and the question of who shall enjoy it. The former issue lies within the field of economic principles, while the latter is a question of social policy and hence does not properly belong with a discussion of these general principles.

Suggested References

- BROWN, HARRY G., "Common Sense and the Common Welfare," Part II, Chap. VI.
 BYE, RAYMOND T., "Principles of Economics," Chap. XIX.
 CARVER, THOMAS N., "Distribution of Wealth," Chap. V.
 ———, "Principles of National Economy," Chap. XXXVII.
 DORAU and HINMAN, "Urban Land Economics," Chaps. XXVIII-XXXIII.
 ELY, RICHARD T., "Outlines of Economics," 4th Ed., Chap. XXI.
 FAIRCHILD, FURNISS, and BUCK, "Elementary Economics," Vol. I, Chaps. XXXIII-XXXV.
 FETTER, FRANK A., "Economic Principles," Chaps. XIII-XV.
 HAYES, H. GORDON, "Our Economic System," Chap. LIV.
 HOBSON, J. A., "The Economics of Distribution," Chap. IV.
 HOLLANDER, J. H., "The Concept of Marginal Rent," *Quart. Jour. Econ.*, Vol. IX, pp. 175-187.
 RICARDO, DAVID, "Principles of Political Economy and Taxation," Chaps. II and III.
 RUFENER, L. A., "Principles of Economics," Chaps. IX-XII.
 TAUSSIG, FRANK W., "Principles of Economics," 3d Ed., Vol. II, Chaps. XLII-XLIV.
 TAYLOR, H. C., "Agricultural Economics," Chaps. XVII-XIX.

CHAPTER XX

CAPITAL AND INTEREST

Nature of Capital.—Interest, as a functional share, is that portion of the income from industry that is imputed, or attributed, to capital. Before we can proceed with an exposition of this subject, it will be necessary to examine into the nature of capital and in pursuing this task we soon discover that this term has more than one meaning in common usage. If we should ask a business man how much capital he has invested in his business, he would doubtless reply in dollars' worth, as \$50,000, or \$100,000. Here, capital is used in the sense of *capital value* and expresses an estimate of the social significance of the things evaluated. Should we ask our business man the nature of his investment, he would likely reply by telling us something concerning the character of the stock of wealth which makes up his business. This usage of the term refers to the *capital goods*, or *capital instruments*, which constitute the tools, machinery, and instruments of production. These two usages are common but they are not always carefully distinguished. When we are dealing with the assets of a business, as in accounting, the capital value concept is used, but when a physical inventory is taken, we are dealing with the capital goods aspect of a business. While both of these usages of the term require consideration in economics, it is with the latter that we are concerned when we think of capital as a factor in production and, therefore, as a claimant on the income from industry. In this sense, capital may be defined as that portion of durable wealth, other than land, that is used as an aid in production.

Social and Individual Capital.—There are some forms of wealth that are regarded as capital by the individual, which, from the social point of view, are consumers' goods. Take, for instance, a dwelling house. It is a consumers' good from the point of view of society, yielding satisfactions directly, but it may be let by the owner to a tenant for a stipulated sum. The owner will regard this dwelling as a part of his capital from which he is deriving a

portion of his income. There are other forms of wealth that are technically finished goods awaiting purchasers, such as the stock of groceries on the shelves of a retail grocery. To the grocer, these goods are regarded as capital from which he expects an income, but from the point of view of society they are in reality a stock of consumers' wealth, ready to yield satisfactions.¹ The term "acquisitive capital" is frequently used to distinguish this type of goods from machines, tools, and other material aids to production. The latter type of goods is called "producers'," or "productive," capital. It will be recognized that in some instances there may be a gradual ripening of one type into the other, while in other instances, as the case of machines and tools, the types are quite distinct. Loan interest will appear in connection with transactions involving either acquisitive or productive capital, but the explanation of interest as a functional share deals primarily with producers' capital, or capital in the social sense. The income that the individual enjoys from acquisitive capital tends to conform to the interest yield of productive capital.

Economic and Contractual Interest.—The most common form in which the phenomenon of interest appears in ordinary business transactions is that of a loan of money, or credit. A man borrows a definite sum of money, say \$1,000, from a friend, a bank, or other lending agency, for a definite period of time at an agreed rate of return or, as commonly called, interest. The borrowing and use of money in this form, while exceedingly important as a part of the business process, is not the fundamental phase in the explanation of interest. All that can be done with a loan of money, or credit, is either to get control of the instruments of production—land, labor, and capital, and combine these to produce a good which consumers will buy, or to use the funds to purchase goods to be consumed immediately by the borrower. As will be shown later, *loan or contractual interest* is likely to be affected by the conditions of the money market or the supply of loanable funds. *Economic interest* is not so affected, as it consists of that portion of the income from industry that is assigned to capital, while contractual or loan interest is the amount agreed to between the borrower and the lender for the

¹ It is not quite accurate to say of a stock of goods on the shelves of the retail merchant that the productive processes have been completed. The work of the merchant in selling and delivering the goods to the consumer is essential to the completion of this task.

purpose of controlling the economic interest and tends in the long run to conform to it. When borrowers and lenders are making a contract, the real basis of that contract is the control over the earnings of capital, or economic interest. The borrower wishes to get control over capital instruments to use in his business, and his ability to pay contractual interest depends upon his success in securing an income from the use of these instruments. The interest problem, then, lies in the explanation of economic interest or in the use of capital in the sense of tools, machinery, etc.

Three Major Issues.—There are three questions around which the interest problem may be set forth. These may be stated as follows: Why is it necessary to pay interest? How is it possible to pay interest? And what determines the rate? We shall now proceed to answer these questions in the order given above.

The Formation of Capital.—The answer to the first of these questions is comparatively simple. The reason why it is necessary to pay interest is because capital is scarce. Put in another way, before one can get control of capital, he must pay a price, which may be in the form of a series of annual payments, or it may be a capital sum equal to the capitalized value of one of these annual payments. In the one case we see the phenomenon of loan interest; in the other, the purchase of the full property rights in the capital which carries with it the right to enjoy the earnings of the capital during the lifetime of the instruments themselves.

But why are capital instruments scarce? Since capital is not a gift of nature, it entails disutilities or costs, the character of which will now receive our attention. There are two steps necessary in the formation of all capital instruments. Tools, machines, and all such forms of capital have to be produced. The production of these instruments requires the combination of land, labor, and other capital instruments. The process, therefore, involves the ordinary costs and expenses that we associate with all productive operations and, in addition to these ordinary operating expenses, there is another cost in capital formation that involves a time element and constitutes what is known as "waiting."

In order to see the importance of this element, let us consider some of the general conditions that surround all production. Before production of any kind can go on there must be an

accumulation of a stock of goods of various kinds. How would it be possible to make an automobile, if the manufacturer could not go into the market and buy iron, steel, glass, leather, cloth, and rubber? In all lines of production before a finished good can be put on the market, the materials out of which it is to be fashioned must have been made some little time prior to the beginning of the work on the article itself. The machines to be used in the productive processes had to be made a considerable period before the manufacturing process could be undertaken. Prior to the manufacture of the machines was the production of the materials from which they were made. In addition to all of the goods of this character, and before labor could be available for work on an automobile or other form of wealth, there had to be a stock of consumable goods that would sustain the men who were engaged in making it. They had to have food and other necessities on which to live while they were manufacturing the machine, tool, or other form of an economic good. It should be evident from this analysis that a reserve or stock of wealth is necessary to the formation of any type of wealth, including that of capital goods.

Where does this stock or reserve come from? How does it get into existence? There is only one way to accumulate such a reserve, namely, by the postponement of present consumption on the part of someone. Someone must forego the satisfaction of consuming present goods. This postponement, or waiting, necessary to accumulate the reserve is known as "saving," and all capital formation involves this waiting element. The form in which the saving is done does not change the essential character of the waiting process. Among primitive peoples, before any tool can be fashioned, their present desires must be gratified and either a stock of goods saved to sustain them while they devote their time and energy to making a tool, or, if not a stock of goods, they must have enough free time above that required to secure the daily necessities to enable them to devote part of their energies to fashioning the tool. However simple this primitive tool, the making of it involves the two essential steps in all capital formation, *viz.*, *waiting* and *production*.

In modern times, the waiting or saving process has, like most every other phase of our economic life, become highly subdivided. We do our saving now by means of money and credit instruments. A simple illustration will suffice to show the present method of

saving. From whatever source it may arise, income is received today in the form of money or credit instruments. Suppose that an individual refrains from spending his entire money income, let us see what he can do with this surplus. He may lay aside a stock of money and keep it in his own strong box, or he may rent a safety-deposit box from a bank and keep his money there. But he will be more likely to deposit it with the savings department, keep it in his checking account, or use it to purchase some form of an investment security. Any one of the last three methods of accumulation would have approximately the same industrial effects, except for minor differences in the manner in which these effects are worked out in modern industry.

Let us assume that the person in question decides to buy an industrial bond—an assumption which accords with experience and has the advantage of bringing into bold relief the relations of modern methods of saving to capital formation. The bond will be purchased from an investment banker who regularly makes loans to industrial enterprises. These loans arise because some entrepreneur thinks that he sees an opportunity to expand his business to advantage and applies to the banker for a loan. After convincing the banker of the soundness of his project, a loan for a specified amount is made, and as security a mortgage is placed on his property. The entrepreneur will now use the proceeds of his loan to hire labor with which to build the addition to his plant and to buy machinery to equip it. The energy of the men so employed will be directed in such a manner that the result of their efforts is a net addition to the capital equipment of his plant. The funds provided by the banker have enabled these men to buy present goods while they were employed in manufacturing the tools, machinery, or other equipment desired.

In the meantime, the banker has issued on the basis of the mortgage industrial bonds in convenient denominations, as \$100, \$500, and \$1,000, which he offers to sell to his customers. Let us assume that the person mentioned above buys one of these bonds. He has by this purchase in reality supplied the entrepreneur, through the banker, with the funds necessary to make the addition to his plant and equipment. By postponing the use of a part of his money income, he has made this purchasing power available for use by others. Those who secure it can demand present goods for their daily needs. If the process ended at that point there would be no addition to capital. But in order

to secure these funds, these men give in exchange labor services which are utilized in fashioning the tools or other equipment. Thus, while the methods of accomplishing the results have changed from primitive times, the essential steps remain the same, namely, the postponement of present consumption and the accumulation of a stock or reserve, and the utilization of this reserve in support of those who produce the capital instruments. All new capital involves these two steps in its formation.

Saving and Hoarding.—From the foregoing statements, the difference between saving and hoarding should be apparent. The postponement of present consumption and the accumulation of a surplus would involve waiting and constitute saving, but it does not represent an addition to the capital of the community. The man who puts his surplus in a strong box, or in a safety deposit vault, is hoarding wealth, but he is not contributing to the capital equipment of society. In like manner, the primitive man might hoard a stock of concrete goods without adding anything to the ability to produce more goods. In either the modern or the primitive case, the individual might call forth a definite part of his accumulated reserve for his immediate use. Hoarding has no bearing upon the interest problem. It is only when the stock of wealth that has been accumulated comes back into general use that the waiting factor becomes a phase of the interest question.

But what bearing do these steps of waiting and the utilization of the reserve so accumulated have on the explanation of the necessity of interest? The significance of this analysis lies in its effect upon the scarcity of capital. Human nature is so constituted that, other things equal, present goods are of greater importance than future goods. Life at best is uncertain so that people generally prefer present goods to the promise of goods at some future date. In other words, people do not like to save, or to postpone present consumption. If everybody liked to save more than to consume, capital instruments would eventually become free goods. It is this dislike to postpone present consumption, this waiting element, that limits the supply of capital and enables it to command a price. Waiting usually involves a disutility in addition to the labor or effort required to produce a good. It is, therefore, necessary to offer an inducement in the form of a price to get people to save. We have to pay interest, then, because the formation of capital requires the accumulation

of a reserve, or a stock of wealth, which can be secured only through the postponement of present consumption, or waiting. The price which is paid for this element of waiting is in reality interest. While this statement is a necessary part of the explanation, it is not a complete statement of the theory of interest.

The Maintenance of Capital.—A distinction should be recognized between waiting as an element in capital formation and the maintenance of capital that has been produced. Capital goods, such as machines, tools, etc., wear out and have to be replaced. Every factory owner, or other user of machines, knows that it is necessary to provide for the wear and tear of his machinery. If he disregards this matter, he will soon find that, instead of a going concern, he has only a worn-out plant. All business men know that it is necessary, as a matter of business policy, to make ample provision for the repairs to their capital instruments and to build up a reserve to replace them when they are worn out. Such a reserve is known in modern business as a *depreciation fund*.

A question of some significance at this point is, whether the setting aside of a depreciation, or replacement fund, involves a cost. Some authorities have argued that once capital has been saved, it has a permanent existence. It is argued that during the lifetime of a machine or tool, there will be created a surplus fund from which the machine or tool may be replaced and, unless the capital instrument produced such a surplus, it could not be said to be truly productive. The better opinion is that provision for depreciation requires a conscious effort on the part of the management of a business in setting up such a reserve. Unless adequate care is exercised, a business may use up its capital by declaring dividends out of capital rather than out of the true earnings of the business.

A nation may experience the same result, unless care is exercised in respect to its capital equipment. For instance, immediately following the establishment of the Russian Soviet System, the demand for present goods was so urgent that adequate provision was not made for the maintenance of the capital equipment of that country. Very soon the need for capital compelled the new form of government to make concessions to outside sources and from current reports there is still in Russia a serious deficiency of capital goods. Our conclusion is that capital maintenance requires vigilance on the part of the business manager, and capital, therefore, does not reproduce itself without

conscious attention. From the point of view of this discussion, depreciation must be sharply distinguished from interest. Interest is a sum in excess of the amount that will be necessary to replace the capital and to keep it in operating order.

Obsolescence.—As a problem in business management, the entrepreneur has to reckon also with *obsolescence*. His machinery may be technically efficient and in excellent operating condition but a new machine may be invented that so lessens unit costs as to make it necessary to scrap his old machines and replace them with new ones. While this is a very important problem from the point of view of the business manager, it does not occupy a large place in the explanation of interest. The scrapping of usable machines to be replaced by more economical ones, has to be regarded as one of the wastes of the competitive system and a part of the general costs of a dynamic and changing industrial order. Our general conclusion, however, is that interest has to be paid because capital is scarce, and the scarcity of capital is explained by the disutility involved in waiting.

The Possibility of Paying Interest.—The previous question has to do with the supply of capital. But the possibility of paying interest has to do with the demand for it. Borrowers want to control capital because they wish to add to their private incomes. They assume that by means of a machine, or a tool, they can produce more goods, or produce the same volume at lower unit costs than without its aid. We might reply to the second of the above questions by saying it is possible to pay interest because capital is productive. Rightly understood, this answer will suffice, but it is not sufficient to think solely of the physical productivity of the capital instruments. It is easy to understand how a man with a tool or a machine, can turn out more units of product than if working barehanded. Productivity in this sense is evident enough. Anyone who has watched the complicated machinery of a modern factory will readily assent to the proposition that labor employed with machinery will turn out more units of product than the same labor working with old-fashioned tools.

In most cases, this form of explanation is sufficient to show how it is possible to pay interest, but it is not quite complete. It frequently happens that more units of a good actually sell for less total value than a smaller volume. The interest question, then, has not been completely answered until we have explained

the source of the additional sum of value that the payment of interest entails. To put the case concretely, let us assume the loan of capital, the value of which is \$100, and the contractual interest promised 6 per cent. At the end of the contractual period, which we will assume as a year, the borrower has to return \$106. In answering the question concerning the ability to pay interest, the problem centers about the source of this extra \$6. Where does this additional increment of value come from? Proof of physical productivity is clearly not a complete solution of the problem and we must, therefore, find the source of what Böhm-Bawerk has aptly called "value-productivity."

The only explanation for the existence of a surplus of value, resulting from the use of capital, is found in wise entrepreneurship. As the chief borrower of capital, the entrepreneur gets control of land, labor, and capital in order to produce a good which he thinks will sell at a price that will cover all operating costs, including a return on the capital. He plans this production in anticipation of the demand for the goods and, unless his judgment of future demand is correct, there will not only be no surplus for the payment of capital, but there may not even be enough to pay the other operating expenses. The number of business failures that yearly take place is sufficient proof that wise direction in the use of the factors of production is a necessary condition for the existence of an income sufficient to offset these costs. But assuming wise entrepreneurship, which means in this case keen foresight as to the course of demand, the ability to pay interest arises out of the productivity of capital. By the wise use of capital a stock of goods will be produced which will sell for a sum that is sufficient to pay all operating costs, including a compensation for the service of waiting.

The Determination of the Rate.—As a functional share, interest has been defined as that portion of the income from industry that is imputed to capital as a factor in production. The question may well be raised at this point as to why we speak of a *rate* of interest, when we are thinking of a division of the concrete goods and services that result from the productive processes? The answer is, that the receipt of interest in modern economic society is primarily through a money contract. The negotiation of that contract involves an evaluation process, both on the part of borrowers and of lenders. The lender is comparing the relative significance of the offer of a sum of value at some future

date with his estimate of a lesser sum expended for present goods. His attitude will be conditioned very largely by the state of his present desires and his estimate of his future needs. The borrower is making similar comparisons, but he is ordinarily placing a higher estimate on his present desires than on future necessities. As a business man, he estimates that if he can command more capital now, it will so enhance his future income as to enable him not only to pay the contractual interest but to leave him a sum in excess of the other normal expenses of production. It is clear from these statements that interest is a phase of the value problem.

While it would be possible to work out a division of the concrete products and services arising from industry among the factors of production, and to speak of interest in terms of a definite quantity of those goods, it is much more common to think of the income of capital as a ratio between a capital sum invested and the capital sum yielded by its use. This is the form in which interest commonly appears and, while the other relationship is fundamental, the reader can probably follow the explanation of the rate of interest in terms of value more easily than in terms of concrete goods. Loans of capital value are made and an agreement entered to pay back a capital sum. Let us suppose that A lends B \$1,000 and B agrees to pay A \$60 per year for the use of this capital value. The rate in this case is 6 per cent. The question to be explained is, Why the amount paid for the use of capital value is 5, 6, 7 per cent, or whatever the rate happens to be? This statement of the case places the problem at once in the field of value, and the rate paid becomes a price for waiting. Our problem really is, how to apply the supply and demand principles to this question. On the supply side is the lender who is willing to supply capital funds and postpone present consumption for a definite consideration. On the demand side is the borrower who is willing to offer a price for the control of present goods by an exchange of rights to a larger share of future goods, the amount of which will depend upon the effectiveness of his use of the capital instruments in production. In the analysis of the determination of the rate, we will begin with the supply of capital available for lending.

We have already seen that waiting is an essential condition to capital formation and that waiting constitutes a disutility or cost which must be compensated. Those who save a portion of

their money income and accumulate a reserve have, by this act, had the effect of accumulating a stock of present goods in the market which they could have purchased with that income. Instead of using it themselves they offer to lend it to others, which really means a transfer of the rights to purchase those present goods. The question of importance for us now is, What determines the amount of savings which the suppliers (lenders) of capital set aside?

In considering this issue, we should recognize at once that there is a vast amount of difference among people in respect to their attitude toward the postponement of present consumption. Some have a high degree of prospectiveness, *i.e.*, they regard their future needs as very significant, hence, they will strive to provide for the future by saving a portion of their present incomes.

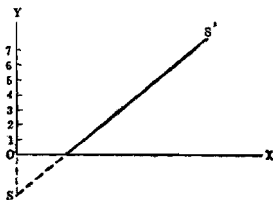


FIG. 33.

Then, there are those who have a low degree of prospectiveness and consequently place a low significance on goods that will be available at a future date. Finally, there are those who have very large and assured incomes, and on this account are in a position to gratify their present wants more fully than those with smaller incomes. Anyone so situated is likely to save a larger amount than those whose present desires are more urgent.

Enough has been said to indicate a wide variation among people in respect to saving. There are some who feel the responsibilities of family or of old age so keenly that they would be willing to save even if they had to pay for the protection of their savings. We can represent this difference in attitude toward saving, and therefore toward the furnishing of a supply of capital, by an ordinary supply curve. Along the line OX in Fig. 33 represent the amount of savings, and along the line OY represent rates of interest. The line SS' will represent the increasing rate that will be necessary to induce an increasing volume of saving. The positive inclination of this line indicates that an increasing volume of saving entails an increasing disutility and that the additional sums can be secured only by encroaching

more and more upon present consumption. It will be noted that there is an area below the line OX . This indicates the saving that would take place even if the saver had to pay for the protection of his savings. Such payments have been referred to as *negative interest*. If it were possible to meet the needs of industry from savings that arise in this way, it is evident that there would be no interest. Capital, in fact, would be a free good, as savers would pay to make sure of a return of their savings. The fact that a payment is necessary to secure capital is sufficient evidence that an adequate supply can be secured only by offering a reward to those who can be induced to postpone present consumption, hence, the price paid for the disutility of waiting is a very important step in the explanation of the rate of interest.

Time Preference.—Most individuals place a higher estimate on present goods than on the prospect of goods at a future date. This difference in the estimate that one places on equal quantities of wealth of like kind in the present and at a future date has been called "time preference."¹ The time preference of an individual depends upon the degree of prospectiveness that he possesses. Since individuals differ in this respect, it follows that their time preferences will differ. An individual will continue to save until he has established a balance between his estimates of present and of future goods. Those who have the highest degree of prospectiveness (those who are willing to pay negative interest) will continue to save until their desire for present goods, plus the cost of protecting their savings, offsets the satisfactions expected from their future income. Those who have a low degree of prospectiveness (marginal savers) will save little when their evaluations are unaffected by a special inducement or compensation for the disutility of waiting. An offer to pay something for the assumption of this disutility will affect their attitude toward saving. If the offer exceeds their time preference, they will forego present consumption and accumulate a reserve which adds to the supply of funds available to lend. The variation in the attitude of different individuals is represented along the line SS' in Fig. 33. As the rate or price offered increases, the number of savers as well as the volume of savings will increase. At any one rate there

¹ Professor Fetter has introduced this term into the literature of economics and explains interest solely by the cost of time preference. Cf. *Am. Econ. Rev.*, Sup., March, 1927 pp. 62-105 for a very full statement of his theory as well as a history of the interest discussion.

are those who are marginal savers, or put in another way, the rate offered is an accurate measure of the time preference of marginal savers.

