TEXT FLY WITHIN
THE BOOK ONLY

UNIVERSAL
LIBRARY

OU_152857

UNIVERSAL
LIBRARY

OSMANIA UNIVERSITY LIBRARY

Call No. 330.1 W35P Accession No. 28488

Author Waugh.

Title Principles of Economics

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

PRINCIPLES OF ECONOMICS

Principles of ECONOMICS

by Albert E. Waugh

PROFESSOR OF ECONOMICS
UNIVERSITY OF CONNECTICUT

First Edition
Second Impression

NEW YORK AND LONDON

McGRAW-HILL BOOK COMPANY, INC.

1947

PRINCIPLES OF ECONOMICS

COPYRIGHT, 1947, BY THE MCGRAW-HILL BOOK COMPANY, INC. PRINTED IN THE UNITED STATES OF AMERICA. *All rights reserved. This book, or parts thereof, may not be reproduced in any form without permission of the publishers.*

THIS BOOK IS

LOVINGLY DEDICATED TO

Alice Vail

Mary Holbrook

Edith Stewart

PREFACE

This book is planned and written for the student who wishes to get a good general understanding of our economic system and the way it works. The author has tried to write for college freshmen and sophomores rather than for his colleagues, and to keep the expression simple and the style easy to follow. The student who goes on to more advanced courses in economics and business will find the treatment here of such a nature that he has nothing to unlearn; yet it is planned for the student who is seeking general education rather than for the specialist.

No single beginner's text can possibly include all the interesting and exciting and important things about economics. The most difficult part of planning this text has been that of deciding what to exclude. In this connection it has been apparent that the student in an elementary course has a right to two things: a description of the most important economic institutions, phenomena, and problems that surround him; and an understanding of the nature of economic analysis and reasoning. Some elementary textbooks slight one and some the other. Some are purely descriptive, failing to give the student any adequate conception of the interrelationships of the various parts or why things work as they do. Others are purely theoretical, including little or nothing aside from exercises in a specialized sort of logic which, when not applied, is of little interest to most people. In either case, the student is defrauded. This textbook attempts to do both jobs, giving the student some intelligent knowledge of the more pressing of the current economic problems that surround him and, in addition, some basis for thinking intelligently about them and for understanding the nature of the economist's approach to problems.

After over 20 years of college teaching, the author is convinced that nothing is included here that is too difficult or abstruse for any student who is qualified for work at the college level and who can give a reasonable amount of time to it. On the other hand, he would be greatly concerned indeed if the book were pitched at such a level that at least some parts of it would not challenge the careful thought and attention of even the better students. The book is intended for college men and women and assumes that its readers will bring to it the intellectual equipment that should be expected at that level.

The chapters in this book differ materially in length and difficulty as a natural result of the fact that the topics considered differ in complexity and importance. The system of numbering the subsections should make it easy, however, to spread some chapters over two or more assignments while lumping two chapters together in a single assignment in other cases. The questions at the ends of the chapters should be thought of as integral parts of the work. Many of them raise problems not directly treated in the body of the text, yet problems on which the student who has read the text carefully should be able to develop a reasoned opinion.

Since economics deals with quantitative ideas, there has been a tendency in recent years to formulate economic postulates in mathematical terms. It has been difficult to know how best to treat this aspect of the presentation since, while mathematical treatment is undoubtedly both easier and more precise for the student who has the necessary background, it is terrifying and confusing to teachers and students who are not familiar with it. The problem was finally handled by means of a set of notes in a mathematical appendix, handily arranged for use by those students and teachers who can profit by it, but so segregated that others can omit the mathematics without breaking the continuity of thought. Students being what they are, instructors will be wise to assume that the Mathematical Appendix has not been read unless it is specifically assigned.

Sections of chapters are numbered serially within the chapter, the number showing both the chapter and the section; for example, a section numbered 18·6 is the sixth section of Chap. 18, while Sec. 6·18 is the eighteenth section of Chap. 6. This method of numbering makes it easy for the instructor to assign parts of chapters if he wishes to break them up for more detailed study, or to omit material if he cares to do so. Similarly, the figures and tables are numbered by chapters, with Table 4·7 being the seventh table of Chap. 4 and Figure 9·2 being the second figure of Chap. 9.

The author is indebted to far too many of his teachers, colleagues, and students for anything like a complete listing here. He has also naturally been influenced by many of the textbooks used in his teaching in the past. In many cases, he cannot recall where first he came across ideas incorporated here. Especially, indebtedness should be recorded to Dr. Edward Hugo Gumbart who, although little known in professional economics circles beyond the walls of his own classroom, was nevertheless a peerless teacher and a gentleman. During the author's first teaching experience, he shared the office of Dr. Gumbart, whose patient and sympathetic understanding and amazing breadth of interest have left their mark.

Dr. S. Howard Patterson of the University of Pennsylvania's Wharton School and Prof. Everett W. Goodhue of Dartmouth College have read

the entire manuscript of this book and have been unusually generous in the willingness with which they have shared their experience with the author. In many cases, the text has been altered to take advantage of their suggestions. It is a privilege indeed to have had the benefit of their long experience and to have this opportunity of acknowledging with thanks the aid that they have given.

ALBERT E. WAUGH

STORRS, CONN.,
July, 1947.

CONTENTS

PART ONE. *Introduction*

1. Wants and Scarcity 2
2. The Nature of Economics 10
3. Economic Development 16

PART TWO. *Consumption*

4. Consumer Incomes 40
5. Utility 57
6. Demand 69
7. Elasticity of Demand 87
8. Control of Demand 108

PART THREE. *Production*

9. The Organization of Business 124
10. The Nature of Production 158
11. Natural Resources 186
12. Human Resources 202
13. Capital 236
14. Variations in Cost; Supply 247

PART FOUR. *Exchange*

15. Exchange; Competitive Market Price 272
16. Monopoly Price 294
17. The Relation of Price to Cost; Normal Price 324
18. Money 352
19. The Commercial Bank 392
20. The Banking System of the United States 426
21. The General Level of Prices 457
22. The Business Cycle 474
23. International Trade 505
24. Foreign Exchange and the Balance of Payments 543

PART FIVE. *Special Problems*

25. Problems of Monopoly 566
26. Public Utilities and Transportation 604
27. Problems of Organized Labor 636
28. Social Security 670
29. Economic Problems of Agriculture 686

PART SIX. *Distribution*

30. The Nature of the Distribution Problem 718
31. Economic Rent 738
32. Wages 760
33. Interest and Profits 779

PART SEVEN. *The Government and the Economic System*

34. Public Expenditures 810
35. Public Revenues 828
36. Types of Economic Organization 858

MATHEMATICAL APPENDIX 883

INDEX 919

PART ONE. Introduction

1. WANTS AND SCARCITY

1·1. Human Wants. Great thinkers have taught from earliest time that "man doth not live by bread alone," and philosophers have urged men to put thought of the afterlife above thoughts of the physical needs of today. Yet, despite this continued advice, most men seem to spend a large part of their lives in efforts to satisfy their wants. The reader probably cannot think of anyone within his circle of acquaintances whose wants are entirely satisfied—who does not want some things that he does not have. To be sure, this does not indicate that the philosophers have been mistaken. On the contrary, it may well be evidence of the importance and truth of their teachings. Be that as it may, it will pay us here at the outset of our study of economics to inquire somewhat carefully into the nature and peculiarities of human wants and their relationship to human activity.

1·2. Individuality of Wants. First, we might note that no two people seem to want exactly the same things. Some like oysters, and some find them positively revolting. Some men wear bow ties, while others would be ashamed to appear in public with them. Some men wear beards, while others go clean-shaven. Some like music; others are tone-deaf. One man drinks whisky, another vodka, another champagne, and another milk. Some like to live in the city, others in the country. And so it goes. No matter what general sort of human want we consider, whether for food, clothing, shelter, amusement, sport, or religion, there are differences in desires. The saying goes that, "What is one man's meat is another man's poison." We even pride ourselves to some extent on these differences, arguing that men should not all be cast in the same mold, but that they should develop their own personalities and their own individualities.

These differences in human wants are naturally most noticeable when we compare people from widely separated parts of the world, who were brought up in different cultures. We remember from our grammar-school geographies the pictures of the strange costumes worn by the people of some foreign lands, or of the unfamiliar kinds of houses in which they live. We

tend, in fact, to typify people of various nations by their real or imagined food preferences, calling some of them "frogeaters," and ascribing to others a fondness for beer and sauerkraut. We recall that the people of some localities eat snails, or snakes, or dogs, or soup made from birds' nests, or locusts, or their next-door neighbors. Some live in igloos, some in huts raised high on slender poles, some in temporary tents, and some in proud penthouses. But, while the differences in human wants are most striking when we compare people of differing cultural backgrounds, nevertheless differences in wants are still there when we compare two people from the same city, from the same block, or from the same family.

Individual differences in wants may arise from differences in climate, from differences in age, from differences in sex, from differences in education, and from many other sources some of which are fairly well understood and some of which are but vaguely sensed and poorly comprehended. We do not expect a six-year-old boy, a sixteen-year-old youth, and a sixty-year-old man to want the same things. We do not expect a college senior to enjoy exactly the same things that are enjoyed by a freshman; we would feel that the educational processes had gone awry if no change in desires were evident with the passing of the college years. Certainly as we start to study the nature of human wants, it is important for us to note these individual differences and to make allowances for them.

1·3. Conformity in Human Wants. Yet paradoxically enough, having noted the differences in human wants, we next turn to the uniformity or similarity or conformity among individuals in the things they want. Although it is true that men do not all want exactly the same thing, yet *within any given community* their wants are remarkably uniform. It is probable that no one in your home town eats locusts or ants, despite the fact that such things are there for the asking and that in some parts of the world they would be considered delicacies. As a matter of fact, the various families in your home town all eat about the same sorts of things at the same meals. Even the order in which the foods are served is fairly uniform. The kinds of clothes worn by people of any particular age or sex on any particular kind of occasion are determined in advance, so that you feel that certain clothes are "proper" and other clothes "improper" for the occasion. The furnishings of our homes are far more uniform than they would be if governed purely by chance. Even the uses of spare time are similar throughout a community, so that in some places people play cricket and in other places baseball, etc.

When we consider all the possible ways of getting nourishment, we would hardly expect to know two people who ate the same combinations of food if chance alone were the determining factor. Yet in our communities it is common to find people eating potato for dinner and less common to find them eating rice; it is common to find them eating wheat bread and

uncommon to find them eating oatmeal bread; it is common to find them eating lettuce and uncommon to find them eating maple leaves; it is common to find them eating beef and uncommon to find them eating horse meat.

1·4. Source of Human Wants. We are accustomed to thinking of our wants as something born in us. The tendency is to think of them as given and to imagine that economic problems consist in finding ways to satisfy these given wants. But the similarities in the wants of different people, which we noted in the preceding section, lead us to realize after more careful reflection that our wants are not inborn. Rather they are made for us by the society in which we live. They are environmental rather than hereditary. Had we been born into and brought up in some other society, we could have been taught to eat snakes with relish or to wear togas and sandals; we could have been taught to abhor beefsteak and strawberry shortcake.

Although it is most certainly and obviously true that men require such broad classes of things as nourishment, clothing, and shelter, it is just as true that the particular foods that we use to furnish our nourishment or the particular kinds of clothing that we use to protect ourselves from the elements are by no means necessary. The need for nourishment was born in us; the desire for beef rather than horse meat was built into us by our families and our neighbors. Nutrition is a matter of biochemistry, but food is also a matter of social acceptance. Some things that are nutritious are unacceptable as food in our society, yet they are completely acceptable as food in, let us say, cannibalistic societies. The difference between the situations is not a physiological one, but a social one.