But we have seen that the time preference of individuals differs, partly because of inherent differences in the persons themselves, and partly by the differences in the amount of income each one has at his command. Whatever the cause, there are varying estimates at any one time and this fact affects the determination of the supply of capital. The demand for savings, as expressed by an offer of a payment for their control, will determine how far savers will go in postponing present consumption. Since there is a variation in individual estimates, there will always be a marginal group of savers, *i.e.*, those who are just on the point of spending, who are weighing the advantages of a present use against a future use. From the side of supply, the estimates of these marginal savers constitute the most important influence in determining the volume of capital. The rate that is necessary to induce them to wait constitutes a price for the service of waiting. This price, or rate, paid to marginal savers as compensation for the disutility involved in postponing present consumption tends to become the rate of interest. It represents the cost of securing the marginal supply of capital, whether capital be thought of as loanable funds (capital value) or as capital instruments (capital goods).

Corporate Saving.—It might be inferred from the previous discussion, that all savings are made by individuals, and that one consciously balances the present against the future. This impression needs correction in two directions. First, the individual is not as calculating as the argument implies. No one is actuated solely by economic motives but, to the extent that these motives affect his conduct, they will result in choices that conform reasonably to the above conclusion, namely, that differences in return will have an effect upon the volume of savings. The second modification is, that a vast amount of savings results from corporate action rather than individual action. It has been asserted that fully 60 per cent of the present assets of corporations have been accumulated by a reinvestment of surplus earnings of the respective enterprises. This means that the boards of directors of these various industrial undertakings have saved for the stockholders vast amounts and, as a result, have greatly increased the amount of social capital. The

desire to build a large and successful business concern is doubtless a more potent cause in determining the policy of such boards of directors than the prevailing rate of interest. Their action is not entirely independent of that rate, however, because the stockholders will demand a return equal to the normal earnings of capital. Reinvestment of surplus earnings usually takes place in those enterprises that are unusually successful, which makes it possible to pay customary returns on investment and enables the directorate to expand the concern without encroaching on the ordinary return paid on loanable funds.

In addition to the savings that arise from the action of boards of directors of industrial corporations, the insurance companies also cause a large amount of saving. The multitude of policy holders contribute vast sums into the treasuries of the insurance companies. Funds so contributed become invested in the more conservative lines of business and, to some extent at least, add to the supply of social capital. The rate of interest has little to do with the contributions which policy holders make to the insurance companies. Protection is the important thing in which the policy holder is interested when he purchases insurance.

The recognition of these important sources of new capital, and the fact that the disutility of waiting plays little part in these two forms of saving, have led some writers to discredit the cost of waiting as an influence in the explanation of the rate of interest. No one familiar with present-day methods of capital formation can deny the significance of these two sources of new capital. But notwithstanding their importance, it is clearly evident that appeal has still to be made to the individual in order to meet the industrial needs for additional capital. The flotation of bonds by investment bankers and other lending agencies is convincing evidence that it is still necessary to offer a return in the form of contractual interest in order to induce individuals to provide an adequate amount of savings.

Our conclusion is that, even though there is compulsory saving, as in the case of the corporation and insurance companies, or even costless savings, as in the case of negative interest, so long as appeal has to be made to individuals to induce them to postpone present consumption as a means of securing new capital, the cost of waiting is an important element in securing the supply of capital. It will be noted from the above discussion that marginal savers were the ones whose influence was potent in

determining the rate. All those below the margin, and here we could include the savings of corporations and insurance companies as well as those willing to pay for the protection of their savings, would not have a determining influence. In fact, they all enjoy a "saver's surplus," because they receive the market rate determined by the marginal savers, whereas they would be willing to save, even if the rate were less. There can be little question, however, that the cost of waiting is an important influence in determining the supply of capital.

Demand for Capital.—The determination of the interest rate cannot be made solely from the side of supply, or cost, any more than the price of a good can be explained by cost alone. Demand for capital is an important influence in determining the rate of interest. There are two types of borrowers—those who borrow for purposes of consumption and those who expect to use the borrowed funds in productive enterprises. While both types of borrowers seek to control purchasing power, it is evident that the purposes are very different. The business man will soon utilize his borrowed funds in expanding his business from which he expects an additional income, whereas the borrowers for purposes of immediate consumption, the so-called "spendthrift loans," have no such expectations. The explanation of interest lies primarily in the use of capital for productive purposes.

The entrepreneur will get additional income from the wise direction of more machines, tools, materials, etc. But he is confronted in the use of these additional capital instruments by the principle of diminishing productivity. As the supply of capital instruments of a given kind increases in number without a corresponding increase in the demand for the products, the significance of each unit of capital tends to decline. Technically, each capital instrument may be capable of turning out as many units of goods as the previous instruments, but the significance of these products becomes less with the increasing volume of goods. As a result we impute to the capital instruments a diminishing product or diminishing share in the income from the industry. This fact has an important bearing upon the demand for capital. It tends to determine what the business can offer as a price for waiting. We may illustrate this influence by the following graph:

Along the line *OX* in Fig. 34 is measured the quantity of capital instruments of a given kind, and along the line *OY* is measured the significance, or value, contributed per instrument. The line *DD'*

represents the declining significance of the service rendered per unit of capital as the number of units increases. This line is the resultant of two influences, namely, diminishing physical productivity of the capital instruments, and diminishing utility of the added units of the product. The former is expressed in terms of physical units of product, while the latter is expressed in terms of the value, or the price, of these products. The quantity of the product times the price which the goods will command will

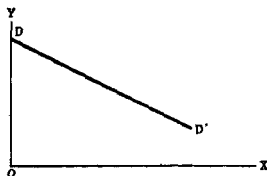


FIG. 34.

give a value expression to the significance rendered by the additional units of capital instruments employed.

This theoretical statement can be expressed concretely as follows: When a business man thinks of adding more machines to his existing equipment, he has to consider whether the additional prod-

ucts obtained by the use of these machines will sell for a sum sufficient to pay for their cost during their lifetime of use, and for any other costs necessary for their operation. He has to think not only of the rate of physical production of the additional machines, but also of the price at which he can dispose of the additional volume of commodities. In connection with the former phase of the problem, he is likely to encounter the principle of diminishing productivity, while in the latter he will be affected by the principle of diminishing utility in the disposal of an additional volume of goods. Consequently, additions to capital equipment, as represented along the line OX in Fig. 34, are accompanied by a declining social significance of the service rendered by these additional units. It is evident, therefore, that the offer which the business man can make to control capital declines as the amount of instruments of a specific kind increases.

It would thus appear that the interest rate is the resultant of two influences that operate in opposite directions, and is, in fact, only another phase of the general problem of value. The securing of a supply of capital, whether we think of it in terms of loanable funds seeking investment, or of instruments to be employed, involves a cost while the use of capital results in a

return out of which these costs must be defrayed. We have here another example of demand and supply which may be presented graphically as in Fig. 35. Along the line DD' is represented the declining significance of added units of capital which control the offers that borrowers can make for their use. The line SS' represents the increasing cost of waiting as the additional units of capital are made available. These two influences are represented in equilibrium at P' , which means that the return from the use of an additional unit of capital just balances the cost of securing it. The line PP' then represents the price paid for the services of capital, or it may be thought of as the yield of capital in use. On the side of demand, PP' represents the marginal productivity of capital instruments used in production and, hence, limits the borrowers' ability to pay interest. On the supply side, it measures the cost of waiting, or time preference of marginal savers and, thus, determines the amount of capital funds available. In this way, the interest rate is determined by the interplay of the forces of demand and supply, as expressed through the marginal productivity of capital on the one side, and the marginal cost of waiting on the other.

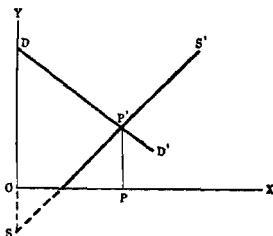


FIG. 35.

Since the rate is the resultant of these two forces, it may be thought of as being measured equally by the productivity of the marginal units of capital in use, or by the marginal cost of securing these marginal increments of the supply. This cost is determined, as should be apparent from the above discussion, by the time preference of marginal savers, *i.e.*, the amount necessary to induce these savers to postpone present consumption.

Since the rate is the resultant of these two forces, it may be thought of as being measured equally by the productivity of the marginal units of capital in use, or by the marginal cost of securing these marginal increments of the supply. This cost is determined, as should be apparent from the above discussion, by the time preference of marginal savers, *i.e.*, the amount necessary to induce these savers to postpone present consumption.

The Rate on Consumption Loans.—The distinction has already been made between economic interest and contractual or loan interest. The forces of demand and supply are expressed in the market by a rate for loanable funds, which is the method of borrowing in modern economic society. Savings are effected by means of money and credit instruments. Borrowers and lenders of loanable funds are brought into the money market,

and rates or prices are made. Not all borrowers are seeking loans for productive purposes. Many loans are made for purposes of consumption as individuals may borrow to pay for current expenses. Such loans in the past were called "spendthrift" loans. In modern times the advent of installment buying, which, after all, is a form of a loan for the purchase of consumers' goods, has greatly extended this field. One needs to cite only the importance of buying automobiles, furniture, musical instruments, and recently men's clothing on the installment plan, to call attention to the significance of borrowing for purposes of consumption. Most government loans are also of this character. The European governments have come to realize to their sorrow that the debts which they incurred in the conduct of the World War did not create a means of paying back the loans at maturity. This fact accounts for the embarrassment of so many of these countries at the present time. It is for this reason that some authorities have referred to the debt of a nation incurred for normal operations, or for the conduct of war, as "spendthrift" loans. However, so far as the immediate effect on the market rate for loanable funds is concerned, the demand for purposes of consumption is as potent as demand for productive purposes, but, in the long run, the rate will tend to conform to the earnings of capital employed in productive lines, because those who borrow for purposes of consumption have to pay as much as those who use their funds for productive purposes. In this way, the ability to pay becomes an important consideration in determining the market rates. Only those who are able to pay the cost of waiting will get loans and, therefore, the marginal productivity of capital becomes important in determining what the rate will be. In other words, the market rate, or contractual interest, tends to conform to the economic interest earned.

In the case of government loans, other things being equal, the rate will be lower than for ordinary industrial loans, because the risk elements are not normally so large. The source of income of most governments is likely to be more stable than the income of most business enterprises and, therefore, this fact will affect the actual rates paid on borrowed funds. This qualification is not absolutely true, as can be seen today in connection with bonds of many European countries. They have to pay higher rates than many industrial concerns of our country. Here the reverse situation is true, which emphasizes the fact that the risk

elements are always potent factors affecting the actual market rates.

The Quantity of Money and Interest Rates.—Since borrowing is done by means of money, many people get an erroneous impression of the relation of money to the determination of rates of interest. They assume that borrowers can get more funds and at lower rates if the supply of money is increased. This fallacy reappears on innumerable occasions, and an increase in the amount of money is frequently advanced as the most effective means of relieving a debtor class. By reference to the chapter on money and credit, the student will find that the only effect that an increase in the volume of money can have, other things remaining unchanged, is to enhance the general level of prices. The borrower does not want money, but what money will buy. An increase in the volume of money, without a corresponding increase in the quantity of goods, can have but one effect, and that is to give the borrower control over a smaller quantity of goods per unit of money.

The impression that the volume of money has an effect on interest rates is given some basis for acceptance in connection with bank discount rates. These rates generally fluctuate inversely with the volume of reserves in the banks. When the banks hold large idle reserves, it is said that money rates are easy and the discount rates tend to fall. When the reserves are falling off, discount rates tend to rise. It is evident, then, that rates for bank credit fluctuate with the supply of credit or the ability of banks to make loans. The fallacy of passing from this experience to the general conclusion concerning the quantity of money and interest rates is that the supply of money in the banks and the total supply in the country are different things. If the volume of money drawn from the banks and put into circulation is large, the banks may have to economize in lending their credit. To discourage borrowing, they raise the discount rates, which not only checks borrowing but puts a premium on the flow of funds to the banks. As these sums accumulate, they build up the reserves and tend to ease the discount rates. To this extent, credence can be given to the relation of the quantity of money and the money rates, which the student will understand are in this case bank discount rates. But should the country as a whole get a larger volume of money, as in connection with the payments of indebtedness by other countries, then the effect of

the additional quantity of money would be to enhance prices rather than to cause a change in the rates of interest.

The interest rate has been spoken of in the previous discussion as a uniform rate. Anyone familiar with the market for loanable funds knows that there is a great variety of loan, or contractual interest rates. The following examples, taken from the *New York Journal of Commerce* for Sept. 8, 1928, will illustrate the variations:

Short-time:

Call loans.....	7½ per cent
Commercial paper.....	5¼ to 5¾ per cent

Long-time:

U. S. Treasury 4¼ per cent certificates, due December 1928, quoted to yield 4.48 per cent.

Brazil.....	8s	market price.....	108¾
French Republic.....	7½s	market price.....	114¾
German External.....	7s	market price.....	106½
Italy Kingdom.....	7s	market price.....	97¾
Japan.....	6½s	market price.....	101¾
Sweden Kingdom.....	5½s	market price.....	103¾
Canadian National Railroad '57	4½s	market price.....	97¾
Canadian Pacific Railroad debentures.....	4s	market price.....	87

The evidence is conclusive that, instead of one rate, there are many rates at any one time for the control of loanable funds. This variation in the contractual rates is explainable by different causes. It is due partly to the nature of the banking business, as a bank must have its investments in a form that will enable it to meet its day-to-day demands for cash, and on this account it will usually accept lower rates for short-time than for long-time loans. But short-time rates, especially call-loan rates, are subject to the most violent fluctuations, as can be discovered by examining quotations during a period of time. The rate may be 2 per cent or less on one date and within a short period run up to 10 per cent or higher, as the pressure for funds causes fluctuations of this character. Long-time loan rates are more stable and subject to much less violent fluctuations.

Another cause for variation is the risk element that is involved in connection with a loan. The risk element is present in all

loans, but is much greater in some than in others. This can be seen in the rates of both industrial and government bonds. It is especially noticeable in connection with the foreign bonds, listed above, many of which were issued during the World War, when the uncertainty of the outcome and the unsettled financial conditions in these countries made the risk element large. In some cases, the actual yield is higher than the nominal rate as the bonds can be bought below par, which has the effect of increasing the return to the purchaser. In other cases, the market price has increased above par, which reduces the yield below the nominal rate. The difference in the rates due to the risk element is not interest in the sense of the above discussion. It is more of the nature of insurance or a price paid for carrying the risk involved.

But even though the risk element be eliminated, there still would be different rates of return to capital. Capital used in some ways is more productive than in others. This difference in productivity will enable some borrowers to pay more than others, because they receive a larger return from the use of the capital. While this condition of varying returns to capital is persistent in actual business operations, nevertheless, the effect of competition is to render the return equal on all productive capital equally situated as to the risk element. The reader should recall the general assumptions underlying the development of our principles and interpret the conclusions concerning interest in the light of these assumptions. So far as competition is operating, there is a tendency for the establishment of a uniform rate of interest on all capital similarly situated in respect to risk.

The Interest Rate and Capitalization.—We have repeatedly referred to capitalization of income as a method of determining the value of capital instruments and land. The rate chosen in the process of capitalization has a very important bearing on the capital value of any productive instrument. It may be laid down as a general proposition that the only present measure of capital value of any productive instrument is the capitalized value of the income from that instrument. We may take land as an instance. Suppose we have satisfied ourselves that the rent per acre that can be annually expected from land is worth \$6. What is the land itself worth? We say that by capitalizing the income we can find the capital value. But what rate shall we use? If we take 6 per cent, the land is worth \$100 per acre, but

if we take 4 per cent, it is worth \$150 per acre. The importance of the rate chosen in determining the capital value of land immediately becomes evident.

But suppose, some one argues, that the owner has paid \$500 per acre for the land and, therefore, he should get 6 per cent return on this capital sum. Such an approach is incorrect. The value of land, or of a capital instrument, should always be approached through its earnings. No capital instrument is worth more than the capitalized value of the income from that instrument no matter what it may have cost to produce or to install it. In the case of capital instruments, however, the capitalized value of the income yielded by the use of the instrument cannot depart very far from its reproduction cost. Put in another way, if the capitalized value of the income exceeds the cost of producing such an instrument, more instruments will be produced and, under the influence of the principle of diminishing productivity, the yield per instrument will tend to decline. This decline in income earned by the instrument will cause a fall in its capital value. For this reason it is correct to say that the value of capital instruments cannot greatly exceed the cost of reproducing them, while the value of land may increase indefinitely with the growth in demand for its uses.

The rate that is actually chosen for purposes of capitalization will depend on two forces, namely, the risk element present in the use of the productive agent, and the time preference of the buyer of the agent. The greater the element of risk involved, the higher the rate that will be chosen for purposes of capitalization. The effect of a high rate will be to lower the capital value of the agent. In the second place, the time preference of some persons is greater than that of others. Those with a low time preference would choose a low rate for purposes of capitalization, while those with a high time preference, would require a high rate of return to induce them to postpone present consumption. Hence, a variation in the rates chosen can be accounted for largely by the variation in time preference, or degree of prospectiveness of different persons, and this fact materially affects the capital sums offered for either land or capital instruments.

Interest as a Cost.—One other topic should receive some attention because of its bearing on private business operations namely, the question whether interest constitutes a cost of production. There has been more or less confusion over this

issue, largely due to differences in the usages of terms in accounting and economics. The accountants frequently distinguish between cost of production and a financial expense. A cost is an outlay, or direct money expenditure, that has to be paid by the business in connection with the actual operation, whereas a financial expense is one connected with getting funds, or money, for use in the business. According to this usage, the payments on capital sums borrowed are classed as a cost because there is a direct payment to be met, but interest earned upon capital furnished by the owners of the business is not a cost, although it might be classed as a financial expense.

A misunderstanding of this classification has led some to argue that charging interest on capital owned as a cost would put a business in an unfavorable position as a competitor. Such an argument, of course, overlooks the fact that the owner expects normal returns on his own, as well as on the borrowed, capital and furthermore, it fails to recognize the opportunity for gain which the employment of his own capital offers in other directions. Unless his capital yields an income that is covered by the price of the good, he is clearly a loser, for by employing it, he foregoes the return that he could have received had he lent it to others. Viewed in this way, it seems evident that interest should be regarded as a necessary part of the cost of producing a good. If it is charged as an expense, it will accrue to the owner of the business as earnings or interest on the capital that he has invested, but if it is regarded as a financial expense, it will appear as a profit item. The purposes involved in accountancy and economics are not identical, but they are not contradictory. For practical purposes in keeping the records of a business, the treatment of interest as a financial expense is ordinarily justified. If it becomes necessary to determine all of the costs that are assignable against the production of a good, as when a utility company is demanding a fair rate before a state commission, then interest on all capital will be charged as a cost. This will be done, even though the accounting records will carry the item as a financial expense; it will be separated and charged against the rendering of the service. Here the two usages approximate each other.

It is also argued that if interest on capital owned is charged as an expense, it will have to be charged also as an earning, or income, and that the same result may be obtained by treating this return as a part of the profits of the business without setting

this item up on both sides of the ledger. It is evident that the confusion is one of usage of terms. To the economist, cost of production includes all expenses that can be assigned against the production of a good, *i.e.*, all expenses that have to be incurred in order to bring the good into existence. In this sense, interest is a cost for, if capital, whether owned or borrowed, received no return, there would be comparatively little waiting and no inducement to utilize any reserves that might be accumulated by those who saved. Consequently, there would be few or no capital instruments produced.

Interest in this sense, therefore, is as distinctly an item in the cost of producing a good, as is the outlay for wages, or for the materials on which the labor is expended. The case in connection with borrowed capital is clear enough, but when capital is furnished by the owners of a business it may be said that, unless this capital earns a return at least equal to that received by loaned capital, there is no inducement for the owners to operate and use their own capital funds in production. Hence, such capital must receive a return as an inducement for its employment, and this return should be regarded as a necessary cost of production.

Suggested References

- BROWN, HARRY G., "Common Sense and the Common Welfare," Part II, Chaps. III and IV.
- BYE, RAYMOND T., "Principles of Economics," Chap. XVIII.
- CASSELL, GUSTAV, "The Nature and Necessity of Interest."
- CARVER, THOMAS N., "The Distribution of Wealth," Chap. VI.
- CLARK, J. B., "The Distribution of Wealth," Chaps. IX, X, and XI.
- DAVENPORT, H. J., "Economics of Enterprise," Chaps. XVIII and XXI.
- EDIE, LIONEL D., "Economics, Principles and Problems," Chaps. XIV-XVI.
- ELY, RICHARD T., "Outlines of Economics," 4th Ed., Chap. XXV.
- FAIRCHILD, FURNISS, and BUCK, "Elementary Economics," Vol. II, Chaps. XXXVI and XXXVII.
- FETTER, FRANK A., "Economic Principles," Chaps. XX-XXV.
- , "Interest Theory and Price Movements," *Am. Econ. Rev.*, Vol. XVII, Sup., pp. 62-122.
- FRIDAY, DAVID, "Profits, Wages and Prices," Chap. V.
- FISHER, IRVING, "The Rate of Interest."
- , "The Nature of Capital and Income."
- GONNER, E. C. K., "Interest and Saving."
- HOBSON, J. A., "Economics of Distribution," Chaps. V, VIII, IX, XXIII, and XXIV.
- KLEENE, G. A., "Profit and Wages," Chaps. I-VI.

RUFENER, L. A., "Principles of Economics," Chap. XXIII and XXIV.

SEAGER, HENRY, "Principles of Economics," Rev. Ed., Chap. XVI.

TAUSSIG, FRANK W., "Principles of Economics," 3d Ed., Vol. II, Chaps. XXXVIII, XL, and XLVI.

WICKSTED, P. H., "The Common Sense of Political Economy," Book, I, Chap. VII.

CHAPTER XXI

WAGES

Definition of Wages.—As a functional share, wages are those portions of the income from industry that are assigned to labor, or, put in another way, wages are a payment for the services of labor. We have previously defined labor as all human effort, whether mental or physical, that is used in production. Hence, when we speak of labor as a factor in production, we mean not only the manual labor of the men in the workshop who work with the tools and the machines of industry, but we include as well the men in the office who handle the financial and accounting records of the business, and those who lay out the plans and direct and manage its various phases. Labor in this sense, then, includes all productive activity of the individual, as contrasted with the services of land and that of waiting involved in capital formation. The definition of wages as a payment for the services of labor, therefore, is intended to include all forms of compensation that go to labor used in production. It will be noted at once that this use of the term is more inclusive than the customary usage, since wages are usually thought of as the payments made to manual laborers. But as a problem in distribution the meaning of the term must be enlarged as indicated above.

Statement of the Problem.—Starting with this definition of wages, there are two problems that immediately confront us. How can we measure the services of labor, and how can we determine the value of these services? Two general observations can be made at once. First, the services of labor are usually expressed through some good or product, which means that their value or social significance is derived from the value of this product. The only exception to this general statement is found in connection with personal services, such as that of a barber, or a masseur, or those whose services yield satisfactions directly. In all other instances, labor services are measured and evaluated in connection with the goods produced. In the second place, there are great differences in the capacities of men to render

services, whether they are rendered directly or indirectly. Consequently, we can not speak, even theoretically, of a general rate of wages applicable to all kinds of labor.

Any attempt to measure the services of labor has to take account of the very great differences in the productive capacities of different individuals. Some men are capable of doing types of work which other men are quite unable to perform. This is especially true of brain workers. There are some who have great organizing and executive ability, or the capacity for discovery and invention, while others are capable of performing only the routine tasks of industry. These unusual capacities of one kind or another that enable some to perform tasks that others cannot do cause variations both in the amount and importance of the services performed. Such differences will be reflected in the compensation received, and are a part of our problem.

Then there are types of work of varying social significance. Some kinds of labor contribute very little to human satisfactions. This may be due to the fact that there are many people who can make the good or render the service; or it may be an insignificant task in the production of an important good, as the work of an office boy, or that of the gatekeeper at a factory. These great differences in the capacities of men and in the kinds of labor performed in modern industry prevent the compensation received from being reduced to a general rate of wages. Some men may get a salary of \$100,000 per year, while others may get only a few hundred dollars. Hence, we cannot speak of a general rate of wages, meaning thereby, a rate applicable to all forms of labor services. The compensation of prize fighters, movie actors, professional men, and unskilled workers cannot be reduced to a common rate. But our wage theory should furnish a basis of explaining all types of remuneration received as compensation for labor services.

Non-competing Groups.—In attempting to measure the services of labor, recognition must be made of the great variety or differences in the types of services to be found in modern industry. A little observation will convince one that the compensation received by one kind of labor is not directly affected by that received by another. The income of the movie actor, mentioned above, has little or nothing to do in determining the pay of a carpenter or the salary of a bank president. The

recognition of this fact has led to the classification of labor into what is known as "non-competing groups." There are of course no hard and fast lines of classification of occupations. In a country like India, where the caste system prevails, the classification is vigorously maintained by long-established social and religious traditions, but in countries of western civilization, and especially in the United States, there is comparative freedom for the individual to pass from one occupational group to another. It is possible, in other words, for a man who has worked as a section hand to become the president of a railroad. There have been many notable instances, particularly in this country, of men who have risen from the lowest ranks of labor to the most highly paid and influential executive positions in modern business. In fact, public sentiment here emulates and exalts freedom of opportunity for the individual that is exemplified by such illustrations.

Nevertheless, while individuals may pass from occupation to occupation with comparative freedom, the fact remains that the pay of the section hand has little to do with that of the railroad president, and *vice versa*. This means, of course, that the existence of non-competing groups is an important influence to be reckoned with in explaining both the method of measuring labor services and the determination of their social significance, or compensation.

If all labor were homogeneous in character and the choice of occupations perfectly free, there would be a gradual gradation of wages from the lowest to the highest. Under these circumstances, any variation in wages would just equalize the difference attached to the goods produced and just offset the differences in attractiveness, or agreeableness, of the separate occupations. As between two occupations that differ in attractiveness, the one which is least attractive would have to offer a compensation that would offset the disagreeableness. But even under these circumstances we could not speak of a general rate of wages, except for those who performed like tasks. However, labor cannot move with perfect freedom from one occupation to another, not only because of differences in native capacity of individuals, but because laborers tend to become specialized and, therefore, incapable of free movement into other lines of work. A skilled watch maker is not likely to become a lawyer. Those who have spent long years of patient study to become lawyers, or chemists,

or engineers cannot readily pass into another trade or profession which requires skill or training different from that which they now possess. Then, too, there are many artificial obstacles to the free movement from occupation to occupation. It may be trade union rules in one case, or it may be laws regulating the practice of medicine, law, or accountancy in another. Whatever the cause, it is perfectly evident that the free movement from occupation to occupation does not exist and, therefore, some classification is necessary and greatly affects the explanation of wages as a share of the social income.

The number of groups of occupations, or lines of productive activity, is not so significant as the fact that some grouping is possible, and that the movement from group to group is limited. It cannot be said that there is a gradual gradation of labor from the lowest to the highest, nor can it be said that there is a definite number of classes into which all labor can be placed. Any classification that is attempted will show some overlapping; that is to say, there will be some members of the various groups who are capable of passing from one group to another whenever the compensation is more attractive. Many instances of this character can be found. Frequently, lawyers, engineers, and accountants become business executives. Similar transferring can be found among the skilled trades, and not infrequently workmen pass from the position of mechanic to that of employer or contractor. Enough evidence has been cited in support of the contention that there are occupational groups and that the existence of these groups must be taken into account in explaining wages. The effect of the overlapping of groups on wages is to prevent the compensation received by the members of any one group from departing far from the amounts necessary to equalize the costs involved in passing from one group to another.

Diminishing Productivity as Applied to Labor.—Because of the differences noted above, and for purposes of developing our wage theory, let us now assume that we are considering the services of a homogeneous group of workers, as the unskilled or common laborers. At the outset in explaining the method of measuring the services of labor, we must recognize that the principle of diminishing productivity applies to labor, just as to land and to capital. Assuming that this labor is equipped with a definite amount of land and capital, a definite product will be obtained. By experimentation some combination of labor with these other

factors will be found that will yield the maximum amount per laborer. By increasing the amount of labor beyond this point, the equipment of land and capital remaining unchanged, the product per laborer will diminish. This decline may be relative or it may be absolute. The amount produced per man will decline before there is an absolute reduction in the units of goods produced, but if the supply of labor is increased sufficiently, production will eventually decline both absolutely and relatively.

While, theoretically, the increase in the number of workmen may be sufficient to cause an absolute decline in production because it is conceivable that the workmen could become so numerous as to get into each other's way, practically, this would never occur, for the entrepreneur would stop hiring men long before this point had been reached. In addition to the decline in physical units produced, the social significance of each unit is likely to fall with an increasing volume of production. That is, if the supply of any good be increased without a corresponding increase in demand for it, there will be a diminishing significance attached to units of that good. If the value of the good declines, it follows that there will be a corresponding decline in the social significance of the labor services employed in the production of the good.

The above principle can be set forth by the following graph:

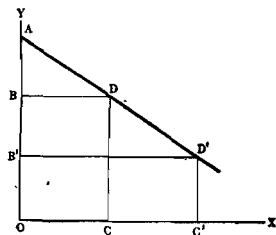


FIG. 36.

Let us measure along the line OX the number of laborers, and along the line OY the productivity per laborer employed upon a given area of land with a definite equipment of capital. If the number of laborers be represented by OC on OX , then the productivity per man would be represented by the line CD . This is

so because by assumption each laborer is technically as efficient as every other, so that the productivity of one is just as important as that of any other so long as they are engaged in producing the same commodity. It follows that the productivity of any laborer in the group is determined by the productivity of the marginal man. If more laborers were added, as indicated by CC' , the other factors remaining unchanged, then the marginal productivity per man would fall from CD to $C'D'$. Hence, we may conclude that the services of a group of homogeneous laborers employed in making identical goods are measured by the services of the marginal man.

The practical way of testing this method of measuring the productivity of labor services would be by setting an additional laborer at work with the existing equipment and existing working force, and by observing the results in output. If the product were increased, the addition could be attributed to the added worker, since his efforts constitute the only change in the conditions of operation.

While the principle of diminishing marginal productivity explains the productivity of the labor services of a group of workmen in terms of physical units, it is not a complete explanation of wages because, as indicated above, the value or social significance of the goods produced is equally important in explaining the claim of these workers for a share of income. If this statement be put in terms of the business manager's experience, it would take some such form as follows: The manager of any enterprise can pay the marginal man a wage not solely because he adds to the physical supply of goods, but because the additional goods have value. The amount he can pay depends on the value of the goods.

It, therefore, becomes necessary to apply the value principle to these services. As the supply of goods of any kind increases without a corresponding increase in the demand for them, their value will decline. This follows from the principle of diminishing utility. As the significance of the goods declines the importance of the labor services used in their manufacture will likewise diminish to the point of marginal use.

This notion can be presented graphically in Fig. 37 by a slight change in the relations shown in Fig. 36. Along the line OX is measured the quantity of labor services of a particular kind. Along the line OY is measured the social importance of a unit of

these services, whether this unit is expressed by a duration of time or by a quantity of products. The line AD represents the declining social significance of the added units of labor services. It will be observed that this line represents the resultant of the two influences—the physical productivity of the labor and the diminishing value per unit of the good. The margin of use of the labor services of the group of workers under consideration

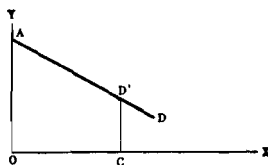


FIG. 37.

is determined by the demand for their products. Let us assume that the demand carries the use to point C on OX , then the value of the marginal unit is represented by the line CD' . Since, by assumption, the laborers are alike in efficiency, the significance or value of the whole

supply is determined by the significance attached to the marginal unit. We conclude, then, that the worth of the services of a definite quantity of labor of the same grade is determined by the value of the services of the marginal man.

The only qualification that is necessary to make this conclusion apply to labor services in the form of personal services is that these services are evaluated directly rather than indirectly through a product. The offers by marginal users for the marginal services will determine the value of services of this character. It will be observed that the addition of more laborers after a certain point has been reached will cause the value of labor services to fall, partly on account of a decline in the physical productivity of the labor and partly because of a declining social significance of the additional units of the goods produced. The net effect, then, is that increasing the number of workmen of a given group lowers the productivity of every member of the group which tends to coincide with the productivity of the marginal man.

This is the application of the marginal productivity theory to labor. As set forth in the preceding paragraphs the theory was discussed in connection with a single group of homogeneous laborers. The reasoning employed with this group, however, applies with equal force to every other group, whatever may be the classification of labor. It follows from this, that the productivity of any group of laborers is determined by the marginal

productivity principle. The social significance of the services of any laborer in any group tends to conform to the services of the marginal laborer of that group. Since, by assumption, all of the laborers in a group are alike in technical efficiency, then any one of them may be regarded as the marginal laborer, whose productivity will vary with the numbers in the group.

Economic and Contractual Wage.—The marginal product, as thus defined, constitutes the maximum claim that any laborer can make on the income from industry. This amount is what is known as the *economic* or *imputed wage*. It is the full value of the services of labor, as derived from the goods produced and attributable to labor. It is the functional share assigned to labor, as contrasted with rent and interest that are assigned to land and capital, respectively. As in the case of the other shares, it is necessary to distinguish between the economic and the contractual wage. The bargain between the employer and the employee is over the control of the services of labor. The employer is interested usually in the product that results from the use of labor services. What he can pay for these services depends on the importance attached to this product by the users of the good, *i.e.*, the contractual wage cannot exceed in value what the laborer has added to the product. If it does, then the other shares will suffer. The value added by any laborer will coincide with the marginal productivity of labor of that particular grade. Hence, we can conclude that the contractual wage cannot long exceed the economic wage and, under conditions of perfect competition and mobility of labor, will tend to equal it.

Application of the Principle.—This principle, it should be remembered, applies to all grades of labor. The services of labor are not worth more than the value that is added to the products of industry, nor can the laborer get more than this amount for any long period of time. It may be asked, How can this sum be practically determined? This problem is complicated by the fact that goods are produced in anticipation of demand and, hence, the derived value of labor services involves an uncertain element. Since the responsibility for the wise use of labor services devolves upon the entrepreneur, we can assume that he will continue to add laborers to his force as long as he finds it profitable to do so or, put in other words, as long as he thinks the value added by the laborer is equal to, or exceeds, the money wages paid. Adding a man and observing the effect

on the output, or *vice versa*, is a fairly accurate method of measuring the contribution of labor of a given grade. The testing of the contribution of labor in accordance with this principle is only another way of saying that the workmen are being paid on the basis of their marginal productivity.

Wages of Management.—In the development of this theory, the discussion has implied that the laborer sold his services in a labor market. This is, of course, the customary form in which wages appear, and some authorities confine wages to the payments made to hired laborers. But in this text, the term “wages” is intended to include the compensation for labor services, whether these services are sold to employers, or whether the share is combined with that of other shares into a common income, as in the case of the proprietor of a small business. The distinction that is being made applies not only to those who sell their labor services, but also to those who furnish labor along with the ownership of capital, or land, or both, in the operation of a business. A farmer may own the land and tools with which he cultivates his crops, and at the same time perform all the work on the farm. In this case and in others like it, the entrepreneur will get a combined income made up of rent, wages, and interest. He may not be able to separate one source of income from another and determine how much each source yields him. The owner-manager of a business has to confront this problem. He may be operating on leased land and borrow a large part of his capital funds, and devote his own time and energy to directing the business. Frequently, a business man of this character does not reckon a payment for his labor at the full market value of his services but will take his income in what he calls profits. This is an incorrect method of calculating profit or of arriving at the costs of producing a good. The wages of management at the full market value of the services rendered should be counted as a part of the cost of producing a good.

The value of these services can be approximated in one of two ways. First, one may figure what the manager could get if he sold his services to an employer or, second, he could charge the industry with a sum that would be required to hire someone else to operate the business just as effectively as he would run it himself. It will be noted that both of these tests rely upon the market as a means of determining the importance of the services of the owner-manager. This method is essentially accurate,

though it does not take account of instances of managers whose ability is distinctly superior to the average. Differences of ability of this character are comparable with differences in the productivity of land, and the extra income which arises from this cause may be regarded as *rent of ability*.

Unless the wages of management are separated from other costs, true costs of producing a good cannot be found, nor can profits of a business be accurately computed. If wages of management are counted as part of profits, and the business is sold on the basis of the capitalized value of its income, including profits thus calculated, part of the wages of the present manager would enter into the sum capitalized and thus give an incorrect figure for the present value of the business. For these reasons, wages of management should be counted as compensation for labor services, the value of which, it will be seen, conforms in the main to the marginal productivity principle formulated above.

The Demand for Labor.—The principle just formulated is an explanation of wages as a functional share of the social income. The actual share received by any workman is, as has been indicated, the result of a contract between a laborer and an employer. From an acquisitive point of view, the employer controls the demand for labor. He desires labor because the services rendered are essential for all forms of production. He will continue to hire laborers just as long as the value added by the use of their services equals or exceeds the money wages paid for them. Demand for labor, in this sense, is the quantity of labor services taken by employers at a rate of wages per grade of labor. There is, however, another point of view concerning the nature of the demand for labor that should receive some attention. But before this topic can be treated adequately, it will be necessary to distinguish between money and real wages.

Laborers, as well as all other persons, get their income in the form of money wages. They are paid so much per day, per hour, or per piece. But the amount of money received does not constitute their real wages, or real income. Here, as elsewhere, we have to distinguish between money income and what money will buy. The fundamental thing here is the supply of consumable goods available for the gratification of human wants. Real wages, then, constitute the amount of goods that money wages will buy. As consumers, laborers are interested in the quantity of consum-

able goods which they can command with their wages. The larger this quantity, the higher will be their standard of living.

We may now return to a consideration of the demand for labor from the social point of view. We have already emphasized the significance of the division of labor as a phase of production, but this subject has an important bearing also on the topic under discussion. One group of workers produces a commodity which is traded in the market for some commodity or commodities that are produced by other laborers. From this point of view, the demand for the services of any type of labor arises from an offer of goods produced by other laborers. Demand, in this sense, consists of the stock of goods produced for sale, so that demand may be increased by increasing the supply of goods offered for sale on the market. Continuity of the productive processes depends upon a delicate balance between production and consumption. The responsibility for maintaining that balance rests in first instance with entrepreneurs. When the productive energies of a nation are wisely directed, it is correct to say that the demand for labor, socially considered, arises from the production of a large volume of goods available for exchange. The more workers there are producing goods that people want, the greater will be the demand for labor. The difficulty in maintaining this situation continuously is the inability of entrepreneurs to forecast accurately the course and rate of consumption. As soon as a mistake has been made, and the balance between production and consumption unsettled, some entrepreneurs begin to slow down or stop entirely their production. This influence radiates in all directions and, if on a sufficiently large scale, will paralyze industry and bring on an industrial depression. In this sense, demand for labor is based upon the continuity of production in all lines. Business men generally are affected in their private operations by these fundamental economic forces.

Supply of Labor.—Up to this point, we have discussed the wage theory primarily from the point of view of demand for labor and its use in production. While it was pointed out in the previous treatment that the marginal productivity of labor was greatly affected by the number of workmen in any group, no attention was devoted to the causes of supply or the influence affecting the same. If the general principles affecting supply applied to labor, an equilibrium would tend to be established

between the wages offered and the expenses, or costs of production of labor. It is evident that the supply of labor presents certain peculiarities that do not conform to the general principle of supply. As between different groups of workers, however, supply does tend to conform to the general principles governing supply, namely, that the supply of labor will increase as the price offered (in this case wages) is increased. This conclusion should be interpreted in the light of what has already been said concerning the difficulties that limit the free movement of labor from occupation to occupation. Nevertheless, a higher wage for any group of workers will not only cause an immediate shifting of some labor but, if continued, will permanently enlarge the supply of this class of labor.

Total Population and Supply.—The general supply of labor is not so elastic as that between groups of workers. In the first place it must be recognized that supply and mere numbers, or total population, are not synonymous. In the United States, in 1920, four-fifths of those engaged in gainful occupations were males. The proportion of males in the total population is greater in the United States than in most European countries, and greater in some parts of this country than in others. Then, the proportion of the population between the ages of fifteen and sixty-five is a fact that has a bearing upon the supply of labor. It should be evident from these statements that the general structure of the population is an important influence affecting the supply of labor.¹ There is at all times an elastic element in the supply of labor that is independent of the structure of population. There are always some idle men who could be added to the forces of labor, if opportunities were found, or the wages were adequate. A much larger number of women might work in industry, if the demand were sufficiently urgent. The experience during the World War demonstrated that there was elasticity in the supply of labor and that the elastic elements came from both males and females who are not normally employed, and also from a more effective utilization of the existing supply. But in normal times the supply is fairly inelastic. It depends upon the natural increase in population, which requires an interval of approximately fifteen years before the child can enter productive work.

Growth of Population and Supply.—While supply of labor and mere numbers of people, as has just been shown, are not synony-

¹ Ely, R. T., "Outlines of Economics," 4th Ed., p. 428.

mous, yet, from the long-time point of view, the growth of population is the most important influence affecting supply. This question leads into the discussion of population and the causes for its growth and decline—a topic that can only be briefly treated in connection with the subject of wages. In his essay on "Population" (1798), Malthus pointed out that there was a very close relation between the numbers of people and the supply of foodstuffs. He showed that it was possible for the human family to increase faster than the increase in food supply and, unless checked by such causes as wars, famine, pestilence, or prudential checks, as delayed marriage or birth control, population would increase to such an extent as to cause extreme poverty.

The Malthusian theory of population was hotly debated during the nineteenth century, and many of its critics asserted that the theory did not explain population growth and, in fact, was contrary to historical experience. In support of their arguments, they cited the high birth rate among the poorer classes in society as compared with that of the well-to-do. The better judgment of present-day writers seems inclined toward the acceptance of the essential truth of the doctrine as enunciated by Malthus. In support of the latter position, the following figures are given: In 1630 it is estimated that England had about 5,500,000 people; in 1730, about 6,200,000; in 1761, just prior to the Industrial Revolution, 6,700,000; in 1831, after the factory system was well established, the population was about 14,000,000; and in 1921, after the losses entailed by the World War, the population was 38,000,000. It will be noted that the great growth from 6,700,000 in 1761 to 38,000,000 in 1921 is coincident with the period in which the productive capacity of industry made such phenomenal strides.

The same phenomenal growth of population is found in Europe during this period also. The population in 1760 was approximately 130,000,000 while in 1920, after the most severe war losses of history, the numbers were about 450,000,000. Of this increase, 200,000,000 have been added since 1820 and 150,000,000 since 1872.¹ Evidence of this character is fairly conclusive that the increase of population has depended very closely upon the increase in the supply of goods. It would seem from

¹ ELY, R. T., "Outlines of Economics," 4th Ed., pp. 430-431, quoted from W. F. WILLCOX, "The Expansion of Europe in Population," *Am. Econ. Rev.* Vol. V, p. 749.

these facts that, in the long run, population figures conform to the general supply forces. That the amount of real income available for the support of a people is an active influence affecting the numbers. But, for our purposes in explaining wages, while recognizing the significance of an increase in population as the fundamental source of the supply of labor, we must consider labor as much less elastic than the supply of commodities. The demand influences, therefore, are more potent in determining the value of labor services.