1·5. Needs and Wants. This leads us to distinguish between what we may call "needs" and what we have been calling "wants." It is common for us to speak loosely of some things as luxuries and of other things as necessities. But if we are willing to talk in terms of actual commodities that are bought and sold on the market, we find that there are very few necessities indeed if we mean commodities that are actually needed for the support of life. To be sure, as was pointed out in the preceding section, men need nourishment—but no man needs wheat bread, no man needs French-fried potatoes, no man needs a pair of low tan sport shoes, and no man needs a convertible sport roadster. Men have lived long and happy lives with none of these things. It would be a good lesson for most students were they to try to make a list of actual commodities which men really need and without which they cannot get along. Even those things which seem at first glance to be absolutely necessary will be found on more careful study to have substitutes. There have been many men, for example, who have lived for years without drinking water. Men's wants are very numerous, but their needs are few indeed.

1·6. Wants Differ in Intensity. We have noticed that different people want different things and that in any given community the wants of people are very similar. We should now call attention to the additional fact that the wants of any particular person for different things differ greatly in intensity. Some things he wants very greatly, and he would feel a distinct sense of frustration if he were deprived of them; other things that he wants are really but minor matters and, if he has to get along without them, he will not be seriously handicapped nor downcast. For example, the typical student may feel desires both for shoes and for a sporty automobile. In fact, he may fancy in his more thoughtless moments that he would prefer the automobile to the shoes. Yet, if he had to go to class barefoot, he would surely feel conspicuous and uncomfortable and humiliated and very unhappy; while the chances are that he can walk to class and not really feel greatly upset over his lack of an automobile. To be sure, if he had the automobile, he could exchange it for several pairs of shoes, and in that sense it would be wise for him, if offered his choice, to take the automobile. But as for his appraisal of the two desires themselves, it is probable that he would rather have the use of shoes than the use of an automobile. Surely, even if this illustration does not fit the reader's case, he will agree that his wants for different things differ in intensity, and that he wants some things much more than others.

Perhaps this difference in intensity of desires is fortunate, for if a consumer wanted many things with equal intensity he would be like the poor ass in the fable which starved to death between two baskets of corn because he couldn't decide which to eat first. In real life, wants are so numerous and resources so meager that it is a real blessing to have many of the wants relatively unimportant and only a few of them urgently insistent.

We shall find it necessary at a later point in our study to inquire about the factors that influence the intensity of our desire for various commodities, but in the meantime it will be sufficient for us to note in a preliminary way that such differences in intensity do exist.

1·7. Wants Change through Time. Next our attention should be called to the fact that human wants are not static—not unchangeable. The things that men want today are in great measure different from the things that their grandparents wanted. This is perhaps most noticeable in the case of what we call "style goods," such as women's hats and dresses. Here we realize that last year's hat, although physically in perfect condition, may well be unacceptable. In other fields the changes in taste and desire may not be so sudden nor so spectacular; but they are nonetheless present. The typical kinds and quantities of food eaten in the United States today are different from those of two generations ago. The houses differ in plan and in furnishings

and equipment. The amusements shift from five hundred to whist to auction bridge to contract bridge. At one period the desired sermon lasts an hour and a half; at another time a half hour seems ample. Then, of course, there are new goods added to the list for consumption, which our forefathers could not have at all—electric refrigerators, radios, television, etc. One of the great difficulties in satisfying men's wants is that, after you organize a production system to make the things that men desire, men no longer want the same things. "Variety is the spice of life," and in consumption it appears to be true that consumer whims call for constant and usually unpredictable change. Although prediction is always hazardous in any field, nevertheless it seems fairly safe to say that our children will want things that we did not have at their ages, and that those of the older generation who do not understand the facts of change will then decry the "foolish whims" of that younger generation as uncomprehending elders have always done.

1·8. The Variety of Wants. Next let us turn our attention briefly to the immense number of human wants. If it would not take so long, it would be informative for the reader to try to make a list of all the different commodities that he actually consumes. To be sure, these would make up but part of the commodities that he wants, since most of us have many unsatisfied wants. Yet, even if we confine ourselves to the things that we consume, we find dozens of different kinds of food, many different kinds of articles of clothing, household goods of various descriptions, etc. The list will surely run into the hundreds. The comment was made a few pages back that human needs are very few. Now we note that in contrast human wants are very numerous. In fact, there seems to be no limit to their number. When we find means to satisfy one set of wants, another set springs up to take their place as the new heads of the Lernaean hydra replaced those cut off by Hercules. Men long since found ways to produce all those things which their ancestors once desired, until even poor men of today have better food and better clothing and more comforts than had wealthy men a thousand years ago. Yet we do not find men contented and satisfied as a result. Rather we find that new wants have been added to the old ones. Those who talk glibly of an "economy of abundance" are overlooking a fundamental fact of human nature, namely, that if you give men what they want you do not satisfy them but merely make it possible for them to develop additional wants.

Of course, the philosophical student may well consider whether or not men want their wants satisfied—whether or not a life would be interesting and worth while if every want were satisfied without a struggle or effort. Professor Frank Knight has long since pointed out that men enjoy struggle; that, if they were trying to carry a football down the field with the least effort, they would put all 22 men on the same side. It is difficulty that makes the game exciting and interesting and, where games are too one-

sided, we purposely impose handicaps to make winning more difficult. The student with a speculative turn of mind may well wonder if complete want satisfaction would be desirable. For our purposes we need not pursue that thought further. We merely note the fact that no society to date has been faced with that problem and that in every society there have been almost numberless unsatisfied wants.

This, too, makes the economic problem a hard one. We enumerate our wants and organize a system to produce the goods necessary to satisfy them. If we succeed, we may find, as noted in the preceding section, that the wants have changed until we no longer want the same goods. Even if the wants have not changed, we now see that men will not be satisfied, because they will want the old goods and new ones in addition. Trying to satisfy men's wants by producing goods is like trying to walk to the horizon: one walks and walks until he gets to the point where the horizon used to be, but he finds that the horizon has receded and is still as far away as ever. Similarly, men work and work to produee goods to satisfy their wants and ultimately succeed in producing the goods that they desired at the start; but by now they have developed new wants, and complete satisfaction is as far away as ever.

Some economists have dignified this principle by formulating it into a law, which they call the "Law of the Variety of Human Wants": *There is no limit to the number of different commodities that men want, since as old wants are satisfied new ones appear.* The common naïve statement that "modern industry has multiplied production possibilities so that an efficient organization of production and distribution would make it possible to turn out an abundance of goods for all" overlooks entirely the basic facts of human nature, at the same time making the implicit but arguable assumption that such a situation would be desirable if possible.

1·9. Scarcity. Having noted some of the characteristics of men's wants, we may well ask ourselves why these wants are not all satisfied. Looking around us, we discover that hardly any of the things that men want appear at all in nature. Nature produces no automobiles, no shoes, no fried eggs nor baked potatoes, no phonograph records, no houses. To be sure, some things, like wild strawberries, are found in nature all ready for human consumption. But even in those cases, the quantities furnished are far too small to satisfy all wants. We would have hardly any of the desirable things of life if we were to rely on nature. We find that it is necessary for men to work to increase the quantities of goods.

There has been a tendency in recent years to talk about overproduction —about surpluses. To be sure, there have been times when enough wheat or cotton has been raised so that producers could not sell their output at a profit. But never in history has there been a time when we have produced

so much wheat or cotton that all wants for them have been satisfied. Always there have been people who were ill fed and ill clad. What people have glibly called an "overproduction" of wheat or of cotton is really overproduction only if we use the term in a very specialized sense.

The economist says that a commodity is "scarce" if there is not enough of it to satisfy all wants for it. If we define scarcity in this way, it is correct to say that almost all the things that men desire are scarce—some people would like to have more of them than they do have. If there are any unsatisfied desires for a commodity, the commodity is scarce.

1 · 10. Causes of Scarcity. If we look a little further below the surface and ask why so many commodities are scarce, we realize that it is in large part the result of limitations in natural resources, or to what economists have often called the *"niggardliness of nature."* Whether we like it or not, there are many natural resources that are found in relatively few areas. Most localities have no gold, silver, copper, or tin deposits. Even the resources that seem at first sight to be plentiful turn out on closer examination to be scarce. For example, what of the supply of arable land? Our grade-school geographies taught us that the area of the earth's surface was approximately 197 million square miles, which seems as though it ought to be plenty. But most of this area is covered by oceans, leaving but 51 million square miles of land surface, or roughly 33 billion acres. With the world's population approximately 2 billion persons, we thus have about 16 or 17 acres apiece. But this includes not only good land, but desert, stony land, mountainous land, jungle, and land in the polar regions. There are perhaps 15 billion acres of arable land in the world, or between 7 and 8 acres apiece. Possibly the cause of scarcity can be more easily visualized if the reader will imagine himself put on 17 acres of land, of which 8 acres are arable, and being forced to get from them everything that he is to consume. This is the actual average situation for the people of the world and, if some fare better, it is because some have less.

This is but part of the picture. Men are limited not alone by the limitations of their natural environment, but by those of their own make-up—by the limitations of their own time and energy. Life is too short to produce most of the things we want. Dr. Edward Gumbart used to illustrate this fact with a telling example: The average man has a productive life of less than 50 years. He does well to work an average of 50 weeks per year, and the average working hours are now well below 50 hours a week. Thus a man's entire productive life can be thought of as 50 years of 50 weeks of 50 hours as a maximum—125,000 hours of time, which can be put to satisfying a lifetime's wants. Let the reader imagine how long it would take him, working unaided, to build a house or to make an automobile, especially if he had to make all his own tools and equipment, and he will see that only

a limited number of things can be made in the limited time available. Men do not have all the things they want because of limitations in natural resources and in their own time and energy. On the one hand we have the endless variety of human wants; on the other side we have the niggardliness of nature. The combination of the two gives rise to, and lies in back of, all those problems which we call "economic." There are no economic problems where there are no human wants, nor are there economic problems where there is no scarcity. We might at this stage even take this as a temporary and provisional definition of our subject matter, to be made more precise when we have progressed somewhat further in our thinking: *Economics is the study of the problems that men face in their attempts to satisfy numberless wants in a world of scarcity.*

Questions

1. Do you believe that there is any relationship between one's age and the satisfaction of his wants? If so, what is it?
2. What are some of the peculiarities of the wants of the people of England which distinguish them from other people? Answer the same question for the people of France, Germany, China, and Holland.
3. Individual differences in wants may arise, your text says, from differences in climate, age, sex, or education. Give at least one illustration of differences in wants arising from each of these sources.
4. What are some other causes of differences in wants besides those enumerated in the preceding question?
5. Distinguish wants and needs.
6. Do you agree with the thesis of Sec. 1 • 6, that you do prefer shoes to automobiles? If so, why do men pay more for automobiles than for shoes? If you do not agree, why not?
7. Make a list of things that your family wants which your grandparents did not want.
8. Would you say that mosquitoes are scarce in winter? Why or why not?
9. In what sense was it true, during the depression of the 1930's when wheat farmers found it hard to sell their crops at prices that would pay their costs of production, that wheat was scarce?

2. THE NATURE OF ECONOMICS

2·1. Subdivisions of Human Knowledge. The field of human knowledge and human interests is so broad and includes so many varied sorts of things, that it is necessary to break it up into divisions and subdivisions for description and understanding. The student is, of course, familiar with the names of some of the subdivisions, such as chemistry, histology, French, or harmony. Great thinkers have long tended to assign these subdivisions to one or another of three major categories: science, art, and religion, which are thought of as comprising, among them, all the problems of human life. They are sometimes called philosophy, aesthetics, and ethics.