Immigration and Supply.—The previous discussion was concerned with the natural growth of population. New countries may have an elastic supply element from immigration. From 1900 to 1914, the United States received approximately one million immigrants per year. The proportion of these immigrants who were males between the ages of fifteen and sixty-five was greater than the sex and age distribution of the entire population of our country. Other things remaining equal, the effect of the migration of these peoples would be to lower the marginal productivity of labor in this country and to raise it in their native countries. How far the tendency to raise the productivity in the home country has been counteracted by increase in births cannot be told, but the two influences have an opposite effect on the marginal productivity of labor. There can be no doubt that the demand for a limitation of immigration in this country, which has been made effective since the World War, was, consciously or unconsciously, a recognition of this principle. As an influence on wages, our conclusion is that an increase in supply, from whatever source, tends to lower the marginal productivity of labor and, therefore, the economic wage.

Subsistence Theory of Wages.—The theory of wages would not be complete without some consideration of the *subsistence theory* that was advanced by the English economists during the first quarter of the nineteenth century. This theory was a corollary of the Malthusian doctrine of population. Ricardo put the theory as follows: "The natural price of labor is that price which is necessary to enable the laborers, one with another, to subsist and to perpetuate their race, without either increase or diminution." It was argued that, if wages were increased much above the subsistence level, the birth rate would be increased and the increase in numbers would bring the wages back to the subsistence level. While there was much in the conditions pre-

vailing in England at the time the doctrine was formulated to confirm it, subsequent history has shown that the earlier writers did not give enough weight to the influence of the prudential checks on the increase in population. Only the most pessimistic social reformer could argue that the economic conditions of the working classes did not improve during the nineteenth century. The capacity to produce has outrun the increase in population during the last hundred years, with the result that the workingmen generally have a higher standard of living than their forefathers. As a general statement this is certainly true, even though there are still many instances of extreme poverty in all countries.

Wages-Fund Theory.—The subsistence theory did not long continue as a satisfactory explanation of wages. It was superseded by the "wages-fund" theory, which was accepted during the second decade of the nineteenth century and continued as the chief explanation of wages until after the middle of the century. Since production of most commodities involved a waiting period, it was argued that the employer had to make an advance out of capital to the laborer and, therefore, wages were paid out of capital. Wages, then, depended upon the "proportion between population and capital." At this same time, it was held that wages and profits, profits being used in the sense of a return upon capital, were complementary shares. If wages were high, profits must be low, consequently, if the proportion between population and capital remained unchanged, the only way by which wages could be increased in any specific industry was by a reduction in some other industry. The proof of this assertion rested upon the following considerations: Since wages and profits were complementary shares, an increase in wages would mean lower profits. A fall in profits in an industry would drive capital out of that line of investment on the one hand, and an increase in wages would attract an additional supply of labor on the other. The competition among the laborers would cause the wages to fall. But, if wages generally were higher, there would follow an increase in population and a consequent lowering of wages as the result of the competition of the larger number of laborers.

It will be observed that this theory rested upon the acceptance of the Malthusian principle of population and on an erroneous conception of the relation of wages and profits (interest). The wages of labor may actually increase at the same time that the

earnings of capital are increasing, if both labor and capital are increasing in efficiency. This general theory confused wage payments with unit costs. High wages may be associated with low unit costs, if the labor used is very efficient. A truer relation than that expressed in the wages-fund theory is that capital and labor are competing goods. If labor costs are high, i.e., if the wages paid per unit of time are high, as compared with the value of the service rendered, then capital is likely to be substituted. At least this condition of unit costs due to wage payments puts a premium on the use of capital instruments. The weakness of the wages-fund theory directed attention to the influence of efficiency of labor as a cause of wages. The recognition of this influence was an important element in the shift in the wages doctrine to the marginal productivity theory.

Wages as Advances.—Some authorities have argued that wages are advances to the laborers; that because of the time element involved in the round-about processes of production, the employer pays the laborers weekly, or bi-weekly in advance, and then often has to wait to secure a return of these advances from the sale of the products. Since the employer will not assume this waiting unless he receives compensation for it, therefore, the true wage is the value of the marginal product, discounted by the going rates of interest for the period the employer has to wait for the sale of the product. This is called the "discounted marginal productivity" theory of wages. The discount, calculated as indicated, is the amount which the employer receives as compensation for the waiting involved in his advance of wages to the workman.¹

In the acceptance or the rejection of this theory, two questions should receive consideration. First, are wage payments advances to labor? Second, should the discount notion be a part of the wage theory? While it is true that employers pay wages weekly, or bi-weekly, and the longer the period of time elapsing between the beginning of production and the sale of the product, the more working capital in the form of money the employer has to have available, does it follow from this that these payments are advances in the true meaning of this term? On first thought it would seem that these payments are advances, but further

¹ This theory was advanced by PROFESSOR TAUSSIG in his "Principles," and has recently been accepted by Professors FAIRCHILD, FURNISS, and BUCK, in their "Elementary Economics," 1926.

consideration will convince one that in reality, in making these payments, there has been an exchange of present values. The employee secures value in the form of money wages in exchange for the value embodied in the partly manufactured goods that have resulted from the use of his services.¹ From this point of view the real advance has been made by the laborer, for during the interval of the payroll period he has contributed his services daily under the direction of the entrepreneur in producing a good that is to be subsequently sold. The payment which he receives at the end of the payroll period is an exchange of money value for the services embodied in the partly manufactured goods, which, by the general organization of industry, belong to the entrepreneur. During the interval of the payroll period, therefore, the only advance that is being made, as between the employer and the employee, is the contribution of the employee to the employer in the form of labor services used in the processes of production.

In the determination of the value of these services, a time element is involved and must, therefore, be considered. It should be remembered that the value of any intermediate good, whether it be a producer's good or a partly finished, consumer's good, is derived from the value of the finished, consumer's good. In the process of evaluating the present worth of a future good, a discount element is involved. Put concretely, the present worth of a dollar to be received a year from today is less than the worth of a dollar today, by whatever amount that is essential to induce marginal waiters to postpone present consumption. If 6 cents is an accurate measure of the time preference of marginal waiters, then the present worth of \$1 to be received a year hence is 94 cents.² This fact affects the present worth of any good that requires time to produce it. The longer the time element involved in its production, the greater will be the discount. Consequently, in the determination of the present worth of the partly manufactured goods which embody the services of labor at the end of any payroll period, this time element is a factor and must be considered. Assuming wise direction of the productive

¹ This argument assumes wise entrepreneurship, for on no other assumption can we expect a good to have value at a future date sufficient to cover all of the costs of production.

² The student is cautioned not to speak of this as 6 per cent. It is evident that there is a difference in percentum between using 94 as a base with 6 as percentage, and that of using 100 as the base.

processes, the present value of these goods is derived from the estimated future value by a discount that just measures the time preference, as indicated above. For our purposes, we may assume that 6 per cent is an accurate measure of this time preference. Then the estimated future value should be discounted by 6 per cent to obtain the present value of the partly manufactured goods.

In this sense, the discount notion is essential in determining the value of all goods in the production of which a time element is involved. But does the recognition of the time element as a factor affecting the evaluating process require the introduction of the discount notion in the wage theory? There seem to be three valid reasons for rejecting this notion as an integral part of the theory of wages. First, its introduction involves a duplication of the discount notion. As just explained, the discount notion is essential in determining the present value of all intermediate goods. Since the present value of the labor services embodied in making these goods is derived from the value of the goods themselves, it follows that the discount notion has already entered into the calculation and, therefore, should not be included a second time in the explanation of wages. The marginal productivity theory, advanced above, has recognized that the value of labor services is derived from the value of the products of industry and is, therefore, affected by all of the influences that in any way affect the value of these products. To introduce the discount notion again seems an unnecessary duplication of this idea. It has already exercised its influence in determining the present worth of the finished goods, and has no more need for special consideration in connection with the explanation of wages than it has in explaining the price of the raw materials on which the laborer is employed.

In the second place, compensation for waiting is clearly interest as this subject has been developed by all leading, modern writers. For this reason, the inclusion of the discount notion as a part of the wage theory tends to confuse two separate things, *viz.*, the services of labor and the services of waiting. It is immaterial whether the waiting is done by the laborer, the employer, or the capitalist, the compensation for it is an interest charge. Should either the laborer or the employer furnish this service, it should result in a compensation in excess of their pay for labor services which they have each contributed.

Lastly, the discount notion can affect only what the entrepreneur can afford to pay in the operation of his business. He has to reckon with the time element and the risks involved in the conduct of his business in determining whether he will hire the labor, capital, and land requisite for its operation. From his point of view the longer the time elapsing between the start and finish of the productive process, the greater the cost involved. On this account, it would seem that the wages in those industries that require the longest time for placing the finished goods on the market would be lower than the wages in industries that require less time. This conclusion is not borne out by experience. *Employers in these lines have to pay wages commensurate with those paid by other employers using similar kinds of labor.* Even granting the time element as a factor affecting this payment, the wages paid would more likely conform to those offered by the employers whose time processes are shorter, for, by the logic of the discount notion, these employers could afford to pay a higher wage for the same type of labor. Experience shows that laborers of a given grade are paid approximately the same wage, irrespective of the time element required in the processes of the industries in which they are engaged.

The better line of reasoning, then, is that the waiting cost is one factor which the entrepreneur has to include in his reckoning, whether he will undertake a given line of production; and, as a part of this waiting process, he has to calculate the exchange of money value in the form of wages for partly manufactured goods that embody these services. If, on the basis of the forecasts that he makes, he concludes that he can afford to pay the wages which the labor needed in the industry is getting in other lines, have a surplus that will compensate him for his own services, and pay for the waiting, either to himself, if he has furnished the working capital, or to others, if he has borrowed it, he will then undertake the project. The entrepreneur would certainly distinguish between his capital costs, including all capital funds requisite for the operation of his business, and his wage bill. It would seem better, therefore, both practically and logically, to keep the wage and interest theories separate, and to eliminate the discount notion from the explanation of wages and confine our attention to the marginal productivity theory which has been formulated above. The employer pays what he has to pay, and since he is in competition with other employers

he will have to pay the market rate for any particular kind of labor. The social significance of the services of any particular grade of labor is measured by the significance of that labor at the margin of its utilization.

Organization and Wages.—The marginal productivity theory of wages assumes the perfect operation of the force of competition. We have already seen that this force never works perfectly, and in the case of labor there are many obstacles to prevent this result. The employer and the employee are not likely to be on an equal basis in estimating the value of the services of labor. Not having access to the records of the business, the workmen have to rely upon the operation of the force of competition in the general labor market as a guide for the price they ask for their services. What other employers are bidding for labor helps to determine, in a rough way, what any grade of labor can expect to get. There is a widespread feeling among the working classes that employers generally take advantage of the ignorance of workingmen and pay them less than their services are worth. As a result of this and other influences, there has grown up in most industrial countries an organization movement for the purpose of bargaining collectively and securing a larger share of the income for labor.

The literature is filled with contradictory statements concerning the effects of trade unions on wages. For this reason, it will be worth while to examine the theoretical aspects of this question. The problem presents two different issues. First, the effect, if any, that organization may have on the contractual wage, and second, the effect, if any, upon the economic wage. There is little doubt that through organization and collective bargaining the contractual wage is modified to the advantage of the workmen. By this means, the group can employ their ablest leaders who can become more familiar with general labor conditions than the average workman and can, therefore, drive a better bargain. In the most enlightened organizations of today, a great deal of attention is paid to the collection and interpretation of wage and other business data that bear upon the labor contract. Some unions employ economists, or statisticians, for the purpose of conducting studies of these subjects, while others usually employ experts in these matters at the time new agreements are being negotiated.

If the workmen bargain as a group, the employer is less likely to take advantage of the men than if they bargain as individuals.

While some employers may follow voluntarily the practice of paying as high a wage as their industry will stand, because of its effect upon the morale of the working force, others will pay as little as they have to in order to get the labor. When the employer knows that refusal to pay a definite wage may result in the loss of his entire force, he will be more inclined to pay the full value of the labor services than when men bargain individually. The loss of the services of one man will not ordinarily cause any great inconvenience to an employer, whereas the loss of his whole force would paralyze his business and cause a loss to him that is commensurate with the loss of wages to the individual workman. For reasons of this character, there can be little doubt that collective action on the part of the workmen affects the contractual wage. They are thus in a strategic position to get the full value of their services in the wage bargain. If they should succeed in pushing the contractual wage above the economic wage, it would result in a loss to some other share. The more intelligent labor leaders recognize this fact, and are unwilling to force the wage rates so high as to "kill" the business. They recognize, in other words, the legitimacy of the other shares and know that if the entrepreneurs are unable to pay their bills they will go into bankruptcy and out of business. Since the operation of modern business depends upon wise entrepreneurship, it is not to the advantage of laborers to force their contractual wage above the value of what they add to the industry, or, in other words, above their economic wage.

The second of these issues is often hotly debated among labor leaders and employers. Nor is there agreement on the question among economists. The issue can be formulated thus—can workingmen increase their economic wage through organization? In other words, will the marginal productivity of labor be increased by organization? Most employers assert that the effect of unions and organization is to reduce the efficiency of labor. They claim that the union men take advantage of their strength to turn out less, or poorer work than they would if they were not organized. This issue is a very complicated one and there is ample evidence on each side of the question. There is plenty of evidence of "soldiering" and "loafing" on the job under the protection of the union. In many instances organization may give to the group a degree of monopoly power. If so, this power is likely to be used by workingmen, as it is

usually used wherever found, namely, for selfish purposes. The effect of organization, wherever collective power is so used, should be attributed to the existence of monopoly rather than to organization *per se*, even if it is difficult to distinguish one from the other.

The effect of organization, however, can be studied on the assumption that there is no attempt to abuse the power afforded by it. Is it possible to enhance the economic wage through organization? The answer to this question is in the affirmative, providing we limit our study to a particular group of workmen. The limitation of the number of men in a given trade will, in itself, enhance the social significance of the services of each man in the group. This conclusion follows from the application of the principle of diminishing productivity. We saw, in the sections devoted to this topic, that mere increase in numbers of workmen capable of performing a definite kind of service would lower their marginal productivity. The converse of this conclusion is applicable here. The mere reduction of numbers in a given trade will enhance the social significance of the services of every member of that group of workmen. In this sense, then, we can conclude that a specific group of workingmen can increase their economic wage through organization.

But one should immediately perceive the limitations of this conclusion which applies only to a single group or trade. It would be quite a different thing to assert that all laborers could enhance their economic wage through organization. In the case of complete organization, the total supply of labor would not be affected; and while some groups might increase their marginal productivity, it would mean that limits had been placed upon those entering these trades, thus enlarging the numbers left for the other groups. The effect on the other groups would be to lower the marginal productivity because of this increase in numbers. Our net conclusion is that organization may result in increasing the economic wage of particular groups of laborers, but will not do so for all labor. Furthermore, the principal gains that labor has obtained through collective action have come through the effect of organization in bringing the contractual wage more nearly into accord with the economic wage.

Standard of Living and Wages.—Actual wages are, as we have seen, always the result of a labor contract. With the growth of collective bargaining, a standard of living theory of wages has

received a great deal of credence. Once any group of men have established a given standard of living, they will likely resist as long as possible any attempt to undermine it. For this reason, we find many who argue that wages are determined by the costs of living. This theory has been advanced in most of the significant arbitration cases that have occurred in this country, and has been a very effective argument in securing a favorable decision. Elaborate figures on this subject are regularly submitted in support of the particular claims made by the representatives of both the employers and the employees. Because the theory has gained such wide acceptance, it will be well to examine its theoretical basis.

We should note first that a standard of living, meaning thereby the amount of consumable goods of one kind or another that a laborer can buy, is a function of his money wage rather than a cause of that wage. For instance, if a laborer and his family secure \$2,000 per year, their standard of living will consist of the goods which can be bought with this sum of money, and it has little to do with the determination of that sum. It is conceivable that the money wage may be so small that the goods purchased are insufficient to maintain the workman and his family in physical efficiency. An increase in the real income to laborers so situated may be followed by an increase in the productivity of their services. But even under such circumstances the worth of the services is determined by the marginal productivity of the labor and not by the cost of the living standard. It may be that resistance to the attempts to undermine the standard of living may result in the postponement of marriage, or other methods of reducing the number of workers and, thereby, affect the marginal productivity of that group of workers. But the recognition of this possibility does not change the fundamental explanation of the value of labor services, but simply stresses the methods adopted as a means of maintaining a given standard of life. Here, as elsewhere, costs can influence the worth of goods only through their effect on the supply. The effect of costs on the supply of labor services would be shown much more tardily than on the supply of other kinds of economic goods. Theoretically, then, we must conclude that the standard of living has no more to do with the explanation of wages, than costs have in determining the value of other goods. The value of any economic good is explained primarily from the side of its utility, therefore, a wage

theory is concerned mainly with an explanation of the worth of labor services rather than with the cost of rendering them.

The conclusion concerning the explanation of wages gives the correct approach to the determination of the economic wage, or value of labor services, but there are practical considerations in its application which deserve some attention. Logically, it is easy to assert that the wage of any grade of labor consists of the marginal productivity of that kind of labor, but practically this amount is not easy of measurement in the degree of definiteness required by a wage scale. It is easy to assert that the employer cannot afford to pay more than the laborers add to the value of his product, but to measure that amount definitely is no easy task. The wage scale must be expressed in terms of a definite sum per day, per hour, or per piece, for each worker. The difficulties of this task will be apparent when one reflects that many workers are cooperating in making a good which is produced in advance of the demand for it, and whose value is determined at the time of its sale. The worth of the labor services embodied in the good is reflected from the value of the good itself, and can only be estimated at the time the labor contract is negotiated. The theoretical statement of wages gives only a general formulation which must be adapted to the specific circumstances of every wage contract.

In the face of such difficulties, it is not surprising that both the employers and employees seek a basis for negotiation that is more definite than the general formulation given above. As a method of adjusting wage disputes, there is something to be said in favor of the standard of living theory. Assuming that the marginal productivity of labor exceeds the value of the existing standard of living, the submission of figures that measure the cost of that standard is helpful in wage negotiations. The costs of any standard of living can be measured quantitatively with reasonable accuracy. There is a greater likelihood of an agreement on these figures than on the worth of the labor services. Consequently, by means of such figures, the realm of uncertainty and, hence, that of bargaining can be somewhat narrowed. For this reason, as a practical method of adjusting wage disputes, the standard of living has a significant place in the determination of actual wage rates. The recognition of this fact, however, does not modify our fundamental principle in determining the value of labor services and, therefore, of wages.

The standard of living theory is, in fact, an extension of the subsistence theory and there is no more justification for accepting these theories in explaining wages than there is for explaining the value of other goods on the basis of their costs of production. The value of labor services must, therefore, be derived from the utility of those services as expressed through the goods produced.

The Saving Wage.—In this country, the labor leaders have not been content with the implications of the standard of living as a wage theory. They see that the general acceptance of the theory would be an obstacle to the securing of an increasing share in the income from industry. They have changed the character of their demands and have asserted that they want labor to participate in the results of invention and industrial progress by the receipt of a constantly increasing share. This argument has gone through several phases, one of which is the demand for a *saving wage*. It is held that the laborers should have not only enough to take care of their present needs, but sufficient to enable them to save for old age, for the education of their children, and for other contingencies of a normal existence. It will be observed that this, after all, is only an extension of the previous theory, and may gain credence in the adjustment of a wage dispute but does not attempt to explain the value of labor services.

Wages as a Source of Demand.—More recently a theory has been advanced that wage rates must be maintained as a means of maintaining a market for the products of industry. It is argued that the cutting down of wages just reduces by that amount the purchasing power of workingmen and that this reduction will eventually make its influence felt by disturbing the balance between production and consumption and thus become an important cause of industrial depressions. Like most arguments of this character, there is an element of truth in it. We have already shown that the production of goods tends to become a demand for labor. What one group of workers produces is traded in the market for the goods that another group has made. Hence, idle labor reduces the volume of goods available to be exchanged. But all of these processes are carried on in terms of money, and the wage rates paid by an employer cannot exceed the value added to the product by the men. If the entrepreneur has made a mistake in kind or amount of goods produced, it will not help him, or even the laborers, to pay the men the same or higher

wages for the purpose of maintaining demand. The only economic cure for such a mistake is a readjustment of the factors of production and the direction of these factors to the manufacture of some good which consumers will buy at a price that will cover all of the costs of production. Wages as a cause of demand, then, come back to the wise direction of productive effort in maintaining a balance between production and consumption. This issue has to be settled by a comparison of the utility of the good in relation to the costs of producing it rather than by attempting to maintain a demand for it by the maintenance of the costs involved in its production.

Conclusion.—Our final conclusion, then, is that wages are determined by the marginal productivity of the particular grade of labor that is under consideration, just as the value of any other good depends upon the estimate placed on it by the marginal users. This principle holds, even though there are many extremely difficult problems in its application to the complexities of modern industry. The task of adapting the principle falls upon the business economist, whose function is to adjust the internal organization and operation of a business enterprise in accord with the external forces that surround it. The task is no more laborious nor complicated than those that confront the engineer, or the industrial chemist, in applying physical or chemical principles to the specific problems which they encounter in modern industry.

Suggested References

- BYE, RAYMOND T., "Principles of Economics," Chap. XX.
 CARVER, THOMAS N., "The Distribution of Wealth," Chap. IV.
 ———, "Principles of National Economy," Chap. XXXV.
 CLARK, J. B., "The Distribution of Wealth," Chaps. VII, XI, XII, and XIII.
 ELY, RICHARD T., "Outlines of Economics," 4th Ed., Chap. XXII.
 FAIRCHILD, FURNISS, and BUCK, "Elementary Economics," Vol. II, Chaps. XXXVIII and XXXIX.
 FETTER, FRANK A., "Economic Principles," Chaps. XVI-XIX.
 HAMILTON and MAY, "The Control of Wages."
 HOBSON, J. A., "Economics of Distribution," Chaps. V, VII.
 KLEENE, G. A., "Profit and Wages," VII-IX.
 RUFENER, L. A., "Principles of Economics," Chaps. XIX-XXI.
 TAUBSIS, FRANK W., "Principles of Economics," 3d Ed., Vol. II, Chaps. XLVII, XLVIII, and LII.
 WALKER, F. A., "Political Economy," Part IV, Chaps. V and VI.

CHAPTER XXII

PROFITS

Summary Statement.—The shares previously treated—rent, interest, and wages—have been called functional shares. They are the portions of the income from industry that are assigned to the factors of production on the basis of a service rendered and are, therefore, regarded as compensation for functions performed. It should be understood that the principles of distribution that have been set forth impute these shares to the factors of production, but they are, in reality, received and enjoyed by the respective owners of these factors. That is, the landlord receives rent; the capitalist, interest; the laborer, wages. We have called the principle by which these shares are determined, the marginal productivity theory of distribution. Underlying the theory and the analysis that led to its formulation is the assumption of perfect competition and perfect mobility of the factors of production. Throughout the whole of this discussion, the reader has been cautioned to recognize that the actual conditions do not coincide perfectly with those assumed, but that the assumption is justified logically as a method of explaining the complicated relations found in our modern economic life. To reduce these complicated relations to some degree of system that will enable the mind to grasp the complexities and the forces operating among them requires the assumption of some arbitrary standard of stability as the starting point of the analysis. Perfect competition and a perfect market have constituted the basic assumptions for this purpose. Modifications of the conclusions reached are necessary in order to bring them more nearly into accord with actual conditions, and the statements given above should be accepted in this light. This caution is repeated at this point because of the nature of the treatment of the subject of profits.

On the basis of the assumption of a perfect market, as just outlined, the whole of the income from industry would go in the form of the three functional shares already treated. In other words, if competition were perfect, there would be no share known as

profits. The actual shares would equal the theoretical shares. Hence, in the treatment of this topic, the reader must remember that we are dealing with certain modifications of our general conclusions. The influence of competition is to distribute all of the income from industry into the three shares mentioned above, but because competition does not work perfectly there arises a share commonly known as profits.

The Entrepreneur and Profits.—The share known as profits ordinarily accrues to the entrepreneur and, in order to understand this fact, attention must again be called to the position of the entrepreneur in modern industry. We have seen that private property, free contract, and individual initiative were among the chief characteristics of our economic order. The entrepreneur occupies a strategic position in the operation of this system. Generally speaking, all productive enterprises are initiated and operated by him, the exceptions to this general statement being those undertakings that have been started and are operated by some governmental agency, or by a socialistic or some other collectivist system of industry. In most of the countries of western civilization, aside from governmental enterprises, industry is conducted by private entrepreneurs. It is the private entrepreneur who forsee an unsatisfied desire, or the possibility of stimulating a new one, and then proceeds to marshal the factors of production necessary to produce the good that will gratify it. He contracts with the landlord for the use of land, either in the form of a lease, or a direct purchase; he borrows or furnishes the capital funds with which to equip his plant and to supply the working capital necessary for its operation; and, finally, he bargains with labor for the control of labor services or does the work himself. In all of these cases, he agrees to pay a stipulated sum of money in exchange for the services of the respective factors of production and thus assumes full responsibility for the operation of the business.

Strictly speaking, the entrepreneur performs no other service than to perceive a need for some good, and then gain control of the factors of production necessary to produce it. In reality, however, he usually invests some of his own capital in the business and performs some of the services connected with its management. He may own the land on which the business is situated, furnish all or a part of the capital for its operation, as well as perform all or a part of the labor required for conducting it.

Such a combination of functions is seen in many small enterprises, such as farming, where the farmer is a landlord and a capitalist, as well as a laborer. But in the corporation there is likely to be a separation of these functions because of the complicated problems connected with large-scale production and the ease with which the separation may take place in this form of business organization. The real entrepreneurs in the corporation are the stockholders who may do nothing more than select the enterprise to be undertaken, appoint the management to direct its operations, and provide the capital funds requisite for commanding the factors of production needed in the business. These funds may be furnished by the stockholders from their own savings, or borrowed from some financial agency, but, in either case, the stockholders assume full financial responsibility for the business.

In those enterprises in which all of the functions of production are centered in one man, it is not customary to separate his earnings into wages, interest, and rent. In such cases, the entrepreneur is more likely to regard the total return in excess of any outlays for leased land, borrowed capital, and hired labor, as profits. While such a procedure may satisfy the ordinary accounting needs of a small business, it is not a scientifically accurate classification of income, nor does it fit the needs of large corporate organizations. In these larger undertakings, some attempt must be made to separate the returns into shares that correspond closely with the functions performed. In such enterprises, the entrepreneurs (stockholders) receive interest in the form of dividends on the capital which they have provided and, in addition, any sums in excess of the normal returns to the other factors as determined by the contracts for their control. In other words, they are the residual claimants on the income from the industry.

In calculating their shares, all expenses of operation are deducted from gross earnings, including payments for the use of land and buildings, for raw materials, for capital funds borrowed, and for all hired labor required in the operation of the business, including payments for the services of management and direction. Stockholders frequently delegate the management of their business to others, in which case the wages of the managers would clearly appear as an expense. If some of the stockholders performed these services, they would receive a salary for their work, and these sums should be reckoned as a part of the necessary

expenses of operation and not included in the net earnings of the corporation. What has been said concerning the single proprietor and the corporation is equally true of the partnership. It would thus appear, whatever the form of business organization, that all surpluses arising from the operation of a business accrue to the entrepreneur by virtue of the fact that he has assumed the responsibilities for conducting it.

Nature of Profits.—From what has been said concerning the operation of modern business, it will be evident that the term profits has a variety of meanings in common usage. It may be used to refer to all of the net returns received by the owner of a small business, including rent for any land that he owns, interest on the capital he has furnished, and wages for the labor that he has performed; or its meaning may be confined to the net surplus, if any should exist after the contractual shares have been paid. In the previous discussion, we have carefully distinguished between the economic and the contractual shares. The former is the full value of the services of the respective factors in use, while the latter is the amount which the entrepreneur has to pay in order to control these services. These contractual shares, in the form of rent, interest, and wages, are expenses to the entrepreneur and are regarded by him as the costs necessary to the production of a good. If his business venture succeeds, any surplus gains are his, but if it fails, he bears the losses. This surplus, or residue, that consists of the difference between the gross receipts from the sale of an article and the expenses of producing it, accrues to the entrepreneur and constitutes what is commonly known as profits. Defined a little more specifically, profits are those sums, if any, that accrue to the entrepreneur under a system of private property and free contract, by virtue of the fact that he takes the initiative and responsibility of operating an independent business, and consist of the difference between gross income and the direct expenses of operation. It will be noted that there is no assurance that a surplus will be present. In fact, many entrepreneurs have had to assume losses, when a business venture did not fulfil the expectations of those who had launched it.

A Concrete Example.—The above definition constitutes what may be called *ordinary profits* but, since the term has several meanings, it may be an aid in distinguishing these usages to give a simple statement showing the operation of a small concern, such as a general grocery store. It will be assumed that, in

this simplified form, the statement represents the operations of the business for a regular interval of time, as a year.

Receipts		Expenses	
Grocery Department.....	\$20,000	Cost of Groceries.....	\$16,000
Meat Department.....	15,000	Cost of Meats.....	11,250
Fruit and Vegetables.....	5,000	Cost of Fruits, etc.....	2,500
		Wages.....	2,250
		Heat and Light.....	200
		Taxes.....	150
		Depreciation.....	600
		Miscellaneous.....	500
Total Receipts.....	\$40,000		\$33,450
Total Expenses.....	33,450		
Net Income.....	\$ 6,550		

From the previous statements, it will appear that the sum \$6,550 represents ordinary profits, or the difference between gross income and the direct operating expenses. Ordinary profits include the remuneration of the entrepreneur for his services, together with the return to any capital and land which he has furnished and used in the operation of the business. It would exclude any payments for borrowed capital and leased land, as these items would appear as a part of the direct expense of operation. *Business profits*, on the other hand, would exclude the wages of management from ordinary profits, but would include the return to capital and land that were owned and used in the business. If the owner-manager of this store should charge the business for his services at their full worth, say, \$3,000, then the business profits would amount to \$3,550.

Suppose, now, that the owner has advanced \$5,000 of his own money in this business, and that 6 per cent is the market rate for funds similarly invested. Suppose, further, that the yield of the land and building used for the business was \$100 per month, which, we will say arbitrarily, is divided into \$60 for the building and \$40 for the land. Under these circumstances, we would have an interest charge of \$1,020 composed of the 6 per cent on the \$5,000 and \$60 per month on the building, together with a rent item of \$480, or a total of \$1,500 as the return on the capital and land owned. Deducting these sums from the busi-

ness profits, as defined above, we get a surplus of \$2,050. This sum constitutes what may be called *pure profits*.

The reason for deducting the above items from ordinary profits should be apparent, for they are all true costs to the entrepreneur in producing a good, or in rendering a service. The work of the manager is an essential task in every industrial undertaking and, in the more highly organized ones, payments for such services are made because a division of functions has taken place and such payments regularly appear among the expenses of operation. Likewise the capital and land owned by the entrepreneur and devoted to a specific use, are just as productive as if they had been borrowed and, therefore, they should receive compensation equal to what these factors would earn in other uses. If they do not yield a return from their use in the business, the owner suffers a loss of income equal to what these factors would yield in some alternative use. The entrepreneur should, therefore, charge his industry with costs that include a return on these items that is equal to their yield in the general market. Unless he makes these charges, he does not have a true cost, nor is he in a position to determine whether his business is yielding anything in the form of pure profits. Pure profits, then, is the sum that accrues to the entrepreneur after all outlays have been deducted from gross receipts, together with an allowance for the wages of management and an income on land and capital owned by him. They are the sums, if any, that accrue to the entrepreneur as the result of his assuming the responsibility for operating an independent business, and are in excess of the normal return yielded by the factors of production in other uses. This is the share that requires separate treatment.

Wages of Management as Profits.—Some writers have made the wages of management the minimum of profits, basing their treatment on the theory that the owner-manager gets his compensation out of the residue, if any is left after the payment of all contractual shares. The objection to this treatment is that it classifies unlike things together. The tasks performed by the entrepreneur are in the nature of labor services, and any compensation received for them should be regarded as a wage or salary. While in many industries a payment is not made for these services in advance of the sale of the good, yet in all cases where production proceeds only on orders, a charge for these services would be included in the estimated costs which form the basis for the bids

or prices at which the work would be undertaken. In addition, it may be as truly said that the owner-manager should not charge a return for the capital and land furnished by him as a cost, since the return for their use appears as a surplus rather than as a payment in advance of the sale of the product. Such a treatment of the wages of management would class these unlike sources of income together, merely because the accounting records show them as a surplus over outlays. Scientifically, this surplus, if any exists, should be broken up in such form as to show whether the several factors are earning as much in a particular industry as they would if used in some other way. For this reason, compensation for the services of the owner-manager should be included as a necessary part of the costs and should not be regarded as a part of the profits of the industry.

While business practice may not regularly exclude the wages of management from profits, the importance of the corporation, as a form of business organization in which managerial services are paid a stipulated salary, emphasizes the logical need for the above distinction. In all such organizations, payment for the services of the manager is deducted before dividends are declared which indicates that the practice here is to regard these payments as costs along with the wages paid for other types of labor services. Unless the owner-manager charges the business with the full market value of his services, he is unable to determine whether it is yielding a net profit. Furthermore, should a prospective purchaser of the business use profits as a basis for determining the price to be offered for it without eliminating the value of the manager's services, he would be led to offer more than the business was worth, because he had capitalized the manager's salary as well as the net earnings.

If the owner-manager does not get a return at least equal to the market value of his services, he is in reality suffering a loss. There are some men, however, who have such a strong desire to be known as "business men" rather than as "salaried men," that they will continue to operate a business without pure profits, and even for an income less than they might secure by selling their services to others. The amount of this loss is, therefore, an estimate of the significance which they attach to the prestige of being known as independent business men. It is a kind of psychic income that does not enter into our money economy but does, nevertheless, have an influence on the choices of marginal

entrepreneurs. However, the better reasoning seems to be that wages of management should be regarded as a cost, and profits should be thought of as a surplus in excess of all costs.

Rent, Interest, and Profits.—The items of rent and interest should be deducted from gross earnings, whether the land and capital are owned or borrowed. If the grocer, in the illustration given above, decided to dispose of his stock and close his store, the capital funds invested in the business could probably earn as much in some other use. Likewise, the land would have an equal rental value if devoted to some other purpose. To determine whether one line of business will yield more profit than another, it is necessary for the owner-manager to charge against each use all costs, including rent and interest. But if rent is charged as a cost, does not this run counter to the previous conclusion that differential rent is not a price-determining item in the cost of producing a good? The purpose at that point in our discussion was quite distinct from the one now before us. We were endeavoring then to determine whether the payment of rent would affect the price of a good and our conclusion was that, except for marginal rent, the market price of a good would not be affected by the payment of rent. It was held that rent was not a social cost as interest and wages and, therefore, could not affect price by limiting the supply of the goods put on the market. In fact, it was held that rent was the result of price rather than a cause of it.

But our problem now is one that concerns the operation of a private business. To the entrepreneur, one of the primary questions is, whether his present line of business is yielding him as large a return as he might receive from some other; whether, in other words, he is securing as large profits by utilizing the factors of production in one direction rather than in another. For this purpose, it is necessary that he charge against his gross receipts all outlays for the use of capital, labor, and land, including the full value of his own services and a charge at going rates for any capital and land which he has furnished. In this sense, rent is a cost that the entrepreneur must recognize. And, since he is engaged in business primarily to secure profits, it is quite evident that the rent which he pays should appear as a cost that is as definite an outlay as his payments for labor and capital. It is, therefore, in this private and acquisitive rather than in the former social sense that we are now speaking of rent as a cost.

Source of Pure Profits.—As defined above, pure profits are a surplus which some business men enjoy that others do not. For this reason, profits do not constitute a price-determining item in the cost of producing a good or in rendering a service. This fact can be seen by examining some of the principal sources of profits. As has been shown, the entrepreneur contracts for the services of the factors of production for stipulated sums. If, in the utilization of these services, he receives in economic shares more than he has paid for them in contractual shares, the differences will accrue to him as profits. It is impossible to measure accurately in advance the full value of the services of any one of the factors, land, labor, or capital. The value of these services has to be derived from the value of the completed good which cannot be measured accurately because the present worth of a future good can only be estimated. No one can know with certainty the course of demand during an interval of time, even for the most stable and consistently purchased commodities. In other words, production in anticipation of demand involves a large element of uncertainty. If in bargaining for the services of the factors of production, the entrepreneur gets them for less than they actually earn, as measured by the price of the finished good, then these gains from bargaining will accrue to him as profits. If he can borrow capital funds at 5 or 6 per cent, and make them earn in his business 8 or 10 per cent, the difference will appear as profits. A similar condition prevails in the case of land and labor. If he can secure the services for a contractual share that is less than the factors in use add to his income, the balance will accrue as profits. Such sources may be called the *gains from bargaining*.

Profits may arise also because of differences in price between two different markets. These markets may be different in point of location, or in time. The man who knows that market prices in one locality exceed those in another by an amount that more than covers the costs of transporting the goods will undertake to move them from one market to the other, and will receive on this account a profit item. Price differences of this kind tend to be reduced by competition to a minimum that just measures the costs of moving the goods from one locality to another. Likewise, the man who believes that a good will command a higher price at a future date than at present, and that the difference will more than pay for all of the carrying charges, including warehouse costs, insurance, and interest on money invested;

may buy now for future delivery and make a profit. The profit item in this case arises on account of the differences in price at two different intervals of time. This constitutes the field of activity of the operators in the great terminal markets, such as the boards of trade or stock exchanges. Here men are regularly buying and selling contracts for future delivery. Those who estimate the course of future prices most accurately gain—others lose.

Another source of profits is the differences that arise on account of price fluctuations that take place after the raw materials have been purchased and the processes of production have commenced. An unusual demand for a good may develop suddenly, especially in connection with goods affected by fashion and style elements. A fad may result in a market price that will yield a profit to those who happen to be in a position to enjoy it. These are some of the principal sources of pure profits. It will be evident that in none of these cases does the existence of profits affect the cost of producing the good. In fact, they are in all cases the result of price conditions that are fixed by other forces. If profits exist at all, they are in the nature of a surplus, or a residue, left in the hands of the entrepreneur after he has paid contractual wages, interest, and rent at figures determined by conditions prevailing in the respective markets for labor, capital, and land. Pure profits, then, are not a cause of high commodity prices, but exist because some entrepreneurs can bring the goods to the market more cheaply than others.

Profits and Losses.—There is plenty of evidence that profits, as defined, exist. Accounting practice, especially in the smaller concerns, does not always show profits in this form, but in the modern corporation, rent, interest, and wages, including managerial salaries of all the officials of the company, are charged as costs and any surpluses that may exist are in the nature of pure profits. It is evident, from the above discussion, that the existence of a surplus is not an assured fact of business. On the contrary, the conduct of business is surrounded by great uncertainty and it is altogether probable, although statistics will not permit the balancing of profits against losses, that, surveying the whole range of industry, the profits gained by some business men are about offset by the losses of others. In the four years from 1916 to 1919, one of the most prosperous periods in the history of American business, the corporations of the United

States reported to the Bureau of Internal Revenue net deficits amounting to \$3,000,000,000 which constituted 8 per cent of the total net income reported by the corporations showing net income.¹ If the losses were this large during these years, they would certainly be much larger during periods of normal business. It will be apparent, then, that profits accrue to some business men; others get only wages of management, plus the normal returns for any capital and land which they have furnished in the operation of the business; while still others suffer losses. In the long run, the effect of competition is to eliminate those entrepreneurs who cannot operate a business and pay in contractual shares the current rates for the services of land, labor, and capital. The number of business failures that are annually recorded is evidence of this fact. The entrepreneur, therefore, has to assume this uncertainty of success whenever he undertakes the operation of an independent business.

Risk and Uncertainty.—It is sometimes argued that profits are a compensation for the risk involved in a business venture. To the vast majority of people, the uncertainties of business constitute a real disutility and it would seem that a definite price must be offered in order to induce men to assume this cost. Before proceeding further with this analysis, we should re-examine the nature of the uncertainty in the conduct of business. In the chapter on Risks in Production, it was shown that some uncertainties are calculable while others are not, and that there are ways of reducing uncertainty so far as the individual is concerned. For instance, in the case of fire, uncertainty can be reduced to a reasonable degree of accuracy by combining individual risks as mass action is much more nearly predictable than individual action. The number of fires that will occur in a hundred thousand instances can be predicted with reasonable degree of accuracy, even though it is quite impossible to foresee what individual will suffer a loss by fire.

But after the most accurate methods known have been applied to reduce risk, there are still uncertainties that the entrepreneur must always assume. The recognition of this fact seems to justify the distinction between *risk* and *uncertainty* as conditions which the business man has to confront. An uncertainty which can be calculated may be called a "risk." The non-calculable uncertainties may be called simply "uncertainty." Insurance

¹ Figures taken from ELY, R. T., "Outlines," 4th Ed., p. 518.

companies are organized to deal with risks as here defined. The payments made to these companies are in the nature of a compensation for a service rendered. If a company assumes a risk for less than it can be carried by the individual, the latter will gain thereby. On the other hand, if the company collects in premiums more than the cost of carrying the risk, the difference between the premium and the actual cost is in the nature of business profits, as defined above, and accrues to the company.

From this discussion the distinction between insurance and profits should be apparent. Insurance deals with the calculable risks of industry, while profits are connected with the assumption of the non-calculable uncertainties that are involved in most business operations. Whenever the uncertainties of business become known, they can be discounted and adjustments made either to eliminate them or to make provisions for distributing them in definitely measured quantities. Not every business man is in a position to forecast all of the predictable uncertainties of his business. Consequently, this service is left to an independent group of entrepreneurs who undertake this task for those who dislike the assumption of uncertainty, or who do not have the facilities to forecast and make provisions for it. The insurance companies will, for a definite charge, carry the calculable risks, and forecasting agencies are in modern times endeavoring to furnish to business men the information on the basis of which they may guide more accurately their business operations in accord with the probable future demand for their products.

Profits as a Functional Share.—After all known methods of reducing and eliminating risk and uncertainty have been applied, there will be some uncertainty that cannot be accurately forecast. The entrepreneur assumes this uncertainty in becoming an independent business man. It might be argued from this, therefore, that profits are a compensation for the services of carrying the non-calculable risks of industry and should, on this account, hold a place as a share of income alongside of rent, interest, and wages. The difficulty with this line of reasoning is that it proves too much. If profits be regarded as a compensation for the assumption of the non-calculable uncertainty in business and are, therefore, a functional share, then, all entrepreneurs would tend to get profits regularly because all business ventures involve some uncertainty of this character. From the point of view of production, profits would be a cost along with

wages and interest; from that of distribution they would be a payment for the disutility involved in the assumption of the uncertainty, just as wages act as a compensation for the disutility of labor. In addition, if this treatment be correct, we would expect profits to be highest in those industries in which the uncertainty is greatest, a condition that does not obtain. Profits are neither a part of the necessary costs of producing a good, nor are there any positive relations existing between the degree of uncertainty and the amount of profits received. If the hope of gain is large, there are many who will be willing to assume a limited amount of this uncertainty, which is only another way of saying that the hope of gain is the motive force that induces men to assume tasks that are otherwise disagreeable to them. But we know that some men will continue to operate their business, even if it pays only normal returns in the form of wages, interest, and rent. In fact, some men accept less for their services as independent business men than they could get on a salary basis. In this sense, then, profits differ from the other shares of income that have been previously treated and should not be regarded as a functional share. They are surpluses enjoyed by some entrepreneurs and not by others. They are the sums received, if any, by the entrepreneur by virtue of the fact that he is operating an independent business and are always in the nature of surpluses.