2·2. Field of Science. Now these three fields are very different, not only including different aspects of human life, but being treated by entirely different methods. Workers in the three fields differ in their purposes and reach their goals by different routes. For example, the scientist is interested in seeking the truth. He wants facts. He does not require that they be pleasant ones nor does he express either approval or disapproval of them. The facts are thought of not as moral or immoral, but as amoral—outside the field of morals. To be sure, the further one progresses in the field of science, the more keenly he becomes aware that facts are elusive and hard to recognize. Cocksureness, complete conviction, and absolute certainty are characteristics of the beginner in a field. At some distant time it may well have been true that the wisest and best informed men in a society were agreed that the world was flat. They would have said, "Yes, the earth is flat. That is a fact." Today the wisest and best informed men in our society may agree that the earth is approximately spherical (unless they dodge the issue by saying, as the geography books sometimes do, that the earth is a geoid which merely means that it is shaped like the earth!). They may say, "Yes, the earth is very nearly spherical, flattened slightly at the poles. That is a fact." But a century hence, or even tomorrow, new knowledge may change our ideas, and to men of that day our present "knowledge" may seem ignorance. They may say, "Yes, the earth is shaped like a corkscrew.

That is a fact." The point is that the "facts" of the scientist are always tentative, subject to change if some other explanation seems preferable. Facts are by no means the final and ultimate and unchangeable things imagined by the high-school sophomore.

With regard to any problem in the field of science, then, the scientist tries to find the facts. He asks, "What is true?" And in attempting to solve these problems he uses reason or judgment.

2·3. Field of Art. In contrast, in the field of art no competent person asks, "What is true?" If we see a great painting or hear a fine symphony, we do not ask, "Is it a fact?" In the field of art we deal not with facts but with feelings. We ask, not "Is it true" but "Is it beautiful?" To be sure, the untrained person sometimes displays his own ignorance in the field of art by asking whether or not a painting is "true to life," without realizing that he has tried to apply in the field of art a test that is valid in the field of science but meaningless in the field of art. The ancients, understanding this, said *De gustibus non disputandum est,* which can be freely translated as, "There is no use arguing over matters of taste." They realized that one could not reason about a fine poem or a sonata, that problems in the field of art are settled by one's taste rather than by reason.

2·4. Field of Religion. Finally we note that, in the field of ethics, morals, or religion, our problems are still different. They deal not with facts nor feelings, but with conduct. We ask, not whether the conduct is true or beautiful, but whether it is good. We form our conclusions not as a result of reason or taste, but by means of conscience. Cannibalism may seem good to some people and bad to others, and one cannot settle the question by reasoning about it. One can think up "reasonable" arguments on either side, but the final decision must be made on the basis of conscience.

2·5. Comparison of the Three Fields. We might summarize the remarks just made in tabular form, as follows:

	Field of thought		
	Science	Art	Religion
Subject matter	Facts	Feelings	Conduct
Question asked	Is it true?	Is it beautiful?	Is it good?
Method of solving problems	Reason	Taste	Conscience

It is not our intent to suggest that one of these fields is more or less important than any other. Unfortunately, students often become obsessed with the importance of some one of them to the exclusion of the others, thinking that nothing is important if it is not scientific, or perhaps feeling that art is the

whole of life. Any well-rounded life touches all three fields, and every major problem of life cuts across them all. But it is important for the student to understand that the fields are different, not only in the character of their subject matter but in the nature of the methods used for the solution of problems. One makes an unfortunate mistake when he tries to attack the problems of one of these fields by methods adapted to another.

2·6. Place of Economics. Most of our work in economics will be more nearly scientific than artistic or ethical. To be sure, most major economic problems raise ethical questions, and it is of utmost importance to society that the ethical aspects of the problems receive thought and attention. We shall spend most of our time in trying to learn the facts—to find out what is true and to reason about cause and effect in the field of economics. The layman often expects the economist to go much further than this and do things that he would never expect the scientist in other fields to do. For example, if it has been decided that we are to build a new bridge, the physicist and the engineer, as scientists, should be able to give us a great deal of help in knowing how to go at it. Their study of stresses and strains and materials should enable them to assist us in reaching our goal. But they are no more capable, as scientists, of deciding whether or not we should build a bridge than is anyone else. When a goal has been agreed upon, the scientist may be able to tell us how to reach it; but he has no special competence in choosing goals. Science is a means to an end, but not a selector of ends.

The economist, however, is often expected both to select the goal and to advise on methods of reaching it. He is expected to know whether or not more equal distribution of wealth is desirable, whether or not the goals of a government are the proper ones, etc. These questions are more nearly ethical than scientific. We do not expect the scientist in other fields to answer them.

There are those who feel that economics should confine itself to an appraisal and evaluation of goals—to the semi-ethical problems involved in deciding where society should be going. Without intending in any way to imply that such appraisal is unimportant or unnecessary, we shall confine our study here mainly to the more "scientific" approach that involves discovery of the facts. We shall naturally from time to time indicate some of the ethical or moral problems involved, however, and surely the student who studies a course in economics without raising in his own mind some of these problems has "missed the boat."

2·7. Kinds of Law. While we have been speaking of economics as a science and of our approach as a scientific one, the student should not be misled by such use of terms. Economics is, or may be, a science in the sense in which we have just been using the term (since it is an attempt to ascertain the facts, using reason), but the term is used in various ways at various times. On the basis of some uses economics is not a science at all. To be sure,

quarrels over matters of definition are fruitless; but it is nonetheless important that we understand our own definitions so that we may know what we are talking about.

Probably the differences between the various sciences can be most easily understood if we notice the differences between the laws of the several sciences. The word "law" is used in several senses, all familiar to the student. Perhaps his first acquaintance with the word was in the sense of what we call "statute laws," which are written down in the law books. They are intended as guides for human conduct and are established by human agencies (such as legislatures or kings or dictators). They do not exist without men and have to be enforced by men. If no attempt is made to enforce them, they do not work at all, as is evidenced by many an archaic "blue law," which technically remains on the books but actually is not respected nor enforced. None of them works perfectly, and some of them not at all. As examples, we might take the laws governing collection of income taxes. These laws once did not exist at all; they were made by men; they are but imperfectly enforced; and they could at any time be done away with by men.

Let us contrast with this meaning of the word "law" the meaning that we convey when we talk about "natural law," such as the law of gravitation, or Boyle's Law, or the periodic law. We do not think of the law of gravitation as having been made by man. It was, of course, discovered by and formulated by men but, as far as we know, it had been always true. It is not thought of as a desirable form of action. When we state that bodies attract each other directly as the product of their masses and inversely as the square of the distance between them, we do not indicate that such a type of attraction is desirable. No legislature or dictator passed the law, nor could they repeal it. The law does not have to be "enforced" by men. In fact, we notice that natural laws are merely descriptions of what happens, intended to be statements of fact. As was pointed out a few pages back, with the discovery of new knowledge, our understanding of the facts may change, and hence our natural laws may change. In this sense these laws are man-made and limited by man's capacity for comprehension. But they differ decidedly from statute laws, and the nature of the difference is surely apparent to every student.

Now let us turn to *economic laws,* which shall engage our attention in this book. The student is, of course, not so familiar with them as with the laws of the natural sciences or with statute laws. We cannot easily take up Gresham's Law, or the Law of Diminishing Utility, and use it as an illustration, since the student is not yet familiar with either. But it is important for the student to understand in advance, as he starts out on the study of economics, that the laws he is to study are much more nearly like natural laws than like statute laws. The laws of economics are not passed by legisla-

tive bodies. They are not man-made, nor do they require human enforcement. They are intended to be statements of fact rather than guides for conduct. In the *social sciences,* which deal with men and their behavior, the problems are much more complex and much less completely understood than are the problems of the natural sciences. In the natural sciences our results take on a form of almost mathematical accuracy. If we want to know how far a body will fall in a given time, we substitute in an equation and find an answer that in practice will turn out to be nearly correct. It will not usually turn out to be exactly correct, because of slight errors in measuring the time and distance involved, or because the body fell in air rather than in a vacuum, or for similar reasons. When we work out any actual practical problem in the natural sciences or in engineering, we always have to make some allowance for the *margin of error.* But these allowances are usually relatively small, and our computations relatively accurate. In the social sciences, including economics, we find that the same situation exists, but the errors are relatively much greater and the accuracy relatively much smaller. The difference is one in degree, but almost big enough to constitute a difference in kind. Thus the student will find that economic laws are usually stated as mere general tendencies, showing the kind of change expected rather than the exact amount of change. Instead of stating exactly what price is to be expected under given circumstances, we may have to content ourselves with saying that under these circumstances we would expect the price to rise. Perhaps we can go so far as to say that the price will "rise sharply." But we cannot say that the new price will be exactly $3.75. We do not get quantitative solutions except in purely hypothetical cases. The physicist may be able to say that a body will fall through 1610 feet in 10 seconds, or through 402.5 feet in 5 seconds; but if he were in the position of the economist he could merely say something of the general nature that, "the longer bodies fall, the farther they go." We describe this characteristic of economic laws by saying that they are *qualitative* rather than *quantitative*; they tell the kind or direction of change that is expected rather than the amount of change. Economic laws are statements of general tendency. Of course, the economist would like to make statements as precise as those of the natural scientist, and perhaps someday he may be able to do so. But after all, meager and approximate knowledge is better than no knowledge at all, and in the complex field of social relationships it is the only knowledge that we have as yet.

2·8. Economics as a Science. Accordingly, we shall understand that, when we talk of economics as a science, we are using the term in its most general sense. We intend to convey merely the idea that we hope to ascertain, as far as we are able, the facts and their casual connections. Economics is not a science in the same sense, or to the same degree, as astronomy or chemistry

or physics. If the student wishes to define the sciences as those fields in which one can make completely exact predictions of actual occurrences, he will quickly find that he has forced himself to rule out not only economics but all these others too. Yet it is important to keep in mind as we develop the various laws of economics that they do differ greatly in degree from the laws of the natural sciences, even though they are more like those laws than they are like statute laws. We shall have to wait for a more complete appraisal of economics as a science, and its relationships with the other fields of thought, until after we have studied more of the field of economics itself.

Questions

1. How can there be both a science and an art of photography?
2. Are there any ethical problems of photography?
3. Suppose that one is studying the facts about which combinations of musical vibrations sound discordant and which pleasing. Is he studying art, a science, or both?
4. Your text points out that men have changed their minds as to facts concerning the shape of the earth. Mention some other cases in which men's understanding of the facts has changed.
5. Are there principles or laws in the field of art, as in the field of science? If so, what are some of them?
6. Science may be so defined that economics is a science, or so that it is not. Give a definition that would include economics and one that would exclude it.
7. List five natural laws and five statute laws.
8. Do natural laws explain things, or merely describe things? For example, does the law of gravitation explain why bodies attract each other, or describe how they attract each other?

3. ECONOMIC DEVELOPMENT

3·1. Man's Intellectual Development. People brought up in different parts of the world, in different cultures, find it difficult to understand each other. Most Americans pity the poor natives of China and India, feeling sorry for their ignorance and their poverty. It is hard for them to realize that the Chinaman or the Indian is, in turn, pitying them for their preoccupation with worldly affairs and material things, for their hustle and bustle and the pressure under which they work, and for their lack of "cultural background." The American and the Chinaman differ relatively little in physical structure, but greatly in custom and habit and background and in outlook on life—in *Weltanschauung.*

This difference between ourselves and those of our contemporaries who have been raised in different cultures is no more pronounced than is the difference between ourselves and our ancestors. They, too, were much like us physically. Whatever evolutionary changes may have occurred in men have been so gradual that their effect is not noticeable within the span of recorded history. But in our own particular stream of social growth, which we call Western civilization, the past thousand years have brought revolutionary changes in habit, custom, and background and in outlook on life. We can perhaps understand our own present society better if we understand how it grew.

The biologist is accustomed to thinking of animals as exhibiting structural forms that adapt them more or less well to their environments. Some animals have developed great speed, others great strength, others keen eyesight, and still others a powerful sense of smell. Man, as compared with other animals, seems rather poorly endowed in these regards. In competition with other forms of animal life his major advantage has been an intellectual one rather than a physical one. While the other animals have been forced to undergo the long, slow, uncertain processes of physical change, men have managed to get speed or strength or eyesight by their understanding of nature. For example, the taming of the horse, the development of the motor car, and the invention of the airplane have all been steps in man's

attainment of speed. Man's intellectual adaptation to his environment has paralleled the physical adaptation of other animals; and a comparison of the typical man of the twentieth century with the typical man of the tenth century would surely show major intellectual differences as contrasted with physical changes so slight as to be almost or quite imperceptible. The weakening of superstition, the turn from the supernatural to the natural, the understanding of the laws of nature, and the consequent bettering of the human lot—all of these are part of the intellectual revolution. We shall now attempt, not to relate all the steps in this intellectual development, but to describe one or two phases of it for contrast with our present situation.

3·2. Direct Appropriation. Although no existing people live without some social interrelationships and without some plans for livelihood, and although we have no records of any such society in the past, still it is possible that men once lived like wild animals, without making any conscious provision for the future.[1] Certainly it is true that the most backward of known people are but little removed from such a crude state, relying for livelihood primarily on what they chance to find about them in nature, without any very complicated system for increasing or controlling their means of livelihood. Societies that depend on hunting and fishing would be good examples of what we call the stage of "direct appropriation." In this stage men take what they can find in nature, appropriating it for themselves but without raising crops or livestock. They may eat wild berries, but they do not grow wheat or corn. They may eat trout or venison, but they do not raise beef or mutton. They trust in the bounties of nature and, since these bounties are not very trustworthy, such peoples have usually had a pretty hard time of it.

Of course, even men in the direct-appropriation stage are far above the level of brutes in their economic arrangements. They do work together to accomplish their purposes. They do use crude tools to help them with their work—bows and arrows, stone implements, etc. The direct-appropriation stage as we know it is already a definite step in advance of animal brutality.

3·3. Pastoral Stage. Domestication of animals, however, appeared in some parts of the world very early. Men learned that, while hunting and fishing were not dependable as methods of getting continuous supplies of food, they could tame and breed animals and have their food supplies always on hand. Similarly they learned to "tame" various food plants, to grow and multiply them for their own purposes. Thus men shifted to a system of agricultural production, which increased the food supply many fold and made it possible to support many more people in a given territory than could be supported under a system of direct appropriation.

[1] An American sociologist, Thorstein Veblen, has well characterized such guesses as to what was true before men kept records as "conjectural history." The student will realize that on such matters any statements are, at best, guesswork.

Those peoples who were really pastoral, who raised herds and flocks, were usually nomadic, grazing their animals in a given locality until the pasturage was used up and then moving on. The Old Testament tribes are good examples of such a type of economic organization. But the people who raised crops tended to settle more permanently. They had to clear and cultivate land and, having expended the time and energy to get their land in good tilth, they naturally tended to stay and use it rather than move on. Thus the societies that relied on animal raising were in many ways different from those which were mainly crop raisers. The two types of enterprise existed in all sorts of combinations.

3·4. Property. Dependence on flocks or on improved land raised new sorts of problems, among them those which we have since come to call the problems of "property." Property, as the lawyer and the economist define it, is a peculiar thing. It does not consist of things themselves, but of rights to use and enjoy them. My automobile is not itself property, but I have property rights in it. I have the right, guaranteed by society, to use and to enjoy it and, if anyone else tries to take it away or use it for his own enjoyment, society (through its police and judicial systems) will attempt to protect me in my right. Property is more than possession. A thief may possess my car, but it is not his property. Nor is a thing my property merely because I am personally strong enough to protect my possession of it. Property is a social thing, established by and enforced by society. Even a weakling unable to protect his own rights will have his property rights protected for him by society. *Property* is the right to use and enjoy goods, established and enforced by society. If, in time of anarchy, social enforcement breaks down, property disappears. The material goods are still there, but they are no longer property. It is important for the student to realize that property is an intangible thing, created by society. It is an example of what we call a "social institution."

This phrase, which we shall use rather often, puzzles most students who run across it for the first time. They are in the habit of thinking of an insane asylum or a university as an institution, and they are puzzled when they hear social scientists talking about property and marriage and junior proms as institutions. As the social scientist uses the term, an "institution" is a commonly accepted pattern of social behavior or usage. It is a fairly permanent and usually rather complex mode of thought or action which has become embedded in the habits of the people. Thus we could say that in American colleges and universities the wearing of academic caps and gowns at commencement is an institution. Commencements could be run in other ways, but in our society there is a certain rather widely adopted and more or less lasting pattern of behavior that characterizes college commence-

ments and makes of them an institution. When we say that private property is a social institution, we are thinking, not of the goods themselves, but of human behavior with respect to goods. It is only by ellipsis that we say, as we often shall, that an automobile is property instead of saying that men have property rights in automobiles.

To get back, then, to our story, when men came to depend on flocks or on improved land, they had to adopt some sorts of conventions or customs with regard to them. Such things might have been held in common, belonging to the community as a whole, and in societies that depend on direct appropriation land often is so held. The Indians who sold Manhattan Island to the white men for a few beads and trinkets were not so simple as property-minded men have supposed. The Indians did not "own" the land nor did they have private property in it. They were simple only in the sense that the confidence man is simple who sells Brooklyn Bridge to a country yokel. But common ownership has its difficulties under a system of settled agriculture. The problem can perhaps best be understood from an actual example in American history, described by Governor Bradford in his handwritten record of "Plimoth Plantation."[1]

> All this while no supply was heard of, neither knew they when they might expect any. So they begane to thinke how they might raise as much corne as they could, and obtaine a beter crope then they had done, that they might not still thus languish in miserie. At length, after much debate of things, the Gov^r (with y^e advise of y^e cheefest amongest them) gave way that they should set corne every man for his owne perticuler, and in that regard trust to them selves; in all other things to goe on in y^e generall way as before. And so assigned to every family a parcell of land, according to the proportion of their number for that end, only for present use (but made no devission for inheritance), and ranged all boys and youth under some familie. This had very good success for it made all hands very industrious, so as much more corne was planted then other waise would have bene by any means y^e Gov^r or any other could use, and saved him a great deall of trouble, and gave farr better contente. The women now wente willingly into y^e feild, and tooke their litle-ons with them to set corne, which before would aledg weaknes, and inabilitie; whom to have compelled would have bene thought great tiranie and oppression.
>
> The experience that was had in this comone course and condition, tried sundrie years, and that amongst godly and sober men, may well evince the vanitie of that conceite of Platos & other ancients, applauded by some of later times;—that y^e taking away of propertie, and bringing in comunitie into a comone wealth, would make them happy and florishing; as if they were wiser then God. For this comunitie (so farr as it was) was found to breed much

[1] BRADFORD, "History of Plimoth Plantation," pp. 162-163, printed by the Commonwealth of Massachusetts, Boston, 1898. The entry is from Governor Bradford's records of the Plymouth Colony in the year 1623.

confusion & discontent, and retard much imploymet that would have bene to their benefite and comforte. For y^e^ yong-men that were most able and fitte for labour & service did repine that they should spend their time and streingth to worke for other mens wives and children, with out any recompence. The strong, or man of parts, had no more in devission of victails & cloaths, then he that was weake and not able to doe a quarter y^e^ other could; this was thought injuestice. . . .

3·5. The English Manor. Records of the early stages of economic organization are fragmentary and unsatisfactory; but by the late middle ages, say about 1300, English society had developed systems of records, which tell us with considerable detail how the people lived. At that time England was primarily agricultural. The people lived, not on isolated farms, but in little clusters of small, rude, one-room, thatched houses in tiny villages or manors. The social, political, and economic aspects of life were all very greatly decentralized, so that each little manor of maybe two or three dozen families was largely independent and self-sufficient. Some few things, such as salt and metals and spices, were brought in from outside, but the ordinary member of one of these communities had little or no contact with people from other villages. The manors were usually rather small clearings in the woods, often on the banks of a stream that could be used to run the local mill, and largely isolated from other settlements. The houses were commonly clustered around a village green, with outlying fields and pasture running out to the surrounding woodland.

Each manor was under the jurisdiction of a lord, who might be a noble or a bishop, a knight, or even the king himself. Some lords controlled many manors, some relatively few, and some only one. These latter lesser lords usually lived in the manor house of their manor, but if the manor was one of several under the jurisdiction of one lord, he was represented at the manor house by a steward or bailiff. Next in the social scale of the manor came the freemen or free tenants. The freemen owned their own land on condition that they pay a fixed money rent to the lord; in addition that they give a few days of "boon work" to the lord each year at such times as planting and harvest. Their liability was mainly a monetary one, however, and the boon work in their case was not particularly onerous. Below the freemen in status were the villeins. They, too, had land assigned to them which it was their right to use; but in return they gave large amounts of boon work to the lord. They might, in addition, pay cash rents, but their major obligation was that of working on the lord's land, or demesne. The records indicate that villeins commonly were required to give up as much as half their working time to the lord. Moreover, the villein had no right to sell his land, to leave the manor, to give his daughter in marriage, or to make

many other decisions which we would consider individual rights, without permission of the lord; and such permission was ordinarily given only on payment of a fine. In a sense the villein's land belonged to him; but in another sense he belonged to the land. He had no right to leave it or to seek a living elsewhere or in a different occupation.

In addition to the lord, the freemen, and the villeins or serfs, there were sometimes other less numerous classes in the manor. Sometimes we find men who had almost no land assigned to them, merely a little cottage and the surrounding yard. These men, called "cotters," were required to give a small amount of boon work to the lord each year but had to earn the main part of their living working for others. Likewise many of the manors had slaves, who were usually either captives of war or men who had sold themselves into slavery. It was not uncommon in these days for men to fall into slavery on the loss of wagers. Although modern descriptions of the medieval English manor usually list slaves among the classes of people, it seems likely that they were relatively uncommon and few in number. The freemen and the villeins were the people who made up the bulk of the population.

To be sure, there were usually a few specialists in the community who fell into none of the classes so far mentioned. There was the parish priest, there was the miller who paid a rent to the lord for the privilege of operating the lord's mill (and who usually charged so much in turn for his services that the miller was traditionally one of the wealthier members of the community), there was the smith, and possibly a few other artisans. But the entire population of the manor was very small, and the organization relatively simple.

3·6. Three-field System. As we have said, agriculture was the major industry, yet it was crude and backward. Little or nothing was known about fertilization of land, or about the selection of crops and livestock. Crop yields were very low compared with modern averages, and the livestock were small and unproductive. The sheep of those days often yielded as little as 2 ounces of wool, compared with an average yield of from 8 to 10 pounds today. Wheat yields were but 6 to 8 bushels per acre, compared with three to five times that much today. As a consequence, the people were very poor.