The Profit Motive in Industry.—There is an element of truth in the conception of profits as a functional share, however, which can best be discussed under the topic of profits as a motive force in industry. It is clear from the above section that profits do not accrue to all who assume risk or uncertainty in the operation of business, but the existence of profits and the hope of enjoying a surplus in excess of the normal returns to the land, labor, and capital used in a business are powerful stimuli to the business man. It is not necessary that all should enjoy extraordinary gains in order to get many to compete as is well understood from the experience in all competitive contests. In intercollegiate athletics, it is not necessary that every athlete win a contest or every team win every match in order to get men and teams to compete. The hope of victory may keep many an athlete and many a team in the competitive struggle.

In the same way, the hope of securing business profits is a powerful stimulus to the assumption of the uncertainties in the operation of an independent business. It is not necessary that

all business ventures succeed in the sense of yielding profits in order to induce men to become entrepreneurs, any more than it is necessary for every athlete to win his race in order to induce him to compete. The fact that some men succeed and secure large incomes, together with the economic power and social prestige which accompany large incomes, serves as a powerful motive in stimulating others to initiate business ventures and to assume the responsibilities and uncertainties which their operation entails. The existence of profit serves a very important social function in the operation of our present economic order by acting as a stimulus to individual initiative in the conduct of all forms of economic endeavor.

Production for Profits *versus* Production for Service.—Another important issue in connection with the profit motive in industry is concerned with the question of whether the pursuit of money profits coincides with the social needs of a people. It is held by some that production for profits is inconsistent with production for service, and they cite in support of their contention, that business men will close down their plants when profits cease to exist, irrespective of the needs of the community for the products. This argument is faulty in certain important respects. First, the concept of profits is not precise, and includes the total income accruing to the business, or ordinary profits, as defined above. We have already seen that business is often continued beyond the point of yielding pure profits and sometimes for less than full wages of management and normal returns to capital and land that are furnished by the business man. The slowing down or stoppage of production will not take place so long as the products can be sold at a price that will cover the operating costs of production for, as we have seen, pure profits are not an operating cost but the result of a variation in costs that yield surpluses to some business men but not to others.

A second fault in the argument is concerned with the concept of the needs of the community. To be sure, consumers could always use more units of a product, if they did not have to think of the price or cost of securing them. The price of a good sold in a competitive market measures fairly accurately the social estimate placed upon the significance of that good in consumption. If this price will not pay the costs of production, the social need for the good cannot be said to be great. Relating this discussion to the topic under consideration, it may be said that

profits are most likely to appear in connection with the production of those goods, the demand for which is most active. In this sense, the pursuit of pure profits in a competitive industry does tend to furnish the consumer with the goods which he esteems most highly, as judged by the price he is willing to pay for them. As a further consideration of this same point, it may be stated as a general principle, that the pursuit of pure profits in a competitive industry tends toward the most effective use of the natural resources in the light of the existing needs of the community.

The third fault with the above conclusion is the fact that no distinction is recognized between competitive and monopoly profits. The argument that production for profits coincides with production for service is predicated upon the assumption of freely competitive conditions. To the extent that there are natural or artificial obstacles to the operation of competition, there will need to be qualifications of the above conclusion. We know that it is impossible to forecast with accuracy the exact course of demand, or for the buyer and seller to be on equal terms in respect to the knowledge of the market. We know also that the costs of producing a good are not identical, and that some producers will enjoy an advantage over others in the production and marketing of this good. This advantage is likely to result in profits to the producer thus favored. Nevertheless, the effect of competition is to dissipate this surplus to the advantage of the consumer, either in the form of a lower price or of an improved quality. To the extent that competition is operating, its effect is to cause the business man to produce the things that people want at a price that just covers the normal costs involved in the use of the factors of production. Any profits that exist under such conditions are obtainable through production for service, with the result that there is a harmony between individual and social welfare.

In the utilization of the natural resources of a country, there needs to be a qualification of this conclusion. The pursuit of profits is likely to involve a much shorter period of time than does the welfare of a nation. This is easily seen in connection with questions of conservation. The welfare of the nation involves more than one generation, and foresight must be exercised in the present utilization of the resources in order that the needs of future generations are not overlooked. Individual welfare, so far as profits are concerned, cannot extend beyond the lifetime

of the existing generation. Pursuit of profits under these circumstances may sacrifice the needs of the future for those of the present and to this extent be out of harmony with production for service. This, however, is not the sense in which the argument is usually used. The meaning commonly conveyed by this argument is that the motive of private profit does not result in bringing to the market the goods which people want, especially at times of industrial depression. Our conclusion on this point is that the effect of competition is to direct the utilization of the natural resources, except as noted, into the channels of the greatest social need, and that any profits arising under these circumstances accrue to those who have had unusual foresight either as to the character of demand, or in their ability to produce the good desired at relatively low unit costs.

Monopoly Profit.—What has been said concerning competitive profits does not hold in connection with monopoly profits. Monopoly always consists of some form of special privilege enjoyed by the individual which usually gives to him power to regulate supply and, through control over supply, to influence the price paid for the good. The purpose of this control is to secure a price that will yield a surplus, or profit, in excess of operating costs. Whatever may be the cause of monopoly, its effect on price and profits is approximately as stated. But the monopolist is not absolutely independent in his operations. He has to produce a good that people want, and has to reckon with the elasticity of demand and with the possibility of substitutes. Consumers can always economize in the use of any good and, if the price fixed is so high as to check consumption, the result may defeat the purpose of the monopolist. Then, again, if there is a possibility of finding a substitute good, the monopolist has to be alert to the potential competition from such a good and its effect on his profits.

In addition to these two influences, unless his monopoly power is very large, there is always the possibility of potential competition. Other entrepreneurs are looking for the same kind of opportunities as those which he enjoys, and there is always a chance in most lines of business that a competing firm may be established. Nevertheless, it cannot be said that production for monopoly profits coincides with the economic interest of a people. Our net conclusion, then, is that so far as competition is controlling, production for profit does tend to coincide with production

for service, but that so far as monopoly or other obstacles to free competition are present, this harmony between the private, acquisitive aspects of business and the primary, social purpose of production does not necessarily exist.

Competition versus Regulation.—The issue just discussed involves questions of social policy of great significance. For instance, should the state attempt to regulate industry and control the profits earned? Or, should the state undertake to tax profits in excess of a definite amount, as was attempted in the *excess profits tax*? Questions like these keep cropping up and are involved in all attempts to regulate industry, as seen in the control exercised over public utility properties. An intelligent judgment upon such questions requires an understanding of the theoretical nature of profits and the place they occupy in the operation of our competitive system. Any attempt to regulate the earnings of an industry, as the railroads or other public utilities, proceeds upon the assumption of a fair return. But the concept of a fair return implies an earning base and a normal annual yield. We may ask whether there is such a thing as normal profits? A few figures will serve to illustrate the complexities of the question.

The net incomes of corporations reported to the United States Bureau of Internal Revenue were \$4,774,000,000 in 1913; \$3,940,000,000 in 1914; \$5,310,000,000 in 1915; and \$10,734,400,000 in 1917. In another field, the earnings of national banks averaged 9.5 per cent on their capital in 1906, 16.4 per cent in 1907, and 9.70 per cent in 1908. In 1918, out of 1,551 concerns in the bituminous coal industry, 337 operated at a loss, while 92 companies earned in excess of 25 per cent on their capital. The average profits for the 1,214 companies that showed a net profit were 18.86 per cent on the invested capital before the payment of federal taxes and 9.72 per cent after the payment, while the range of profit was from 1 to 500 per cent.¹ Figures of this character show clearly that earnings of companies vary greatly from year to year, from industry to industry, and from establishment to establishment. The above figures show business profits in the ordinary acceptation of that term, and include earnings on the total capital invested in the various industries. Any attempt to regulate profits, or to determine what constitutes excess profits should take these facts into account.

¹ Figures taken from Ely, R. T., "Outlines of Economics," 4th Ed., p. 520.

If profits are to serve as the stimulus to the initiative of entrepreneurs, the losses incurred in bad years must be included in the reckoning. In addition, the total earnings of a business should be separated into a return on the factors of production that equals their normal yield, and into pure profits, or the sum in excess of the normal contractual shares. Any attempt to regulate industry, or to tax excess profits, should make a distinction between these two classes of income. If ample provisions have been made for losses during bad years and due account taken of the effect of profits as a stimulus to the effective utilization of the factors of production, then, pure profits may be considered as a subject for special treatment. But, in formulating a policy of regulation, care must be exercised not to lose sight of the effect of this control upon the efficiency in the operation of business and upon the motive for industrial experimentation. The significant issue is not the mere receipt of profits, but whether these profits have been received during the course of free and open competition, or as the result of some form of monopoly or special privilege. In the former case, as we have already seen, the unusual returns will likely be diffused as the result of competition to the general benefit of the community whereas, in the latter, the surpluses will likely be enjoyed only by the monopolists. Monopoly profits seldom coincide with general welfare, and should on this account be subject to special regulation.

The Concept of Fair Return.—In connection with this discussion, an ethical standard is implied. All attempts to regulate public utilities, and all discussion of excess profits imply a fairness of return. *Fair value* of utility properties, *fair and reasonable rates* of return are familiar expressions in the literature of regulation. The public thought on questions of this character is not always clear as to the standards of fairness to be used, or the source of those standards. An analysis of the discussions concerning fair rates for railroads, and for gas and electric power plants will reveal an unformulated assumption that flows from the theoretical, economic relations which have been assumed as the basis for the development of the general principles previously set forth. In a freely competitive market, the return to the various factors of production would not exceed, for any long period, the value which each added to the products. When this idea is translated into contractual shares, the market price of a good is fair when it just pays the competitive costs of producing

it. A rate, whether for railway service or for gas or electricity, is fair if it earns the normal return on the capital invested in the business and pays all other operating costs at their full market price. The source of this standard of equity, so far as return is concerned, has clearly grown from the conditions that would prevail in a perfect, competitive market. The working out of the ideal into a social policy is difficult, partly because of the complexity of modern economic conditions and partly because of a lack of clarity concerning the fundamental assumptions underlying the ideal in the minds of those on whom the responsibility for regulation rests. The recognition of this fact does not make the problem of regulation easy, but it tends to clear the issue and to pave the way for a more intelligent development of a social policy toward industry.

Suggested References

- BYE, RAYMOND T., "Principles of Economics," Chap. XXI.
 CARVER, THOMAS N., "Distribution of Wealth," Chap. VII.
 ———, "Principles of National Economy," Chap. XL.
 CLARK, J. B., "Insurance and Business Profit," *Quart. Jour. Econ.*, Vol. II, pp. 40-54.
 EDIE, LIONEL D., "Economics, Principles and Problems," Chap. XII.
 ELY, RICHARD T., "Outlines of Economics," 4th Ed., Chap. XXVI.
 FAIRCHILD, FURNESS, and BUCK, "Principles of Economics," Vol. II, Chap. XL.
 FETTER, FRANK A., "Economic Principles," Chaps. XXVI-XXXI.
 FOSTER and CATCHINGS, "Profits," Chaps. II-VI, XVII, and XVIII.
 HARDY, CHARLES O., "Risk and Risk Bearing," Chap. XVIII.
 HAWLEY, F. B., "Enterprise and the Productive Process," Chap. VI.
 ———, "Enterprise and Profit," *Quart. Jour. Econ.*, Vol. XV, pp. 75-105.
 KNIGHT, FRANK H., "Risk, Uncertainty, and Profit," Chaps. I-III, VII, IX, and X.
 HAYNES, JOHN, "Risk as an Economic Factor." *Quart. Jour. Econ.*, Vol. IX, pp. 409-449.
 RUFENER, L. A., "Principles of Economics," Chaps. VIII and XL.
 MACVANE, S. M., "The Theory of Business Profits," *Quart. Jour. Econ.*, Vol. II, pp. 1-38.
 SEAGER, HENRY R., "Principles of Economics," Rev. Ed., Chaps. XII and XIII.
 TAUBSIG, FRANK W., "Principles of Economics," Vol. II, 3d Ed., Chaps. XLIX and L.
 WEBLEN, THORSTEIN, "The Theory of Business Enterprise," Chap. III.
 WALKER, F. A., "Political Economy," Part IV, Chap. IV.

CHAPTER XXIII

GENERAL SUMMARY

In the preceding chapters, a survey has been made of the leading principles of economics. These principles are tendencies that operate within the present economic structure of society. Private property, free contract, individual initiative, and private enterprise with voluntary exchange, are the chief characteristics of this economic order. The principles must be understood and interpreted in the light of these general conditioning circumstances. Many of the principles that have been set forth would not hold in a social order that was based upon different institutions. If we lived, for instance, in a society where all property were socially owned and where incomes were distributed equally, or distributed in accordance with some standard of needs, economic life would be quite different from what we now know. If, for individual initiative, there were substituted the initiative of some collective agency, such as the department of public works of a government, the method of satisfying our daily wants and the whole structure of industry would be very different from the present organization. Many of the principles that have been set forth above would be either inoperative, or their course of operation would be different from what now exists. The motives that stimulate the individual now would be greatly modified under a system of social ownership.

A further illustration of the dependence of economic principles upon the structure of economic society may be given. The central theme of economics as a science is the theory of value. While the processes of evaluation would not be affected by the laws of property that prevailed in some other economic structure, the exchange power of any individual commodity would be greatly affected by these laws. Suppose, for instance, there were no laws protecting the individual in the ownership and enjoyment of wealth, and the only basis of protection was that of the personal force of the individual, how much would one give for a house, or an automobile, or any other form of wealth? Only

those things that were immediately usable would have much value, if it were not for the laws of property protecting the individual in the enjoyment of his wealth. If voluntary contracts were not respected and enforceable, how much business could be conducted? These illustrations suggest how fundamental property and contract are in the operation of our economic life and, therefore, in setting limits to the operation of economic principles.

It must not be inferred from this, however, that all of the principles studied are dependent for their existence upon the structure of our economic society. There are some principles that approximate the laws of the physical sciences in their general application. Take, for instance, the principle of diminishing utility. The desire for the necessities of life is common to all mankind. Hunger, thirst, and other bodily needs are present as long as life exists, and create desires for the goods that are capable of gratifying them. The type of goods consumed may differ, as between different peoples, or the range of goods that makes up the standard of living of different groups of people may vary widely. But, in the consumption of wealth, the reaction on the desires is approximately the same, irrespective of the general structure of the order in which men live. Rightly understood, the principle of diminishing utility holds under all conditions of human life and is independent of the economic structure. The consumption of successive units of a good within a short interval of time will yield diminishing satisfactions.

There is ample evidence from human experience that this generalization expresses a fundamental truth. It is true that there are difficulties in measuring the satisfactions derived from the consumption of wealth because satisfactions are subjective in character and, therefore, do not lend themselves to exact objective measurements. The fact, however, that men do shift in the use of wealth from one good to another is evidence of choice on their part. Even if it be true that some purchases are made as the result of habit or custom, without any conscious thought on the part of the purchaser, there is a large range of purchases within which the individual makes a conscious selection or choice of the good bought. Choice of this character expresses preferences on the part of the individual, which can be explained only on the assumption that the one good gratifies a keener desire than the other. Put in another way, the capacity of one

good to gratify a desire, i.e., its utility, is greater than that of another. Furthermore, as the number of units of a good increases, the capacity of each unit to gratify a desire declines. This principle is independent of the structure of economic society and is as extensive as human life.

The same conclusion may be drawn concerning the principle of diminishing productivity and the law of variable proportions. The physical productivity of land is also independent of the economic structure of society. While the extent to which land is utilized in the productive processes may be affected by the character of property rights that exist, the recognition of this fact has no bearing upon the observation that, after a certain point has been reached in the application of labor and capital to land, the returns in product may increase, but the increase will not be in the same proportion to that of the labor and capital used. This principle of diminishing productivity operates when any one of the factors of production is held constant and the other two are varied. The principle expresses a fundamental truth and is independent of the structure of society.

A similar conclusion may be expressed in connection with the law of variable proportions. It is always possible to attain an economic result by a variety of means. We may use a small amount of labor and capital on a large area of land, or the same amount of labor with more capital and less land, or more labor and more capital and less land, and get the same result in physical units of product. This fact of production is independent of the character of the economic structure. The principle is of general application, although its actual operation may be affected by the social order within which it works—that is, the laws of property may have a bearing upon the particular combinations of the factors used to get a definite result. But this fact does not change the nature of the principle as formulated. While the results of the application of these principles to human action cannot be measured with exactness, they do constitute fundamental tendencies that must be included in any attempt to develop a science in this field. Any attempt to predict human conduct, economically considered, must reckon with these tendencies.

After the general structure of economic society was described, and the effect of the chief characteristics noted, we proceeded to a study of the nature of value and its causes; of price and the relation of money and credit to the determination of price; and,

finally, to a study of distribution of income into the various shares. Throughout this discussion, the central theme has been value. The existence of value as an economic concept is dependent upon the fact of exchange. If there were no trading, there would be no value in the sense in which it has been developed in the previous chapters. The process of evaluation or, in other words, the subjective process of determining the relative significance of goods in relation to the desires for them, would be present and operating even though there were no system of exchanging goods. But as an economic phenomenon value is the result of the exchange of goods.

The term "value" has come to have other meanings of economic significance as a result of our experience in connection with the attempt to regulate and control the conflicting rights to wealth, as, for instance, in the field of public utility regulation. Here, there is a problem of social control and the endeavor is made to reconcile individual rights with a general interest of the users of the service. The "value of the property" for rate-making purposes is a common usage of the term. As an issue of practical social economy, this question of adjusting rights to wealth is of great significance, and the determination of the value of a property is fraught with many complexities concerning which there are many differences of opinion, not only as to what constitutes that value, but also as to the method to be used in its determination.

Another instance of practical social economy in which the term "value" is used, lies within the realm of taxation. Wealth may be assigned a value as a method of distributing the general expenses of government. It should be apparent that the significant problem here is uniformity in the value as between different types of wealth rather than the absolute sums assigned per unit. It is immaterial to the individual tax payer whether he pays a high rate on a low value assigned to his property, or a low rate on a high value. What is of greater significance is that all property be *valued* on a common basis. It is not a part of our purpose to go into these questions, however important they may be. An adequate discussion of these issues would carry one, as will be shown presently, into specialized fields of study. At this point, we are concerned only with the fact, that while the primary meaning of value is that of power in exchange, there are other uses of the term that have economic connotations with which the student should be familiar.

Then, too, in other realms of learning the term "value" has a usage that should be carefully distinguished from the specialized meaning in the field of economics. For instance, we might speak of the value of a given social policy, as that of conservation in which we are comparing the economic needs of the present generation with those of future generations. Or, we may think of ethical, religious, or artistic value in which case we are using the term to express a standard of worth in these respective fields. The term "value" is used in all of these realms and one should be aware of the differences in usage and know just what the economist means when the term is used in a modern text. Care must be exercised in reading the literature of economics because of differences in the concept of value and its usage by the various writers. In this text, the term is intended to mean power in exchange. If used in any different sense, the context will show what modifications of the fundamental idea have been made. Care has been exercised at all points to reserve value as an economic concept and to eliminate other philosophical uses of the term by substituting other words when the occasion seemed to demand it.

The principle of value runs throughout the whole range of economic activity. When dealing with prices and with money and credits, we were concerned with a special phase of the value principle. Price was defined as an expression of value in terms of money. An explanation of commodity prices involves the application of the value principle to money. But all types of money are not identical in character. Credit instruments perform a large part of the money work in all countries where banking has been developed. To explain commodity prices, then, required that the theory of value be extended to all forms of money, including all credit instruments used to perform money work. We went no further into the field of money and credit than was necessary to understand the nature of money and credit and to see the bearing of money and credit instruments upon the expression of value in terms of price.

We next considered the distribution of the income arising from industry. While the subject of distribution approaches the ultimate purpose of all economic activity, it also involves the application of the value principle. Consumption is the end of economic endeavor, but before wealth can be consumed, it must be divided. Individuals get their right to consume through

the receipt of some kind of income. In distribution, we were concerned with the forces that determine the shares of income received by the various agents that have participated in its creation. In practice, nearly all personal shares are received in the form of money income, and as the result of a contract of one kind or another. These contracts involve the services of land, labor, and capital, and the measurement of the social significance of these services involves the application of the principles of value. When one sells his personal services, the value principle is at work; likewise when one buys, or leases land, or borrows or lends capital, the principle of value enters into these transactions as a means of determining the payment to be made. These illustrations should be sufficient to show the pervasiveness of value as a phenomenon of our economic life. Even when we are dealing with the claims on wealth, as preliminary to consuming it, the value principle is continually at work in the processes that distribute wealth among individuals.

In most texts on economics, it is customary at this point in the discussion to pass over into a consideration of a variety of topics of practical social economy. The topics treated vary somewhat in accordance with the author's estimate of the social or practical significance of the subjects chosen. But it is common to find an extended treatment of the various systems of banking existing throughout the world; a few chapters on labor problems, trade unionism, workingmen's insurance, and other forms of labor legislation; an historical treatment of the regulation of railroads and other forms of combinations; and, in many texts, a treatment of public finance and taxation with a concluding portion devoted to the general question of social reform or proposals for modifying the present structure of economic society. In nearly all of these specialized topics there are refinements of the general principles of economics already developed. In most of them, there are *concrete issues that not only involve the application of these principles but illustrate as well the kind of modifications that have to be made when adapting the general formulas to particular economic situations.* These topics present what is frequently known as applied economics. But in these days of specialization in the development of economic principles, it seems much wiser to confine the textual material of the first course to a consideration of the simpler and more general principles of the subject, and to leave the more specialized topics for subsequent treatment in

other courses. But, before leaving the subject, it may be well to point out to the inquiring mind the direction which subsequent study may take and the way in which a few of these topics tie into the treatment of the more general principles.