The arable land of the manor was usually divided into three great fields,[1] and any individual man held small pieces scattered through these three fields. Instead of having a compact area to till, he held perhaps 30 acres in 30 small pieces, 10 pieces in each of the fields. One of these three fields was given over to a crop that was planted in the fall and harvested the next summer, such as wheat or rye. Another field was planted to crops,

[1] There were other systems, such as a two-field system, but the three-field system described here seems to have been most common.

perhaps peas or oats, in the spring and harvested the following fall. The third great field was left unplanted and idle, lying fallow, with the idea that the soil would thus rest from its labors one year in three and recuperate. The fields were rotated from year to year, but in any particular year all plantings in one of the fields must be a spring crop, all plantings in another field must be a winter crop, and no planting was allowed in the third field. This can perhaps be more easily visualized by reference to Fig. 3·1, which shows the plan of an imaginary manor. In this figure the student will see the houses along the main street, which in practice was a narrow, rough lane. The three fields are evident. The northwest field is in wheat and other winter crops this year. The south field this year is given over to oats and other spring crops. The east field is fallow. Since all the land in a given field was plowed at the same time and since the cattle were turned in to pasture the field immediately after the harvest, it was necessary for every landholder, whether lord, freeman, or villein, to grow crops that were planted and harvested at the same time. Thus the landowner's choice in the use of his land was greatly restricted.

The large fields were usually divided into sections, called "shots." The figure shows the northwest field divided into four such shots. The individual landholdings were broken up into small strips, usually an acre in area, and these strips were then scattered through the various shots of the various fields. Any landowner thus had a number of small, scattered holdings. The villein might have 20 or 30 strips, and the lord might have five times that, but in either case the landholdings were broken up and scattered. This, too, is evident in the figure, where one man's holdings in the northwest field are indicated. This was supposed to have the advantage that no one got all the good soil while others got all the poor soil. Each man's holdings were so scattered that he shared in all sorts of soil, getting some that was stony and some that was not, some that was wet and some that was dry, some that was sterile sand and some that was fertile loam, etc. In some manors the holdings seem to have been rotated from year to year; in others they seem to have been more or less permanently assigned. In some manors all villeins seem to have been assigned holdings of the same size, while in others there seems to have been little uniformity. It would not do to give the impression that the three-field system was universal nor that it was uniform, but the commoner characteristics were those here described. There are many additional things that could be said of the English manor, which has here been described in barest detail. It had its own system of justice under the lord of the manor, and it had its own military establishment in which the freemen and the villeins were bound to serve at need. The people within the manor were thrown into constant and intimate contact with one another,

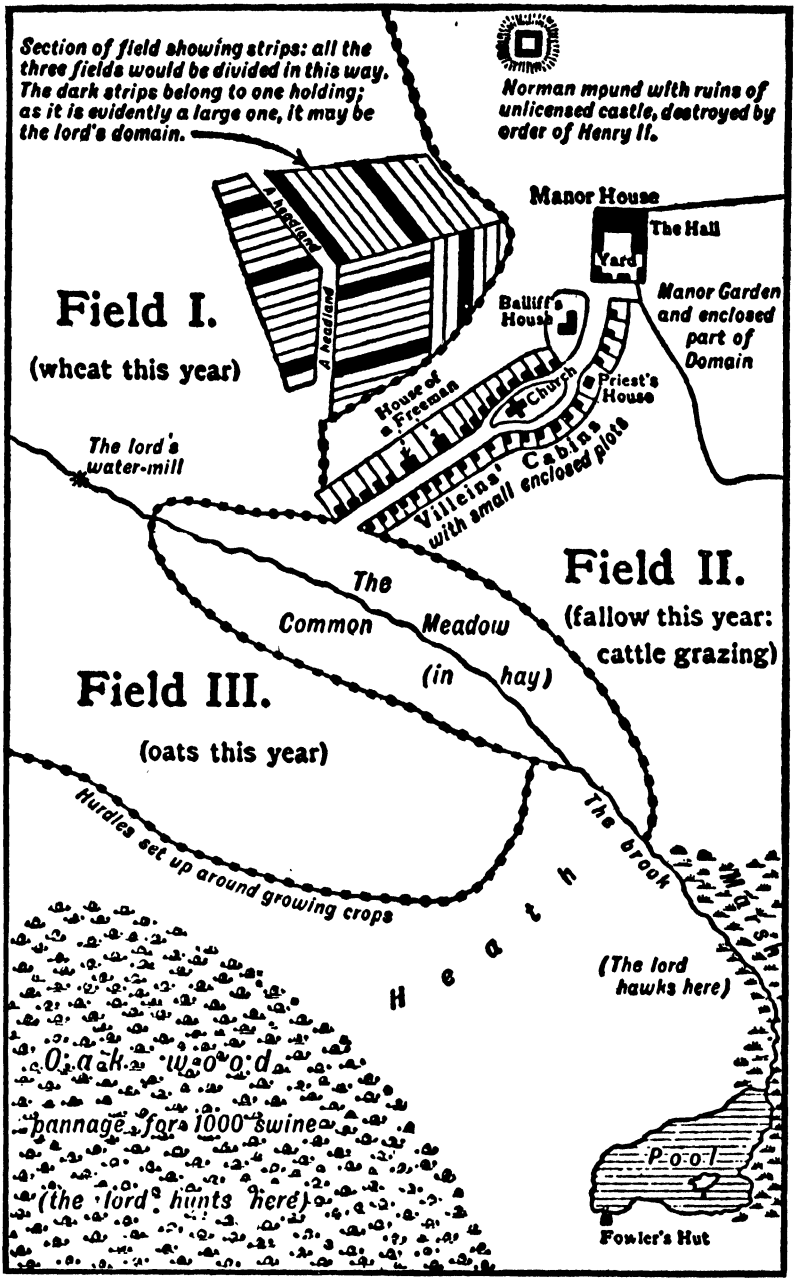


FIG. 3 • 1.—Idealized map of a medieval English manor. (*Taken, with permission, from G. M. Trevelyan, "History of England," Longmans, Green and Company.*)

but their contact with outsiders was almost nonexistent. It was a day of real "isolationism." One student of the times[1] has described the situation in these terms:

> Government was so weak that robbery was common; people were so ignorant of everything outside the narrow sphere of local interests that they suspected every stranger, and too often with reason. There is a whole series of English laws, beginning about 700 and continuing for centuries, of which this is an early example "If a man come from afar, or a stranger, go out of the highway, and he then neither shout nor blow a horn; he is to be accounted a thief, either to be slain or to be redeemed."

3·7. Handicraft Stage: Town Economy. By the fourteenth century the towns of England had begun to assume a position of some importance. Usually chartered by the king, a nobleman, or the church, the towns were few in number and small in size by modern standards. Most of them had populations of 2,000 to 3,000 people, but the bigger ones ran up to perhaps 10,000 or 15,000. At this time London had a population of about 25,000.

The towns were trading centers and consequently were established on the sea coast, on rivers, or at the junctions of the major trade routes. This is evident in the names of such towns as Staf*ford,* Hert*ford,* Walling*ford,* Ox*ford,* Chelms*ford,* or Stur*bridge,* Cam*bridge,* etc. The amount of trade was very small as measured by modern standards, but it was great as compared with that of the isolated manors. It must not be thought that agriculture entirely disappeared, even in the towns; but, as compared with the manors, trade and industry were relatively more important. The towns contained small shops and markets, but no large factories.

Also significant in the town development was the increased specialization as compared with the life of the manor. In the towns there were cobblers, smiths, weavers, tailors, carpenters, and workers in many other crafts. These craftsmen, specializing, produced more of their products than they could use themselves, and their surpluses gave rise to trade, both among the different specialists within the town and with the people in the surrounding countryside. Moreover, the town craftsmen were free and independent, unlike the villeins of the manors. One of the noteworthy effects of the growth of towns was their effect on freedom in general. The custom grew up that, if a villein fled from his manor to a town and was able to stay there for a year and a day without being caught, he became free. Prior to this time the villein had had no place to which he could escape. If he left one manor, he had to go to another. But the freedom of the towns soon began to have a noteworthy effect on the life and freedom of the manors.

[1] DAY, CLIVE, "A History of Commerce," p. 33, Longmans, Green and Company, New York, 1912.

Each town set aside a market place, at which trade was carried on weekly or semiweekly. On market day the merchants brought their wares for sale, and the townspeople and the people of the surrounding countryside came to buy. These markets were usually sponsored by some nobleman or churchman, perhaps the one who had chartered the town in the first place; and this sponsor sought to protect the merchants in their trade and to get for them freedom from the tolls that were so common in those days. One must remember that in those days government was still largely decentralized. A man from the next town was looked upon as a "foreigner," and, just as nations today levy import duties on goods coming across their borders, so it was common for towns in that time to levy tariffs and to charge tolls. Similarly, tolls were charged for the use of roads, bridges, fords, etc. In fact, there were many cases where men built bridges on dry ground, forced travelers to cross them, and then charged them tolls for the use of the bridge! The sponsors of the town markets attempted, as far as possible, to get relief for the merchants attending their market, and to get more freedom of trade.

The growth of trade in the towns also had another effect. This was the increase in the use of money. To be sure, money had been known centuries before but there was little use for it in the manor because there was little trade. A man who is self-sufficient needs little money. Now, as trade and commerce became more common, the use of money increased.

3·8. Fairs. Some markets grew much larger, more important, and more widely known than others; in a few of the major towns there grew up the custom of holding a great market day once a year, or sometimes even semiannually or quarterly. These occasions were called "fairs," and soon grew to combine recreation and entertainment with trade. Each fair was owned by someone, usually as a grant from the king. The owner did what he could to encourage merchants to come to the fair and set up booths which he rented to them for the display of their wares. There were also side shows and many of the side lines that have been common at country fairs here in the United States down to the present time. Some of these fairs ran for a week or two, and they were gala occasions in the countryside. Usually, while the fair was in session, no merchant was allowed to sell goods anywhere in the surrounding territory except at the fair. For the period that it lasted, the fair had a complete monopoly of trade. Usually the fairs set up their own courts, which had power to try such cases as might arise out of trade and commerce at the fair—cases of misrepresentation, adulteration of goods, short weights, etc. These courts built up a great body of law, called the "law merchant," which was less complicated and involved than the ordinary law of the town and manor courts, and which was the forerunner of much of the commercial law of today. The larger fairs were without question the best markets of the day, and merchants came from afar to buy those

kinds of goods which were produced locally, and to sell things from a distance.

3·9. Merchant Gilds. The towns, as we have said, were trading centers. To be sure, the trade was largely confined to market days or, in those centers where fairs were held, to the time when the fair was open. But the merchants were the typical people of the towns—the people who gave the medieval town its character and flavor. They early formed societies, some of which were purely social in character but some of which, the *merchant gilds,* were to a considerable extent economic in character. These merchant gilds had as their basic purpose the protection and regulation of trade. They were by no means "national" organizations but were limited to the confines of a town, to which all merchants in the town could and must belong. The merchant gild, as a matter of fact, was a part of the government, or perhaps we might better say that it was a kind of government. Just as the town governed judicial and administrative affairs, so the gild governed trade affairs, and within this sphere it had the right to make laws and enforce them. No one was allowed to buy or sell goods in any town unless he was a member of the local merchant gild, or unless he complied with whatever rules they wished to lay down for trade by "foreigners." It seems clear that the fundamental purpose of the merchant gild was to obtain and to preserve for its members a monopoly of trade within the town.

3·10. Craft Gilds. As time passed, and specialization became more marked and more important within the town, the differences between the various kinds of producers and merchants came to be more important than their similarities. Under these circumstances the merchant gilds grew slowly less important, and associations of people within the same trade or occupation, the *craft gilds,* came into being and became more and more important. These, also, were confined to the members of the craft within a particular city, but in the larger cities there came to be as many as 40 or 50 different craft gilds—organizations of fishmongers, dyers, glovers, weavers, fletchers, bowyers, and many others. A man might be a member of a merchant gild and also of a craft gild; but his craft gild regulated the production and distribution of the product that he made, and came in most cases to be much more important than the merchant gild, which either died out or became largely social in character.