Business Organization.—We have already seen that little wealth exists ready made for human use. Most wealth has to be produced in the sense in which this term is used in economics. The general idea of production is simply the creation of utility. But as a process, it means the organization of the factors of production in such a form and in such a combination as to transform the raw materials of nature into consumable goods. The productive process is mainly a private venture, and as a private business there are problems of internal organization that require a large amount of the time and thought of business executives. These problems become more complicated as the size of a business unit increases and, hence, the principle of specialization is likely to be developed here as in many other phases of economic activity. The purchasing of raw materials, the organization, the layout, and the operation of a factory, the selling of the product, the credit and collections, the keeping of adequate office records of business transactions, the finding of the necessary funds to operate the business, the maintenance of an adequate personnel and labor force—all of these problems present themselves in some form to the executive head of every business concern. The larger the business, the greater the likelihood that the responsibility for decisions will be delegated to executives who are then held accountable for results.

This whole question of the internal organization of a business has large economic consequences, and there has developed a specialized literature devoted to its various branches and problems which constitutes a special field of research and study. It is significant to note that this new field branches off from that of the general theoretical principles as a phase of the subject of production and is commonly known as business organization and management. The subject is especially significant from the acquisitive point of view, as every efficient entrepreneur is concerned with the question of how to get results with the least cost. The accomplishment of this purpose requires the most careful attention to the combination of the factors of production and involves a multitude of decisions on the part of the entrepreneur. We find in practice that the subject of business organization is

divided into fairly definite fields along functional lines, such as, factory management, sales management, labor management, office management, and other divisions of a similar character.

Marketing.—The primary purpose of a private business is the production of goods for sale. If any business concern cannot sell its products, there is no excuse for its existence. This fact accounts for the amount of energy that goes into the finding of a market by all business firms. Because of the significance of sales, most business establishments have organized a specialized department for the promotion and control of their selling policies. Special training is frequently given the salesmen, acquainting them with the qualities of the good to be sold and with the most effective methods of convincing the prospective buyer that he should place his order with a particular firm. In addition, advertising in all of its modern alluring forms is made a part of the function of marketing, and millions of dollars are spent yearly in this method of finding a market. As a result, there has grown up an extensive literature on this phase of business organization. Those who wish to understand the working of our economic system should familiarize themselves with this literature and its bearing on the marketing function.

Finance.—The business executive must find the capital funds to finance his operations, whether the form of organization be a single proprietorship, a partnership, or a corporation. The funds may be paid in by the owners themselves, or they may be secured, in part at least, by borrowing from others. This subject of financing is vital to all business and it may be studied from two points of view. First, the problems confronting the executive seeking funds, which are covered by a literature on business finance and more especially corporation finance; and second, from those of the investors who provide the funds, a subject that is treated under the topic of investments, which deals with the various methods used in financing a business and with the equities that are created thereby.

The customary method of securing the necessary money for the operation of a business is to borrow from lending agencies. If the business is a single proprietorship, a loan may be made and carried by an investment bank on the personal note of the owner, but it is more likely, especially in the corporate form of business, for securities to be sold. These securities consist of two different types, with many variations that cannot be

discussed in this text. The two principal types are stocks and bonds. A stock is merely paper evidence of partial ownership of the business and is a means of distributing entrepreneurial risks, as well as a method of raising capital funds. There are two principal kinds of stocks, namely, common and preferred. Common stock is one that merely conveys the rights of ownership with the rights to income if earned. All other obligations must be met before the common stockholders receive any benefits from the operation of the business. Ordinarily, the right to vote and thus to participate in the direction of the business is confined to the common stockholder, but this privilege is not always so restricted. The preferred stockholder may also enjoy this privilege—a right that is determined by the articles of incorporation.

The preferred stockholder as compared with the owner of common stock, has a prior claim over the income and sometimes over the assets of the company in case of a liquidation of the business. The preferred stockholder is paid a stipulated dividend, if earned, before the common stockholder can participate. Cumulative preferred stock carries over this right to income from year to year and the holders of this type of stock will receive their full stipulated share before the common stockholder can enjoy any income. This means that if the earnings in any one year are not sufficient to pay dividends on the preferred stock, the balance must be made up in subsequent years. It thus appears that preferred stock occupies a place between common stock and bonds and has some characteristics of each.

There are many variations in these simple illustrations of securities which cannot be discussed here. The financing of an organization is important to the investor and lender, as well as to the borrower. Familiarity with the rights and equities that lie back of a security is an important consideration for the purchaser of stocks and bonds. The field of investment has its own specialized problems, some of which are economic in character, while others are legal. The property rights involved are often complicated and technical in nature. To understand these equities thoroughly, requires a knowledge of accounting and of law in order that the records of the various business units may be analyzed and the ownership rights properly allocated.

Much that has already been said concerning stocks applies equally to bonds. A bond is of the nature of a mortgage. It is

a lien on a part or all of the property of a company, or it may be a lien on its income. Bondholders have prior claim on both the income and the assets of the business. They get a stipulated share, as 5 or 6 per cent, and ordinarily have no voice in the determination of the policy of the business. In the event that the interest payments are not met, the bondholders can force a receivership as a means of protecting their interests in the business. This whole phase of private business operations is now generally treated under the subject of corporation finance and investment. A very large and important literature has developed in this field and a thorough understanding of its technical phases requires specialized study.

Money and Credit.—Closely associated with the problems of finance that have just been presented is another phase of business that is also financial in character, namely, the questions of money and credit. The banking institutions of the country occupy a strategic position in the operation of modern business, as all transactions are carried on in terms of money. Banks are institutions that deal in money and credit. There are two main types of banking institutions, the commercial banks, or those that receive deposits and discount short-time bills of business concerns. These institutions are closely regulated in this country either by federal or state laws. Then, there are the investment banks that undertake to finance the long-time loans. These banks float bond issues, which means that they advance loans to borrowers on the basis of a mortgage or other pledge of property, and, on liens of this character, they issue bonds in convenient denominations which they sell to their customers. They may also underwrite a new venture by providing the necessary funds to start it, or reorganize a concern that has become financially embarrassed. Financial institutions of this character exercise a dominant influence in the operation of modern business.

In addition to the financial institutions that have already been mentioned, there are the great trust companies that make a business of managing estates, and the insurance companies whose combined financial resources mount into billions of dollars. These, together with the banking institutions, are the sources from which credit in the form of money loans is extended to business concerns, and the conditions under which the loans are made constitute what is commonly known as the money market. This phase of finance has, as can be easily seen, both a general and

a private significance. The accumulation of funds seeking investment is a great convenience to borrowers, but centralization of control over the extension of financial assistance in the form of credit to business activities is fraught with grave social consequences. Whether this power, exercised through the money market, is beneficent or not, depends to a large extent upon the integrity of the individuals who decide upon the disposition of credit. It is because of the power exercised by the financial institutions of the country that there has developed a general fear that finds expression in the condemnation of Wall Street, which is synonymous in the public mind with financial control. The legislation restricting financial institutions is evidence of this fear. This field of finance has also developed its specialized literature which may be followed further in the study of money, banking, and credit.

Personnel and Industrial Relations.—The operation of all business requires the services of labor. In the small establishment where only a few men are employed, this problem does not present many difficulties, but as the size of the plant increases the problems multiply. The questions of selection, training, and supervision become more complicated as the size of the plant increases. Since labor services cannot be delivered without the laborer working under the conditions provided by the employer, there is a human and social problem that does not ordinarily arise in connection with the use of any other form of wealth. As a result, the conditions under which the laborers work have long been recognized as worthy of general, social regulation. Labor laws are numerous and in many cases very complicated, involving conflicting rights and obligations of both the employer and the employee.

In addition to the regulation of the conditions of labor by law, there have grown up voluntary organizations, known as trade unions, that undertake to protect and promote the economic welfare of their members. We have seen that the actual wages received by a workman is the result of a bargain with an employer, and since the bargain involves a division of income between the two parties to the labor contract, this becomes a hotly contested field. The unions endeavor to negotiate with the employer over the conditions of employment through representatives of their own organizations. The conditions surrounding negotiations of this character are partly economic, partly psychological, and

partly diplomatic. There has developed here an exceedingly extensive literature devoted to the various phases of the subject. Because of the human factors, this whole subject of personnel and industrial relations has come to receive increasing attention, both as a problem of business organization and as a question of general interest. One can find in this specialized field an opportunity not only for the development of scientific principles as guides for human conduct, but also a chance to develop judgment in dealing with questions involving conflicting interests and in applying principles to this phase of social politics.

Accounting and Statistics.—The conduct of business requires records as a basis for determining the policy to be followed and as a method of measuring the success of its operations. These records may be kept in the head of the proprietor, if the transactions are few and simple, or they may require a large and trained staff, as in the case of corporations. Accounting and statistics deal with the records of a business. Both are necessary tools for the operation of modern business and for an understanding of its complicated transactions and relations. A specialized literature and a special technique have developed in both subjects which furnish, in addition to an excellent discipline in accuracy, an intimate insight into the structure, operations, and formation of policies in modern private business concerns.

Social Control and Regulation.—In the previous discussion, attention has been directed primarily to the private and acquisitive interest in production. In the development of the subject the conflict between the general and social points of view in industry was noted. It is because of this conflict that social control comes to be exercised over industry. When the spontaneous forces operating in the market become heavily weighted in favor of private interest, then the collective will of the social group is expressed in some form of regulatory control. If competition is not sufficient to keep the productive forces functioning in harmony with the general social needs, it gives way to some kind of regulation.

Regulation of industry on a large scale began in this country with the railroads, and has subsequently been extended to other public service industries, such as gas and electric power plants, telephone companies, electric railways, and other industries similar in character. It will be noted that in the operation of such industries there is a large element of monopoly. It was on

his account that they were the first to be brought under legislative control by both the federal and state governments. With the passage of the Interstate Commerce Act in 1887, our country shifted its traditional attitude toward private business and, since that date, there has been a steady growth of legislative interference and administrative control of private industry. To understand the nature of this change in policy a brief summary of the economic changes during the last century and a half will be necessary.

The industrial changes that took place during the last half of the eighteenth century, commonly known as the Industrial Revolution, fostered the acceptance of what was then a new philosophy of non-interference on the part of the state with private business. The new machinery and artificial power processes required greater freedom in combining the factors of production than had been accorded the private entrepreneur for a long time. The economists and statesmen of that period argued that competition was the true regulator of industry and, if unimpeded by the state, it would naturally direct industrial activities into those channels that would contribute most to the common welfare. Free competition and non-interference by the state became accepted as the proper administrative philosophy during the first three-quarters of the nineteenth century and is still the basis of much of our industrial law.

However, important industrial changes occurred during the past century that have limited the effectiveness of competition as a regulator of our economic relations. The railroads by greatly widening the market area, the corporate form of business organization, and the tremendous extension of banking and credit, have all contributed to make possible the formation of industrial concerns on a scale hitherto unknown in the history of the world. In many industries, monopoly began to reach proportions that threatened the general welfare, and created a new problem. To combat this new force in one field of endeavor, Congress reluctantly adopted the Interstate Commerce Act. Since then there has been gradually developing a large body of administrative law. The purpose of this law is to harmonize private and public interests in the regulated fields. It is an attempt to set limits by law within which the competitive forces may operate, and in some instances, to check the force of monopoly. Conscious regulation and control have been substituted for

the unconscious and spontaneous influences of self-interest and the freely operating forces of the market. Administrative bodies clothed with increasing power to control private interests in accord with the common welfare have been set up. A knowledge of economic principles, of accounting, and of law is essential for an understanding of the problems in this field. A very extensive literature has been developed that covers this whole range of socio-economic problems and constitutes a fruitful field for specialization and research that promises results of large social significance.

Trade Associations and Combinations.—Closely associated with the generally recognized public utilities is another fairly well-defined field, namely, that of trade associations and combinations. The line of cleavage between a public utility that is recognized as subject to public regulation, and the problems connected with many of the large industrial corporations that have not been so classified is not always a clearly defined one. The social consequences from the operation of such large concerns as those in the packing industries or the steel mills are much greater than those of a small gas or electric plant. The effect upon the economic welfare of a unified policy among large concerns that are producing similar products may be extremely significant. The policy pursued and the code of ethics governing their conduct toward each other and toward their customers may make or break the economic welfare of a country.

In this field, we have adhered to the doctrine of competition as the regulator of trade and have written it into the Sherman Anti-Trust Law of 1890 and with some modifications into the Clayton Act of 1914. The Sherman Act aims to prohibit all combinations in restraint of trade, and any attempt at price fixing among competitors is interpreted as restraint of trade. The difficulties experienced in the application of this legal theory to modern industrial conditions led in 1914 to the passage not only of the Clayton Act but also of the law creating the Federal Trade Commission. This body is an experiment in applying administrative control over industries that have not yet been declared to be public utilities, but have been feared because of their size and of the potential monopoly power that they might exercise. This experiment has run into serious difficulties in attempting to apply the theory of competition to an industrial condition in which the older philosophy can no longer operate

effectively. Here is a field that is growing in social significance. It requires an understanding of economic principles, of law, and of accounting. The whole question of combinations and trade association activities furnishes an opportunity for specialized study and constructive work.

Public Finance and Taxation.—We turn now from the field of private to that of public enterprise. Not all wealth is furnished by private business. It is thought that many needs can be better served by the state than by the individual, such as the maintenance of order, the provision for public highways, the postal system, the operation of the public school system and public parks, and those economic ventures that extend beyond the span of one generation, such as reforestation and other conservation activities. The maintenance of government is expensive. It calls for the raising of vast sums of money and the appropriation and expenditure of the same. This is the field commonly known as public finance and taxation and was one of the first subjects to receive the attention of early economists, and from it have come the beginnings of economics. The raising of a revenue and the administration of the same were the beginning of economic thinking. Not only is it socially significant to know where to raise the funds to support governmental activities and how these funds are expended, but the effect of taxes upon private business is of equal importance. The incidence of taxation, that is, the final resting place of any tax, is a question that requires the most rigorous analysis and familiarity with economic principles. This field, because of its long historical interest, has developed an enormous literature. It is of vital concern to every member of every community, but because of its complexities the subject has been neglected as a topic for specialized study. The public-spirited citizen knows in a general way of the need for expert opinion in the realm of public finance and taxation, but there is a dearth of those who know this literature thoroughly. Here, one will find an opportunity for the exercise of the most comprehensive grasp of history and familiarity with the principles of economics. The tax expert does not often occupy a position of social prestige, but he performs one of the most important functions connected with the collective activities of governments of all kinds.

Public Ownership and the Readjustment of the Social Structure.—Before closing this review of the opportunities for further study and specialization in the general subject of economics, some

attention should be given to the proposals for a modification of the fundamental structure of economic society. Among the proposals for a change are various forms of collectivism. A collectivist economic order would substitute some form of social ownership over the material instruments of production for our present economic system. Instead of land and capital instruments being privately owned, as now, they would be owned by some social group, the nature of which would vary according to the economic philosophy that is accepted as a guiding principle. There are several forms of collectivism, such as socialism, syndicalism, bolshevism, and cooperation. The socialists propose that the central governing agencies should own and operate all of the instruments of production. Private property would be confined to property in goods to be consumed by the individual owner. No sale for private gain would be permitted. Individual income would be derived from the sale of services to the various governmental agencies, much as postal clerks or other government employees at present.

The profit motive would be eliminated from the operation of industry and initiative exercised through the government would be substituted for private initiative and enterprise. The socialists have less to offer on the question of distribution than they have on the side of control over production. There are great differences of opinion among socialists as to the proper basis for distributing the income arising from industry, but needs occupy a larger place in most of their schemes than they do in the present system of distribution. There is nothing inherently inconsistent between social ownership of productive wealth and compensation in accordance with service rendered. The difficulty in following this principle of compensation under socialism would be to find a satisfactory method of measuring the value of services. But even more serious than this would be the effect that social ownership would have upon the productive energy of the rank and file of people. It is a matter of common knowledge that the drive in government activities is generally much less than in private business. Even in private business, especially in the large corporate concerns, the task of keeping up efficiency is a difficult one. If pursuit of money profits were completely removed from industry, it is difficult to see what other motive could be substituted that would have as constant a driving power. Perhaps, through a long period of education and adaptation, a people

might be developed who would respond to other forms of stimuli. But, certainly, no sudden change in the system of industry could hope to modify greatly the attitude of the present generation. Socialism has served a useful social purpose, however, by pointing constantly to the weakness of our present economic order and thus has paved the way for adjustments.

Syndicalism and bolshevism both advocate some form of collective ownership, but both differ from socialism. The syndicalist proposes to have the instruments of production owned by those who actually use them. The term comes from the French *syndicat*, which is the word for trade-union organizations. The men who belong to a *syndicat* and are employed in a given factory, or plant, would own the industry. The syndicalists have in recent years been regarded as much more radical in their proposals and in their methods than socialists. However, it is interesting to note that, if the proposals of some of the most conservative business men in this country, such as the sale of stock to employees, were carried to their logical conclusions, the result would be a form of syndicalism, or employee-owned industry. It is said that the employees of the Philadelphia Rapid Transit System now own one-third of the stock of this company. Perhaps this is a method that will eventually re-enlist the ownership motive in industry to the benefit of the whole social group. But here again the proposals are more vigorous and definite on the side of production than on that of distribution.

So far as bolshevism is concerned, we have the Russian experiment. The plan is a workingmen's organization of the whole structure of society which is divided into a hierarchy of directing social groups, or soviets. Industry, finance, production, and markets are all controlled and directed through these soviet organizations. Property is owned by the soviets and in this regard there is some similarity between syndicalism and sovietism, but bolshevism carries through the organization from local groups to a central government that controls and directs the general policy of industry. In this particular, the two forms of collectivism differ.

Cooperation is a voluntary joining of people to conduct some common project. There may be consumers' cooperation, or producers' cooperation. The former is a grouping of consumers for the purpose of saving the merchants' profits. Producers' cooperation is extending group ownership over the

productive processes for the purpose of controlling the producers' profit. As a scheme of social reform, the most significant experience is found in England. Cooperation there began as consumers' cooperation in what is known as the Rochdale Plan. This plan provides for the raising of the funds for operation by the sale of stock among the members. The number of shares per person is usually limited to five, and voting privileges are limited to one vote per member, irrespective of the number of shares owned. The sales are approximately at current prices and a rebate is allowed on the amount of purchases. This plan has had remarkable success in England. It has been extended from the field of retailing to wholesaling and also to the management and direction of the production of the goods sold. The annual business by the combined stores runs into millions of dollars. It is said that these cooperative establishments come into contact with at least one-fourth of the population of England. The cooperators have developed a high degree of loyalty to the system and many instances exist of men who have been willing to serve as managers of stores for remarkably low compensation. Where a spirit of loyalty to an idea has been firmly developed, the possibility of substituting some other system for the present economy is within the range of attainment. The English cooperators have amply demonstrated the possibility of this type of social change.

The proposals that we have just examined by no means exhaust the plans for modifying the existing order. We have argued that the present system exists solely on the basis of social expediency. There is nothing inherently right or wrong in the system, and when a different one is found that may reasonably be expected to perform the economic functions more effectively than they are now performed, or when weaknesses in the present system are discovered and satisfactory remedies proposed, the modifications or changes should be welcomed. Their adoption would contribute to the economic welfare of all. In the meantime, the existing system must be given the credit of having built up a great economic structure that daily furnishes us with the necessities and comforts of life. Before abandoning the present system, its merits and defects should be weighed against the merits and defects of any new proposal. Here is a field for study that will require the most comprehensive understanding of human nature and thorough familiarity with the history of the human family, as well as a constructive imagination to foresee the

results of proposed social changes before they have been inaugurated. This field leads into the whole realm of social organization and social economy, and needs the thought of the best minds that can be persuaded to study it.

In this general summary two things have been attempted. First, a brief survey of the subject matter previously covered by the discussion of the general principles that have been found to be operating within the structure of our economic system. Second, the more important phases of specialization that branch off from the general field have been suggested. No attempt has been made to make an exhaustive classification of the opportunities for specialization, but those mentioned conform fairly close to the ones that are commonly treated separately, both in the literature of economics and in the formal instruction given in the subject. Economic activity has become so extensive and complex that it is no longer possible to include a complete survey of its scope within the bounds of a single volume. However, all specialized fields, sooner or later come in contact with the general principles. One cannot go far in any of the subjects mentioned above without confronting a value problem. This is the theme that gives unity to the whole range of literature devoted to the economic activity of mankind.

Within these specialized fields are found the great social and applied problems of both private business and general social economy. The opportunity is offered here, whether it be in the realm of private business or in that of practical, social politics, for the exercise of intellectual qualities of the highest grade, as well as an appeal to the desire to render service to mankind. Those endowed with capacities for analysis and abstract reasoning may make their contributions to the science through generalizations that more accurately explain the phenomena of economic life. Those with strong desires for action and less for reflection can find ample opportunity for the exercise of their powers in the competitive game of business, or in the less active, but equally important, task of maintaining an operating harmony between the social and acquisitive interests in our economic system.