Most of the craft gilds obtained their authority from the town government, and they usually enjoyed a complete monopoly of their occupation within their town. At first membership was not restricted; anyone who was competent at the trade might join. Later some of the craft gilds became exclusive and sought to take advantage of their monopoly position by limiting membership and raising prices. The rules governing apprenticeship and new members were in the hands of the gild, as were rules concerning

the amounts and qualities of the product that could be made, where and under what conditions it might be sold, etc. Some of the restrictions were doubtless intended to protect consumers against inferior products but, as is quite natural when any economic group makes regulations governing its own business, many of the restrictions were made selfishly with the purpose of improving the lot of the members without serious regard to the social consequences. The craft gilds, like the merchant gilds, often combined social and recreational and sometimes educational activities with their economic activities. They played an important part in the life of the medieval town.

3 • 11. Domestic System. The period that we have just been describing, during which the towns were characteristic features of social organization, has often been called the "handicraft stage." Slowly it tended to give way to another sort of organization which is usually called the "domestic system," or sometimes the "putting-out system." Although such a period has no precise point of beginning or ending, we usually think of the domestic system as having run from about the beginning of the sixteenth to about the middle of the eighteenth century. During this period the gilds slowly weakened, and we get the first faint signs of the emergence of capitalism. Manufacturing during this period continued, as before, to be carried on in the home, by a man and his family and sometimes with the assistance of one or more apprentices. But instead of carrying on all the manufacturing and trade themselves, the family now relied on men who carried on a new sort of work. These new functionaries bought and owned raw materials, which they supplied to families for home manufacture. Then they took the finished products from the family for sale. Sometimes these "capitalists," if we can give them that name, also owned and supplied the simple machinery or equipment that was used in manufacture, so that the worker owned and supplied merely his own time and effort, while the materials on which he worked and the machines with which he worked belonged to others. Thus came the beginning of that separation of the workman from the ownership of his equipment which is a commonplace characteristic of modern industrial society.

The processes of production, as we have said, continued to be done at home, and it is this characteristic that gives the period the name "domestic system." The family often continued its farm work and carried on the processes of industrial production as a side line. Production units were small as contrasted with the modern factory, since the workers were usually confined to the members of one family with possibly a few hired men in addition.

3 • 12. Mercantilism. Students of government will recall that, during this period about which we are now speaking, as the powers of the feudal lords and of the gilds and towns weakened and as the church lost much of the

secular authority that it had exercised during the Middle Ages, the power of national governments grew. The state took much of the authority that had theretofore been wielded by the other groups just mentioned, and the feeling of nationalism grew apace. Men were interested in learning how to increase the power and strength of the state, both internally as compared with the manors and towns and other local authorities, and externally as compared with the power of other national states. The policies that were then thought to strengthen the state we call "mercantilism," and the men who formulated and directed them we call "mercantilists." Mercantilism involved a large amount of internal regulation of life by the state, including many sorts of restrictions on business and commerce. It favored the growth of colonies and the establishment of empire. The state was interested in the growth of its population, and mercantilism included various sorts of measures intended to stimulate the growth of population. Perhaps the most persistent and lasting of the ideas of the mercantilists—an idea that unfortunately still actuates many a modern statesman—was that the prosperity of a nation depended to a very considerable extent on the amount of money within it, especially on the amount of gold and silver. The mercantilists thought that individual men were better off if they had more money and that the same thing should be true of governments. Likewise, they thought that the best way to get money was to sell things, and the best way to keep money was to forego buying things. Consequently, if the nation was to increase its monetary stocks (barring the mining of precious metals at home), it could do so by selling goods abroad without buying abroad. The mercantilist believed in what he called a "favorable balance of trade"; that is, a foreign trade in which the value of the exports exceeded the value of the imports. Consequently, the mercantilist favored such policies as might discourage imports (tariffs, etc.) and might encourage exports (bounties, etc.).

The mercantilists thought of the benefit to the nation, rather than of the benefit to the individual, as the aim of economic activity, and as a result they imposed all sorts of onerous restrictions on businessmen. In this particular regard the mercantilists were forerunners of modern fascism and communism, both of which rely on minute regulation of the individual by the state.

3 · 13. Situation in 1750. At the middle of the eighteenth century industry was for the most part in the hands of small independent businessmen, many of whom combined manufacturing with farming. In the seasons when farm work was slack the family turned its energies to the production of manufactured goods. The employees were few, possibly a journeyman or two and sometimes an apprentice, and the "employer" worked along with his men, sharing their problems and maintaining intimate personal relations with them. Production processes were practically the same as those which had been

used for centuries. Textiles, for example, were still carded, spun, and woven by farmers' wives in scattered homes, using the old-fashioned spinning wheel and the hand loom. To be sure, even this was an improvement over the still more old-fashioned distaff, and the spinning wheel had been improved by the invention of a treadle which freed the hands of the spinster and enabled her to twist two threads instead of one.[1]

3 • 14. Industrial Revolution. If, in the United States, a student is asked to name the most important events between 1750 and 1800, he is apt to select episodes connected with the Revolutionary War, overlooking completely the fact that his daily life is much more influenced by other and less familiar events of that period. As a matter of fact, the century from 1750 to 1850 saw the development of a great many of our modern institutions, and we might with considerable justice cut the period down even more to the 50 years from 1775 to 1825. Prior to that period, life was on the whole pretty much as described in the preceding paragraph; by the end of it, life was much as it is today.

The student who has read in the earlier parts of this chapter about life on the medieval manor, and the gilds, fairs, yeomen, serfs, the strip system, and domestic manufacture, realizes that his life differs sharply from that about which he has been reading. Each of us has a tendency to date things in accordance with his own experience, thinking that things were old-fashioned before the first World War, or before the "turn of the century," or before the Civil War, depending on when we came upon the scene. Yet life in the countries of Western civilization is in most fundamental respects still pretty much the same as it was 20 or 50 or 100 years ago, and radically different from what it was 175 or 200 years ago. The period from 1775 to 1825 witnessed not only an important political upheaval and the development of new concepts of government, but also basic changes in the status of land ownership, the control of industry, the techniques of production, the status of the family and the church, the accumulation of capital, the position of wage earners, and the relationship of the government to economic enterprise. Changes came so thick and fast that men have come to call the period the "Industrial Revolution." The name is, perhaps, a bad one. In the first place, the changes were by no means confined to industry; in the second place, the student should realize that society has never been static, that change is the rule, that the changes of this half century were in many ways logical outcomes of the long, slow, evolutionary changes that had occurred in the preceding century or two. Moreover, the changes did not come to an abrupt stop in 1825 but have continued ever since. Yet despite these qualifica-

[1] The making of textiles was so exclusively woman's work that the distaff came to represent women's activities in general. Even today we speak of the "distaff side" of the household. Unmarried women are still called "spinsters" as a carry-over from this period.

tions, the phrase "industrial revolution" has become firmly entrenched as a description of those changes which occurred at the end of the eighteenth and the beginning of the nineteenth century—changes, which we shall now describe in somewhat more detail.

3·15. Causes of the Industrial Revolution. As has been said, society is undergoing constant changes, and the Industrial Revolution consisted merely of a sudden speeding up of the rate of change. This acceleration occurred at different times in different parts of the world, and has not reached some societies even today. Although its progress was uneven, many of its developments occurred first in England, and it is perhaps fair for this reason for us to examine the English situation in an effort to understand the causes that underlay the movement.

In the first place, modern economic organization depends on large-scale production, which in turn, depends on the development of large markets. England's markets had been growing, both at home and abroad. At home there had been a movement of population to the cities, and the gradual disappearance of customs barriers, as the manors lost their power, left a large area of relatively free trade. Abroad, England had been establishing and developing a colonial empire that yielded raw materials for English manufacture and a market for English products. The English merchant marine was reaching every major port. World trade was changing from occasional or even unusual isolated ventures to routine day-by-day transactions.

The growth of trade and the development of markets were important in themselves, but they also gave rise to an increased accumulation of capital which made possible the growth of large-scale enterprise. In addition, the restraints imposed by the gilds were rapidly disappearing in England, leaving businessmen free to try new methods and to develop new products. The decay of the manorial system and the subsequent movement of population to the cities made available a reasonably large labor supply, which was augmented by the immigration of considerable numbers of skilled artisans fleeing from religious persecution on the Continent.

Finally, England was fortunate in the possession of those natural resources which were necessary for the coming of a machine age. She had rich deposits of iron ore close to forests that supplied the charcoal for smelting; by rare good fortune, just as the forests were giving out and the supply of charcoal seemed about exhausted, methods were developed for smelting with coal and coke. England has reasonably large quantities of these latter fuels, and they were fortunately situated near the iron mines. This combination of favorable circumstances together with a series of inventions set the movement under way.

3·16. Inventions in the Textile Industry. The manufacture of woolen cloth had long been one of the major industries of Britain. As has already

been said, this manufacture was a small-scale domestic industry, carried on primarily by hand methods on a family basis. Despite the replacement of the distaff with the spinning wheel, spinning was still the bottleneck of textile production. It took five or six spinners to supply one weaver, even with the old loom in which the shuttle was thrown by hand. When, in 1738, John Kay invented the flying shuttle, by means of which the weaver could jerk the shuttle back and forth by pulling a cord without the aid of an assistant, the disparity became even greater. It was impossible to produce thread and yarn rapidly enough to meet the demand.

In the face of this difficulty James Hargreaves developed, in 1764, his spinning jenny, which carried eight spindles and thus made possible the spinning of eight threads at once.[1] Further improvements soon doubled the number of spindles, and before Hargreaves' death his jennies were operating as many as 80 spindles, thus doing the work of many spinners.

The yarn spun by the jenny was too coarse and soft to be used for warp, the long threads which run lengthwise of the cloth and between which the cross threads or "woof" are woven. In 1769, Richard Arkwright invented a system of spinning which involved the passing of the thread between pairs of rollers revolving at different speeds. These rollers drew out the thread to the requisite fineness, and the product was strong enough to use for warp. Since Arkwright's original machine was operated by water power, it came to be called the "water frame." In 1779, Samuel Crompton perfected what came to be known as "Crompton's mule" because of its hybrid origin. It combined features of Hargreaves' jenny and Arkwright's water frame and turned out a thread both finer and stronger than its predecessors.

The invention of the jenny, the water frame, and the mule, quite upset the balance (or reversed the lack of balance) in the textile industry. Whereas before these inventions it was possible, using Kay's flying shuttle, for one weaver to keep up with perhaps 10 or 12 spinners, it now was possible to spin far more thread and yarn than weavers could possibly manufacture into cloth. This matter was called to the attention of a British clergyman, Edward Cartwright, and despite the fact that he knew nothing whatever about weaving, he decided to see what he could do about it. It took him three years but, at the end of that period, in 1785, he had developed a power loom with an automatic shuttle. Improvements on this machine enable a single worker

[1] It is said that Hargreaves stumbled on to the idea of his spinning jenny quite by accident. He entered his house suddenly and unexpectedly one day, startling his wife, who was spinning at an old-fashioned high wheel. His wife started up and upset the wheel. Hargreaves noticed that the wheel continued to turn, although now in a horizontal position, and the spindle continued to revolve. It occurred to him that it would be possible to connect several upright spindles to one horizontal wheel, and he soon developed such a machine, naming it "Jenny" after his wife. Pictures of Hargreaves' early machines show that they actually did look much like the old high spinning wheel tipped on its side, and to this extent the old story gains support.

to weave more cloth than could be turned out by 200 men on the old-fashioned hand loom.[1]

3·17. Other Inventions. The series of inventions in the textile industry just described was in many ways one of the most remarkable series of inventions in history. Yet it was not in this industry alone that the Industrial Revolution was felt. Watt took out his first patent on a steam engine in 1769, followed by improvements in 1778 and 1781. About 1785, this new steam engine came into use in factories, very slowly displacing water power. Early in the nineteenth century it was adapted to water transportation. Fulton experimented with his *Clermont* on the Seine in 1803, at New York in 1806, and on the Hudson River in 1807. In 1814, George Stephenson perfected his steam locomotive. And, while these changes in transportation by water and by rail were taking place, McAdam was perfecting his new methods for paving highways which began to come into use about 1818. It was an era of canal building. The Erie Canal had been started in 1817, and its immediate success led to almost a mania for the construction of others.