INDEX

A

- Abstinence in relation to interest, 455-458
- Accounting, 534
- Accounting and interest, 475-476
- Advances to labor, 493-497
- Advertising, form of production, 70-71
- Agents of production, 65-73
- Aldrich-Vreeland Act, 315
- Arbitrage, defined, 380
- three-cornered exchange, 379

B

- Balance of payments, 365-367
- Bank deposit defined, 296
- Bank discount, 293
- distinguished from interest, 471
- Bank notes, 272-274
- federal reserve bank notes, 318
- federal reserve notes, 273, 319-321
- Banks, investment, savings, and commercial, 291, 532-533
- number of private 1837, 305-307
- statements of explained, 295-299
- Banks of the United States, First, 305
- Second, 305, 314
- Barter, defined, 259
- disadvantages of, 259-260
- "Bears," defined, 152, 237
- Bequests, relation to private property, 30
- Balance of trade, meaning of, 378-379
- Bill of exchange, 374
- commercial bill, 375
- right and time, 376
- Bimetallism, defined, 282-284
- effect on prices, 282

- Black, J. D., 80, 125, 139, 370
- Böhm-Bawerk, 461
- Brassage, 269
- Bulk line costs, 225
- "Bulls" defined, 152, 237
- Burns, A. R. and E. M., 28, 66
- Business cycle, 102, 333-335
- causes of, 342-350
- four phases of, 336-342
- remedies proposed, 351-353
- social and economic effect of, 350-351
- Business economics, defined, 15
- distinguished from economics, 3
- nature of, 15-18
- Business organisation, 529
- Business profits, 508
- Bye, R. T., 46

C

- Cairnes, J. E., on field of economics as a science, 11-12
- Capital, defined, 72, 453
- demand for, 467
- diminishing productivity and, 467-468
- distinguished from land, 73
- effect of corporation savings on, 465-467
- formation of, 455
- goods, 75
- maintenance of, 459
- productivity of, 460
- rent and, 417
- social and individual, 453
- value, 75
- Capital goods, 453
- Capital value, 453
- Capitalization, the rate of interest and, 478-474

- Carver, T. N., 2, 4, 47, 84, 134, 208
 Chapman, S. J., 7
 Check, a substitute for money, 260
 Chicago Board of Trade, 169
 Clark, F. E., 106, 157, 161, 170
 Class price, 252-255
 Clay, H., 28, 87
 Clearing houses, 299
 certificates, 314
 Coinage system, 268-270
 gratuitous coinage, 269
 Combination, highest profit, 138-139
 horizontal and vertical, 109-110
 least cost, 135-138
 pools, trusts, holding companies, 110
 Commercial banks, 532-533
 Commons, J. R., 357
 Communism as a form of economic organization, 23
 Comparative advantage (*cf. Principle of comparative costs*).
 Competition, as a market influence, 200
 as an organizing force, 37
 combination *vs.* monopoly in relation to theory, 173
 fair, 38
 in relation to the development of theory 39-43
 perfect competition, 171
 Constant costs, 218-219
 monopoly and, 251
 Consumer's goods, defined, 51
 Consumer's surplus, 106
 Consumption loans determination of the rate on, 469-470
 Consumption, end of economic activity, 527
 means wealth-using 9
 Contract, effect on distribution, 412-413
 right of, 31
 Control influence of the state, 35-37
 kinds of, 34
 meaning of, 32, 33
 monopoly, 35
 regulation of private control, 37
 Cooperation as a form of economic reorganization, 539
 a feature of the present economic order, 27-29
 spontaneous, 29
 "Corners," effect of on value, 238
 monopoly and, 257
 Corporate saving, effect on interest, 465-467
 Corporation, advantages of, 120-123
 form of business organization, 119
 importance of, 121
 nature of, 119
 voting privileges in, 120
 Cost of production and value, 216-231
 Costs, bulk line, 225
 direct and indirect, 68
 distinguished from expenses, 210
 joint cost, 233-236
 kinds of, 211-212
 nature of, 210
 "overhead" or "burden," 68
 Credit, bank, 291-301
 effect on value of money, 301-303
 kinds of, 290
 nature of, 289
 supply of gold and bank credit, 303
 Creditors as risk bearers, 148
 Crises (*cf. Business cycles*).
 Currency, elasticity of under federal reserve system, 318-321
 kinds of, 271-274, 285-288
- D
- Decreasing costs, 227-231
 monopoly and, 252
 Demand, distinguished from desire, 182
 elasticity of, 204
 in a market and schedule sense, 201
 nature of, 182
 wages as a source of, 502
 Depletion cost, 227
 Depreciation, meaning of, 459
 relation to interest, 459

- Depression (*cf. Business cycle*).
 Desires, balanced, 53
 classification of, 47-49
 importance of, 45-46
 satiability of, 52
 Dewey, D. R., 268-306
 Dickinson, Z. C., 46
 Differential rent, 447
 Diminishing desirability, 187
 Diminishing average output, 127
 applied to labor, 481-485
 effect on the use of capital, 467
 highest profit combination, 138-139
 increasing costs, 212
 least cost combination, 135-138
 marginal output, 128
 productivity defined, 124-131
 relation to distribution, 407-409
 relation to rent, 422-424
 total output, 127
 a universal principle, 131-134
 Diminishing returns (*cf. Diminishing productivity*).
 Diminishing desirability, 187
 utility, 168-188
 Discount, bank, 293
 Distribution defined, 402-404
 economic *vs.* contractual shares, 410
 effect of legal institutions on, 410-413
 functional and personal distribution, 404
 general theory of, 409-410
 a phase of value, 405-407
 relation of diminishing productivity, 407-409
 Disutility, 210
 derivative disutilities in relation to land, 443
 Divisions of labor, contemporaneous, 82
 division of processes, 80
 forms of, 78
 importance of, 26, 76
 limits of, 77
 localization of industry, 90
 Divisions of labor, occupational, 78-80
 relation to money, 259
 successive, 82
 territorial, 81
 Dorau and Hinman, 421
 "Dumping," monopoly and, 254
 Dunbar, C. F., 20
- E
- Earned income, 410-411
 "Economic friction," 241-243
 Economic goods, defined, 51
 Economic principles, basis of, 5-7
 express tendencies, 5
 relation to economic structure, 524-526
 Economic order, autonomous, 23
 characteristics of, 25-39
 communistic, 23
 individual exchange, 24
 meaning of, 22
 socialistic, 24
 Economics as a science, 4
 method of treatment, 19-21
 nature of social economy, 19
 pure and applied science distinguished, 16-18
 a social science, 7
 Economics, business, 1, 3
 competition *vs.* monopoly as a basis for economic analysis, 173, 175
 defined, 1
 distinguished from political economy, 2, 3
 method of treatment, 19-21
 Economics, relation to ethics, 14-15
 physical and natural sciences, 10
 political sciences, 13
 psychology, 10
 sociology, 12
 Economy, meaning of, 3
 social economy defined, 19
 Edie, L. D., 33, 266, 351, 371
 Elasticity of demand, 204

- Ely, T. R., 1, 67, 125, 245, 246, 303, 333, 342, 405, 416, 443, 490, 514, 520
- Emery, H. C., 399
- Entrepreneur, as risk bearer, 147
 defined, 112
 forms of, 116
 functions of, 114
 influence of lender on, 115
 the place of management, 115
 profits and, 505
 selection of, 113
- Epstein, R. C., 94
- Equation of exchange, 281-282, 301-302
- Evaluation as an economic concept, 176
 relation to value, 176-177
- Expenses of production, 210
 items entering, 212-216
- Extensive margin, in measuring rent, 424
- F
- Fair value, concept of, 521-522, 526
- Fairchild, F., 202, 275, 333, 334, 493
- Favorable balance of trade, meaning, 381
- Federal reserve note, 273, 318-321
- Federal reserve system, 316
 note issue, 318
 other functions of reserve banks, 321-323
 rediscount privileges, 317
 reserve provisions, 316
- Fetter, F. A., 47, 464
- Fiat money, 270
- Finance, as a field for special study, 530-532
- Financial panics, 339
 relation of banks to, 339-340
- Fisher, I., 53, 59, 202, 281, 335, 353
- Forecasting, relation to risk, 146
 as a remedy for business fluctuations, 355
- Free goods, defined, 50
- Free trade, meaning of, 385
- Future contracts, nature of, 152-154
- G
- Gains of bargaining, 512 (*cf. Profits*).
- Gide, C., 5, 50
- Gold certificates, 271
- Gold dollar, as standard money, 268, 269
 weight of, 267
- Gold points, 377
- Gold, as standard money, 268, 269
 relation of the supply of gold to bank credit, 303
 stock of, 276
 use in international settlements, 380
 value of, 276-280
- Goods, capital, 75
 classification of, 50
 consumer's, 51
 defined, 49
 economic, 51
 free, 50
 producer's, 51
 public, 51
- Greenbacks, 270-287
- Gresham's law, 282
- H
- Hamilton, Alexander, 267
- Hamilton, W. H., 359
- Hardy, C. O., 145
- Hedging, effect on prices, 159-160
 defined, 155
 illustration of, 156-159
- Horizontal combination, 109-110
- I
- Immaterial wealth, 53
- Immigration and the supply of labor, 491
- Improvements, effect of on rent, 423, 437
- Incidence of tariffs, 398-400
 of tax on land value, 450-451

- Income, defined, 56
 - distinguished from wealth, 56-59
 - earned, 440
 - money and real, 58
 - net and gross, 58
 - psychic, 56
 - Income, economic *vs.* contractual shares of, 410
 - Increasing costs, 220
 - monopoly and, 221
 - Index numbers, defined, 326
 - kinds of, 327
 - Individual-demand schedule, 183
 - Individual initiative, 25-26
 - Individual proprietor as form of entrepreneurship, 116
 - Individual supply schedule, 198
 - Industrial depressions (*cf.* *Business cycles*).
 - Industrial relations, 533
 - Industry, localization of, 90-96
 - Inelastic demand, 204
 - Infant industry, argument for protection, 389-391
 - Inheritance, right of, 31
 - Insurance companies, deals with calculable risks, 515
 - relation to money market, 292
 - Insurance, method of distributing risk, 162
 - mutual *vs.* stock companies, 165
 - Intensive margin, in relation to rent, 426
 - Interest, ability to pay, 460
 - as a cost, 214, 474-476
 - as a functional share, 404, 453-455
 - determination of the rate, 461-469
 - economic and contractual distinguished, 454-455
 - problems in the determination of, 455
 - Internal and external economies, 229-230
 - International trade, basis of all trade, 262-264
 - favorable balance of, 378
 - International trade, methods of making settlement, 373-379
 - principle of comparative costs, 368-370
 - use of gold, 380
 - visible and invisible items in, 364-368
 - Investment banks, 291, 532-533
- J
- Jevons, W. S., on the cause of business cycles, 344
 - Johnson, A. S., 77
 - Joint costs, 233-236
 - Joint demand, 236-237
- K
- Keynes, J. N., 10
- L
- Labor, classification of, 479-481
 - defined, 67
 - demand for, 487
 - growth of population and supply of, 489-490
 - immigration and supply of, 491
 - productive and unproductive, 67-72
 - supply of, 488
 - Laborers, as risk bearers, 148-149
 - Land, classification of, 416
 - defined, 66
 - distinguished from capital, 73
 - investment in, 443
 - "made" land, 73, 442
 - marginal land, 424
 - how measure services of, 421
 - private property in, 437-440
 - services of, 419
 - value of, 74, 431-433
 - Large-scale production, 96
 - causes of, 98-103
 - combination and, 109
 - costs and, 230-231
 - economies of, 102-103
 - limits to, 103
 - Law of variable proportions, defined, 135-139

Legal institutions, effect on distribution, 410-413
 Legal tender, definition of, 275
 Limited liability, 121
 "Limping" standard, 284
 Loans, international, effect on balance of trade, 365-367
 Localization of industry, causes for, 90-96
 meaning of, 90

M

Macleod, H. D., 282
 Malthus, T. R., on theory of population, 490
 Manipulation, effects of, 161, 238
 meaning of, 160
 Marginal productivity theory, as applied to labor, 484
 definition of, 409-410
 implications of, 414-415
 Marginal rent, 446-450
 Marginal utility, 188-191
 as a market influence, 190
 Marginal vendibility, 195
 Market-demand schedule, 184
 Market-supply schedule, 199
 Marketing, as a division of economic analysis, 530
 Markets, as a coordinating agency, 167-168
 competitive, 171
 defined, 169
 functions of, 170
 imitation of, 170
 monopolistic, 172
 perfect, 171
 Marshall, Alfred, 20, 46, 53, 57, 62, 66, 178
 Middle men, and market operations of, 168
 as risk bearers, 160
 Mill, J. S., 53, 67
 Mines, rent of, 433-434
 Mint Act, 1792, 267
 Mint ratio, 267, 283-284
 de Monchretien, Anto

Money and credit, a field for special study, 532-533
 Money, defined, 260
 functions of 260-264
 good money material, 264-266
 kinds of, 270-275
 monetary system, 266-268
 stock of, 272
 value of, 276-281
 Monopoly, classification of, 244-249
 defined, 243
 monopolistic market, 172
 monopoly vs. competition in relation to economic theory, 173
 natural monopoly, 247
 social monopolies, 245
 Monopoly price, 255-257
 Monopoly profit, 519-520
 Monopoly value, 249-255
 Motives, desires as motives, 45
 profit as, 516-519
 self-interest, 46

N

National banks, 307-308
 decentralized, 312
 inelastic currency, 309
 proposed remedies, 313
 strength of, 308
 surplus reserves, 310
 National dividend, 404
 National Monetary Commission, 315
 Negative interest, 463-464
 Non-competing groups, 479-481
 Normal value, defined, 231
 relation to costs, 232-233
 Notes, banks, 272-273
 federal reserve, 273
 promissory, 289, 290
 Nystrom, P. H., 106

O

Obsolescence, defined, 460
 Open-market operations of the federal reserve banks, 321

- Ophelimity, suggested in place of utility, 50
- Opportunity cost, 211
- Organization and wages, 497-499
- Outerbridge, A. E., Jr., 84
- Overhead cost, 213
- P
- Pain cost, 211
- Panics (*cf. Business cycles*).
- Paper, commercial, 293
- "single-name," 293
- "two-name," 293
- Paper money, kinds of, 270-274
- irredeemable, 285
- Par of exchange, 374-377
- Pareto, 50
- Parker, Carleton, 46
- Partnership, advantages and weakness of, 118
- form of business organization, 117
- Phillips, Chester A., 299
- "Pittsburgh plus," 95
- Political economy (*cf. Economics*).
- Political science, field of differentiated from economics, 13
- Population, growth of, 489-490 (*cf. Malthusian theory*).
- supply of labor, 489
- Price, causes for the changes of, 325
- defined, 179
- distinguished from value, 179
- level of, 280
- Price level, relation to money, 280
- Principle of comparative costs, 368-370, 384
- Principle of variable proportions explained, 134
- highest profit combination, 139
- least cost combination, 135
- marginal choices, 139
- Private enterprise, a characteristic of the present economic order, 25
- Private property, effect on distribution, 412-413
- importance of, 30
- Private property, property in land, 443
- Producer's goods defined, 51
- Producer's surplus, 431-433
- Production, defined, 9, 64
- diminishing productivity, 124-134
- factors of, 65-75
- large-scale, 96-109
- risk in, 142-144
- Profits, as a functional share, 515-516
- as a motive in industry, 516-519
- effect of competition on, 504-505
- losses, 513
- monopoly and, 519-520
- nature of, 507
- ordinary, business, and pure, illustrated, 507-509
- pure profits defined, 509
- regulation of, 520-521
- relation to risk, 514
- source of pure profits, 512-513
- wages of management and, 509-511
- Property, relation to wealth, 59-61
- Protection, meaning of, 385
- natural, 386
- specific and ad valorem duties, 385-386
- tariff for revenue defined, 387-389
- Protective tariffs, beginnings of, 388-389
- by-products of, 400-401
- difficulties in removing a tariff, 397-398
- diversified industries, 394
- equalization of costs, 396
- home market argument, 392-394
- incidence of, 398-400
- infant industries argument, 389-392
- labor arguments for, 394-396
- political argument for, 397
- Public finance, a field for special study, 537
- Public goods, defined, 51
- Public ownership, 537-538
- Psychic costs, meaning of, 210

Psychology and economics, 10
 method of measuring objectively a
 mental state, 11
 nature of desires and motives,
 10, 11

R

Rate of interest, capitalization and,
 473-474

determination of, 461-469
 different types of, 471-472
 quantity of money and, 471

Reasonable rates, concept of, 521-
 522

Rediscount, 317-318

Reserves, bank, 295, 308

federal reserve provisions, 316-317

Regulation of economic activity,
 534-537

Regularization of industry, 355-357

Rent, as a cost, 215

as an earned income, 440-445

as a functional share, 404

as a marginal cost, 448-450

capital and, 421

capitalization of, 421

defined, 416

diminishing productivity and, 422
 effect of growth of population on,
 435-437

effect of capital expenditure on,
 442

extensive and intensive margins,
 424, 426

formula, 425-426

how measure, 421

location and, 421

marginal land and, 424

private property in land, 437-440

relation to price, 445-446

relation to value, 425

taxation of, 450-451

uses of, term, 418

Rent of ability, 216, 430, 450, 457

Rent of mines, 433-434

Representative firm, 222, 225

Representative money, 271

Risk, assumption of, 147-149

causes of, 144-145

insurable risk, 164

involved in production, 142-
 144

methods of dealing with, 145-166

methods of shifting, 149-162

nature of, 142-144

non-calculable uncertainties, 514-
 515

relation to profits, 514

Ricardo, definition of rent, 417, 431

theory of wages, 491-492

Ripening costs, meaning of, 444

Robbins, Lionel, 222

Royalty, distinguished from rent,
 416

S

Saver's surplus, 467

Saving, corporate, 465-467

distinguished from hoarding, 458
 in relation to capital formation,
 456

Saving wage, 502

Say's law, 361

Scarcity, causes of, 208-210

definition of, 51

relation to value, 178, 208

Science, economics as a science, 4
 meaning of, 4

Secrist, H., 35, 107, 214, 221, 222,
 224, 326, 328, 329

Secular trends, 335

Seigniorage, 269

Self-interest, as a motive, 42

"Selling short," meaning of, 152

Services, 53 (*cf. Immaterial wealth*).

Silver certificates, 271

Silver dollar, weight of, 267

Single tax, effect of, 451-452

meaning of, 451

Site rent, 420-421

Smith, Adam, 2, 83

Social control of economic activity,
 534-537

Social economy, defined, 19

Social monopolies, 245
 Social reform, economic structure
 537-541
 Socialism, as a form of economic
 organization, 24, 538
 Sociology, field of and economics
 differentiated, 12
 Specialization, advantages of, 82-85
 defined, 76
 develops international depend-
 ence, 371
 evils of, 85-89
 functional, 82
 limits of, 96-98
 phase of the division of labor, 26
 Speculation, "corners," 238
 effect on value, 237-239
 nature of, 237
 "Spendthrift" loans (*cf. Consump-*
tion loans).
 Standard of living and wages, 499-
 502
 Statistics, 534
 Stabilized dollar, 353-355
 State, the, limits private initiative,
 36
 in relation to industry, 35
 renders services, 36
 Steuart, Sir James, 2
 Subsidiary coins, 274
 Subsistence theory of wages, 491-
 492
 Supply, actual supply, 197
 as a market influence, 199
 in market and schedule sense, 202
 nature of, 197
 potential, 197
 Surplus, producers', 431-433
 Syndicalism, 539.

T

Tariff, relation to monopoly, 246
 Tariffs (*cf. Protective tariffs*).
 Tausig, F. W., 96, 97, 99, 101, 106,
 109, 195, 221, 230, 235, 239,
 264, 274, 330, 493
 Tawney, R. H., 412

Taxation, a field for special study,
 537
 Taxation and rent, 450-451
 Taylor, F. M., 23, 29, 82, 120, 157,
 197, 361, 414, 443
 Taylor, H. C., 430
 Terminal markets, relation to risk
 bearing, 151-162
 Time preference, in the determina-
 tion of the rate of interest,
 464-465
 meaning of, 464
 Thorndike, E. L., 46
 Total utility, 196
 relation to consumer's surplus, 196
 Trade association, regulation of,
 536-537
 Trust companies, relation to money
 market, 292

U

Undivided profits, 296
 Unearned increment, meaning of,
 440
 rent as, 440-445
 Urban land, services and rent of, 427
 Unfavorable balance of trade, 381-
 383
 Utility, defined, 49-50
 diminishing utility, 186-188
 marginal utility, 188-191
 methods of creating, 64-65
 relation to value, 178
 total, 196

V

Value and evaluation, distinguished,
 176
 Value, constant cost and, 218-219
 decreasing costs and, 227-231
 defined, 175
 distinguished from prices, 179
 fair value, 521-522, 526
 fixed supply and, 216
 general rise of, 181
 increasing cost and, 220

- Value, inequality of income and value, 191-193
 intrinsic value, 180
 joint costs and, 233-236
 joint demand and, 236
 monopoly and, 249-255
 relation to price, 61-62
 relation to utility, 178
 relation to wealth, 61
 speculation and, 237
 temporary and permanent differences in cost, 221
Vanderbive, H., 13, 147
 Vertical combination, 109-110
- W
- Wages, argument for protection, 394-396
 as advances, 493-497
 as a functional share, 404, 478
 as a source of demand, 502
 defined, 478
 diminishing productivity and, 481-485
 economic and contractual wages distinguished, 485
 effect of organization on, 497-499
 non-competing groups, 479-481
 saving wage, 502
 standard of living and, 499-502
 subsistence theory of, 491-492
 wages-fund theory, 492-493
 Wages of management, as a cost, 215
 a form of wages, 486-487
 relation to profits, 509-511
 Waiting, effect on interest, 455-458
 relation to capital formation, 455
 Wants in relation to industry, 9
 Watson, John B., 46
 Wealth, characteristics of, 54-55
 defined, 53
 individual and social, 56
 material and immaterial, 53
 relation to income, 56-59
 relation to property, 59-61
 relation to value, 61
 services and, 53
 Weld, L. D. H., 106, 107, 108
 Willcox, W. F., on growth of population, 490