Along with all these improvements in transportation came fundamental and far-reaching ones in communication. Morse perfected his electric telegraph in 1837, and Bell's telephone dates from 1877. Electricity was just beginning to make itself felt as a source of power.

To be sure, other extremely important inventions still lay ahead: the electric light, the radio, the development of the internal-combustion engine, and the automobile. The Industrial Revolution has continued down to the present day and shows no sign of slackening. Yet it was in the short span of a generation or two bridging the end of the eighteenth and the beginning of the nineteenth centuries that men's lives underwent the most radical changes.

3·18. The "Wealth of Nations." The great changes that were taking place in technology between 1750 and 1850 were matched by a similar change in men's habits of thought. The mercantilists had built an economic system that was regulated in great detail by the government. Authority was highly centralized, and regulations were omnipresent and onerous. Mercantilism had gone so far in this direction that there came a natural reaction—a revulsion against government interference in men's lives and the control of industry by the state. The emphasis on personal liberty became evident all through men's lives, in the church, in the growth of political democracy, and in the

[1] There is dispute as to priority of invention of every one of the machines mentioned in the text. These disputes are of no moment to us here. Yet the student should realize that every great inventor builds on ground prepared by hundreds of unknown predecessors, and that presumably there were just as great and important industrial revolutions back before the dawn of history when some early innovator developed the wheel, the treadle, the use of water power, the use of animal power, etc. What a difference must have been made in the lives of men by the discovery of methods for using fire or by the taming of beasts of burden!

economic sphere. Perhaps the most important single force in molding men's opinions on this matter in the economic sphere was a famous book published in 1776 by a Scotch philosopher and economist, Adam Smith. This book, "An Inquiry into the Nature and Causes of the Wealth of Nations," urged greater liberty in the economic world at the same time that men were taking up arms in support of increased political liberty. The book was one of the most influential ever published, having a considerable effect not only on contemporary thinking but also even on that of the present day.

3·19. Laissez Faire. Adam Smith and his followers believed that, if we allowed men to follow their own self-interest, we would find that their actions would usually turn out to be socially advantageous—that it did not take complicated government regulations to make men work for the social advantage; they would do so by themselves if we just let them alone. Thus we got the expression, *laissez faire,* "let it alone," which was the rallying cry of the opponents of mercantilism. They argued that if a manufacturer made high-quality goods and sold them at low prices he would get the trade and the profit; while, if his competitor made goods of low quality or sold at high prices, he would lose the trade. Thus it was to the personal and individual advantage of the manufacturer or of the merchant to do those things which society as a whole wanted done. Society gained by allowing the individual to follow his own selfish interests, so there was no necessity of government interference. "He governs best who governs least."

Actually, the advocates of *laissez faire* built up an important and formidable following, and for many years the practical decisions of most commercial nations were greatly influenced by their ideas. Many of the restrictive acts of the earlier periods were repealed, and private enterprise was given much greater freedom. To be sure, the change should be thought of as one of degree and not of kind. Under the mercantilist system, or under any other system that has ever been in use, many decisions are left to individuals without interference by government. On the other hand, no society has ever existed that did not impose some restrictions on the actions of individuals. We never get complete freedom, nor entire lack of it. We realize today that there is little advantage in arguing over the extreme cases, but that we have to make decisions as to the degree of regulation or the degree of liberty that is most desirable.

3·20. Rise of Capitalism. Along with the technological changes that constituted the Industrial Revolution and partly because of these changes, there came far-reaching developments in the social structure and institutions of society and in the philosophy of the people. Slowly there developed a social system which we now call "capitalism." The term is perhaps most often applied to certain institutions within the economic field, but capitalism is a pervasive phenomenon which runs throughout modern life.

Capitalism is altogether too complex for anything like complete understanding at this stage in our study, and detailed consideration of its characteristics will have to be postponed to our very last chapter. Yet even here we can understand something of its nature.

In the first place, capitalism emphasizes the individual rather than the group. To the end that the individual may achieve the maximum realization of his desires, capitalism involves a maximum of individual liberty: freedom of contract as contrasted with serfdom, the rights of private property, and the stimulation of private individual enterprise. Men are motivated by acquisitiveness—by the desire to make a profit—and the government steps in only to the extent of making the minimum number of rules that are necessary to make certain that the individual, in his quest for profits, does not do too great damage to others. Instead of building a society centered around the church or the nobility, capitalism builds a society centered around want satisfaction—around the fulfillment of the desires of the common man.

The fact that capitalism emphasizes the satisfaction of wants has led many people to call it "materialistic," yet there is no necessity that it be so. Insofar as men's wants are for material things, the capitalistic system is materialistic; but to the extent that men crave things of the spirit rather than things of the flesh, and to the extent that they make use of worldly means to satisfy their spiritual desires, capitalism is not materialistic. The church, the school, the government, all have economic problems, and the capitalistic system is one means of attempting to solve those problems, whether of a material nature or not.

The capitalist system has also been one in which trade and commerce have played an important part, and one in which the price system has been paramount. Things are valued, not by government fiat, but by the free interplay of economic forces on the market. Buyers and sellers are left free to buy and sell as they choose. In fact, capitalism is the system in which individual liberty has reached its highest point, and in which restriction and regulation are reduced to a minimum.

Although it is by no means the most important characteristic of capitalism, nevertheless, as the name implies, capitalism is a system that uses capital on a large scale. It was under capitalism that great factories arose and that investments in productive equipment reached levels that stagger the imagination. The typical capitalistic enterprise requires such stupendous investments that the ordinary man cannot own the tools and equipment necessary to produce. Production is taken over by huge aggregations of capital, contributed by hundreds or thousands of persons. As a result, the contributions to production tend to become separated, some people contributing their own toil and effort while others contribute their property or their skill at organizing and directing industrial ventures. Capitalism is

marked by specialization, by the disappearance of the Jack-of-all-trades and the emergence of the specialist. Most often remarked is the separation of the worker from the ownership of capital; just as marked in practice are the distinctions between different kinds of workers and between investment on the one hand and management or control on the other.

Of course, society did not step at once from a medieval economy to a full-fledged capitalism. The change was a slow and fitful one. But the technical changes of the Industrial Revolution hastened greatly the growth of capitalism which had started a century or so before, and today most people think of capitalism as dating from the Industrial Revolution and as having been caused by it.

3·21. Summary. We have now brought our story down to at least the beginnings of modern industry. The development of new machines and the harnessing of new sources of power have continued down to the present. The scale of production has continued to grow, until there are now single firms having an investment of millions of dollars and employing thousands of men. We shall continue to touch on the problems connected with these modern developments as we go along, for the Industrial Revolution is still with us, perhaps even still gaining momentum.

The student is apt to think of the customs and habits and institutions with which he is familiar as being somehow "superior" or inherently "right." He is apt to think of the society in which he has grown up as being "proper" or "natural" or as being based on some fundamental laws of reason. Yet, it is important for him to remember that in all times and places people have naïvely made the same assumption. The inhabitant of the medieval English manor doubtless thought of the relationship of the villein to the land as natural and proper. Whoever sought to upset it was seeking to undermine society. Similarly, the ordinary individual who lived in the English town at the time of the handicraft stage probably thought that it was right for the gilds to have such powers as they exercised—it was natural to do things that way.

The most important thing that the student can learn from any study of history, even from such a short and cursory review as that given in this chapter, is that most of our own social institutions are relatively new and that our descendants will almost certainly think of them as peculiar and by no means founded on "natural laws." Most of us live today in ways that would have seemed undesirable or even abhorrent to our ancestors, and presumably will also seem so to our descendants. We should try to develop a sympathetic understanding of our present problems without assuming that our present methods are inherently any better than any others. Social objectives and goals change from time to time, and it is quite possible that the very methods that we find most useful in attempting to reach our goals will

be useless or even harmful to our descendants who wish to reach other goals. As we describe in later chapters the activities and institutions of our present-day society, the student should think of them, not as ultimate perfection, but as positions which have so far been reached in our "progress" along a path that leads we know not where.

Questions

1. What characteristics of modern American life might make people of other cultures feel sorry for us?

2. What are some of the characteristics of modern American life of which our ancestors of 10 generations back might disapprove? (Ten generations would be roughly 300 years.)

3. The text suggests that men are rather poorly endowed physically as compared with other animals—that other animals are swifter, stronger, keener of sight, have better sense of smell, etc. What purely physical advantages do men have over other animals?

4. In what ways has the weakening of superstition since the Middle Ages been of economic advantage to men?

5. Have superstitions entirely disappeared in modern times? List some modern ones.

6. We think that those tribes which tried to cure disease by invoking spirits though their medicine men were superstitious. In what way does our own modern field of medicine differ from those primitive methods, so that you would be justified in saying that the doctor who removes an appendix is not thereby practicing superstition?

7. Name some illustrations of direct appropriation in modern life.

8. What does the text mean by saying that property is immaterial and intangible? Is a dwelling house property? Is it material and tangible?

9. College commencements, we are told, are an institution in the sense in which social scientists use the term. Name some other institutions. Does your own college have any peculiar institutions of its own? What characteristics do they have which enable you to say that they are institutions?

10. A quotation from Governor Bradford appears on pages 19 to 20. What, as you understand it, were the reasons that actuated him in allowing some degree of private property in land, rather than continuing the common ownership of land? What have been some disadvantages of private ownership of land?

11. In what principal ways did the medieval English manor differ from a small American village of today? Assume that both communities have the same population.

12. What were the advantages of the three-field system which probably led the people of the manor to adopt it? What were its disadvantages as compared with present agricultural systems?

13. Why did the specialization of the town economy tend to increase the amount of trade and the use of money? What is the connection between specialization on the one hand, and trade and money on the other?

14. The medieval fair was in some respects like the country fair of modern times, but in some respects different. What were the points of similarity? The points of difference?

15. The medieval gilds had some of the aspects of modern trade-unions. How were the two alike? How different?

16. What were the major differences between craft gilds and merchant gilds?

17. In time of war, governments tend to take on certain activities and functions and to make certain regulations that are not common in peacetime. Is a wartime government more or less mercantilistic than a peacetime government?

18. Would you say that the policies of *laissez faire* have been getting more or less popular in recent years?

19. The text says, "Most of us live today in ways that would have seemed undesirable or even abhorrent to our ancestors." Give a few examples.

20. Why do you suppose the word "progress" appears in quotation marks in the last sentence of this chapter?

PART TWO. Consumption

4. CONSUMER INCOMES

4·1. Kinds of Income. As we begin our study of consumer problems, it is natural for us to start with some attention to the problem of consumers' incomes, because the kinds and amounts of things that people consume depend largely on their incomes. The income is usually at least roughly fixed and beyond the control of the family. They have to adjust their consumption habits to their incomes.

First, perhaps, we should ask what income is. Since the term is a common one, this may seem unnecessary, and we might take it for granted that the meaning of the term is known. One man, let us say, has an income of $1500 per year and another man an income of $3000. The second man gets twice as much income as the first. But immediately we see that such figures on *money income* do not give a complete picture. Possibly the first of the two men just mentioned gets $1500 in cash, but in addition gets board and room worth $500 per year. In such a case the money incomes are not reliable indexes of the relative status of the two men. In addition, perhaps the first man lives in the country where prices are low, while the second man lives in the city where food, clothing, and shelter are more expensive. Again it would be obvious that a direct comparison of the two money incomes would be misleading. This same problem would arise if we were comparing the income of a man in 1933 (when prices were low during the depression) with his income in 1945 (when prices were high during the war). If the man received the same money income at both times, he would find himself much worse off at the latter period than at the former, because his money income would not buy so much.

For these reasons economists are apt to be more interested in men's *real incomes* than in their money incomes. A man's real income consists of the goods and services that he actually consumes. Possibly our country man with the $1500 income consumes just as much as the city man with an income of $3000. In such a case, their real incomes would be the same. Our difficulty

with real income lies in the fact, of course, that when men consume different things it is impossible to say which consumed more. One man attends the movies and the big-league ball games. Another spends the same time reading, fishing, or shooting clay pigeons. Which has consumed more? Which is greater, 2 movies or 2 clay pigeons? It is patent that we are dealing with quantities that are not commensurable. We can, of course, evaluate them in monetary terms, and say that one of the men consumed goods and services that cost $1500 while the other consumed goods and services that cost $3000. This gets us back to money incomes and all the difficulties encountered in the preceding paragraph. Either concept—that of money income or that of real income—has its drawbacks and its advantages.

Yet neither of these concepts is final or complete. When we come down to it, we realize that two men might have exactly the same money income and exactly the same real income, yet one might be happy and the other discontented. Suppose we were to take a typical American to Labrador, leaving him there with an Eskimo family to live exactly as they did. Let him have exactly the same money income (if any) which they have, and give him the same food, clothing, entertainment, etc., which they have. The chances are that the American will be unhappy. He will miss the things to which he has been accustomed. His money income and his real income will be the same as those of the Eskimo, but his *psychic income* will be very different. Basically the most important income that men get is the satisfactions obtained in life. Money income and real income are but a means to an end—psychic income. The money income is wanted only to buy real income, and the goods and services that make up the real income are wanted in turn only for the satisfactions that they yield—for their psychic income.

Unfortunately psychic income is even harder to measure and appraise than is real income. As we move from money income to real income to psychic income, we move to concepts which are harder and harder to measure and compare but which are more and more important to the people who receive them. An intelligent man, offered two jobs, does not by any means always take the one that offers the higher money income. Not only in theory, but in cold practice, he tries to evaluate all the imponderables, such as the hours of work, the type of men with whom he is to associate, the social prestige and standing of the position, etc. These will all be parts of the psychic income, even if they are not included in the money income.

4·2. Money Incomes in the United States. Yet after this emphasis on the importance of the nonmonetary sorts of income, we turn back to a description of the distribution of money incomes, because these incomes are the only ones which can be compared quantitatively and on which we can get some actual data. To be sure, our data even on money incomes are in-

complete and inaccurate, but they are good enough to give us a general picture of the income situation in the United States.

Table 4·1 shows percentages of the families in the United States with incomes of various sizes at three different periods. The figures are computed by somewhat different methods and so are not entirely comparable, but they will give us at least some idea of the situation relative to money incomes. Of course, no one expects the student to remember all the detail of a table such as this; but it is important to get from it, as from other similar tables in the text, a general idea of the situation depicted. This we can do more quickly and easily by studying Figs. 4·1 and 4·2, which are based on Table 4·1. We see right away from the charts that the conditions differed

Table 4·1. Distribution of Incomes in the United States, 1929, 1935-1936, and 1942*

Income group	Percentage of families		
	1929	1935–1936	1942
Under $500	7.6	17.0	8.5
$ 500–$ 1,000	13.8	29.5	16.1
1,000– 1,500	20.9	22.1	16.0
1,500– 2,000	17.1	13.1	14.5
2,000– 2,500	11.6	7.5	11.3
2,500– 3,000	7.2	3.7	8.0
3,000– 4,000	8.9	3.4	11.2
4,000– 5,000	4.5	1.2	6.4
5,000– 7,500	4.4	1.0	4.6
7,500– 10,000	1.6	0.6	1.5
10,000 and over	2.3	0.8	1.9

* The data in this table are adapted from the following:
For 1929, from Maurice Leven, H. G. Moulton, and Clark Warburton, "America's Capacity to Consume," p. 54, Brookings Institution, 1934.
For 1935–1936, from "Statistical Abstract of the United States, 1942," p. 363, U. S. Department of Commerce.
For 1942, from "Statistical Abstract of the United States, 1943," p. 393, U. S. Department of Commerce.

at the different times. In fact, we note that a far larger percentage of American families was in the low-income brackets in 1935-1936 than in either of the other two periods. Yet the most interesting thing about the chart is not the differences, important as they were, but the similarities. We see that in each of the three periods the great mass of American families received what most students would consider pitifully low incomes. The bulk of the families is concentrated far toward the bottom of the income scale, with a few families stretched out at the upper end to incomes of astronomical size. Although the most common, or modal, incomes were somewhere in the neighborhood of $1000, nevertheless there were, even in the depths of the

depression, some 20 or 30 families each year with annual income of over a million dollars; and in 1929 there were over 500 such families. These highest incomes cannot be shown on our charts, which run up to incomes of only about $10,000. Fig. 4 • 2 shows us that in 1929 and in 1942 half the families had incomes below about $1700; while in 1935-1936 half had incomes below about $1000. To be in the top tenth a family needed an income of about

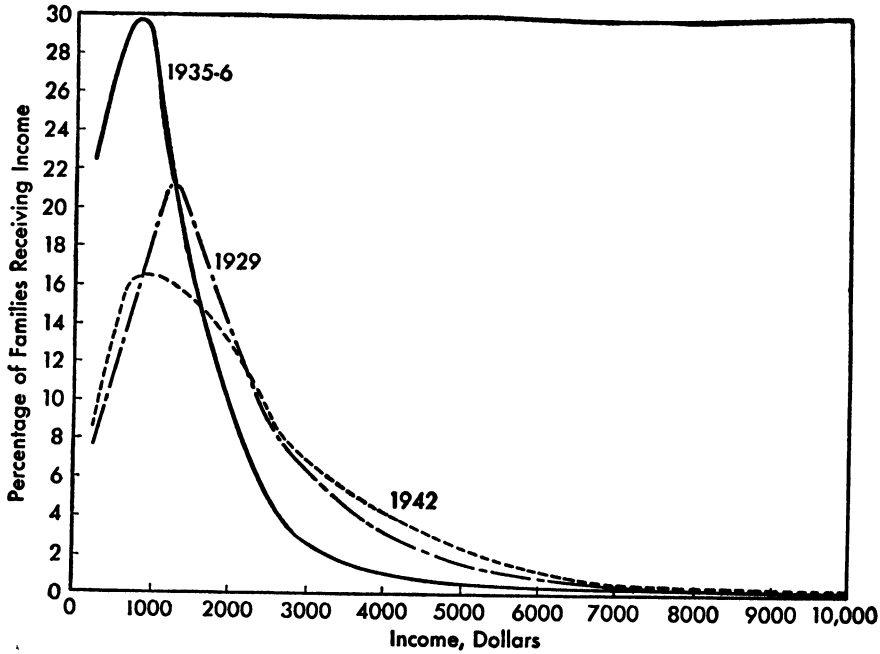


FIG. 4 • 1.—This chart shows the distribution of family incomes in the United States in 1929, 1935-1936, and in 1942. To read the chart, select any desired income on the base line. Erect a perpendicular from this point on the base line to the curve, and where the perpendicular cuts the curve, draw a horizontal line to the vertical scale at the left. The figure on this vertical scale will show approximately what percentage of the families had incomes within $250 of that selected. For example, about 10 per cent of the families had incomes within $250 of $2000 in 1935-1936; while in 1942 about 13 per cent of the families fell within this same range. It will be seen that a larger percentage of the families fell in the low-income classes in 1935-1936 than in 1942.

$4500 or more in 1929 or in 1942; one could have had the same distinction in 1935-1936 if his income had exceeded about $2500.

Numbers are not, by themselves, particularly exciting; but, when we realize what these mean in terms of men's lives, they take on new significance. The economic problem faced by a family with an income below $500 is a difficult one; yet even in years of high income from an eighth to a tenth of all American families find themselves in this position. It might be a good

exercise for the imagination to compute the cost of a year's college education, and to see what part of an average family's annual income would be represented by the figure.

In good years or in bad, most incomes are "low." After we get up to incomes of $1000 or $1500 a year, we have passed the peak and the number of families with higher incomes becomes rapidly smaller, until finally we reach those enormous and spectacular incomes which sometimes run to as much as $10,000 a day!

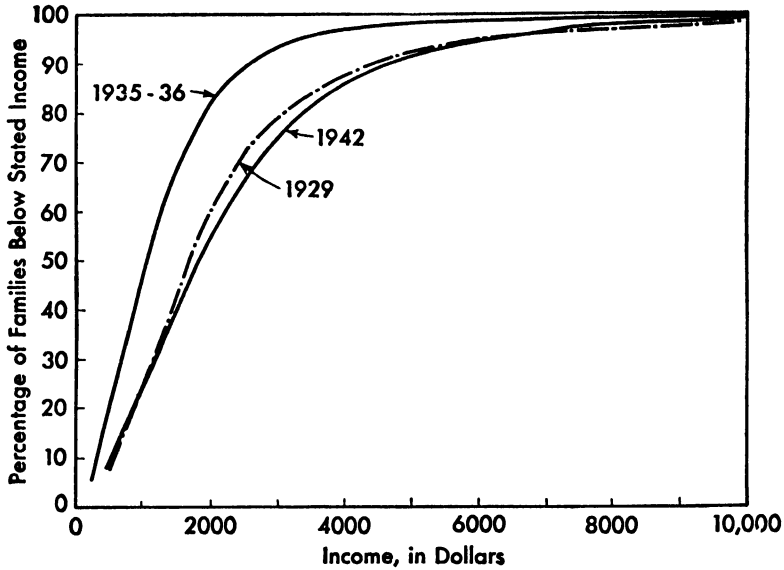


FIG. 4·2.—This chart answers the question, "What percentage of American families got incomes lower than given amounts?" For example, in 1935-1936 about 93 per cent of the families had incomes below $3000; while in 1942 about 74 per cent were below $3000. The chart shows clearly that smaller percentages of families were at the low income levels in 1929 and in 1942 than in 1935-1936.

4·3. Inequality in the Distribution of Incomes. Study of the data of Table 4·1 on page 42 would show that the average income in 1935-1936 was about $1500, while in 1929 and in 1942 it was roughly $2700. Had the incomes been equally divided, these are the amounts that each family would have received. But reference to Fig. 4·2 indicates that about 70 per cent of the families of the United States received less than the average income in each of these three periods; while in each of the three periods some families received far more than the average. The figures show that there is actually great inequality in the distribution of incomes.

There is no simple, easy way to measure the degree of this inequality, nor to make accurate comparisons between countries and periods. Economists

have been in the habit, however, of getting a rough and approximate picture of the degree of inequality by means of the Lorenz chart, such as that shown in Fig. 4 • 3. This chart shows along the base line percentages of the population, starting with the poorest, and along the vertical axis percentages of the total income received. For example, it tells us that in 1929 the 80 per cent of the families at the bottom of the income scale received just over 40 per cent of the total income. Other points of the chart are read similarly. Obviously, if there were equal distribution of incomes, any 10 per cent of the families would receive 10 per cent of the income, and any 50 per cent of

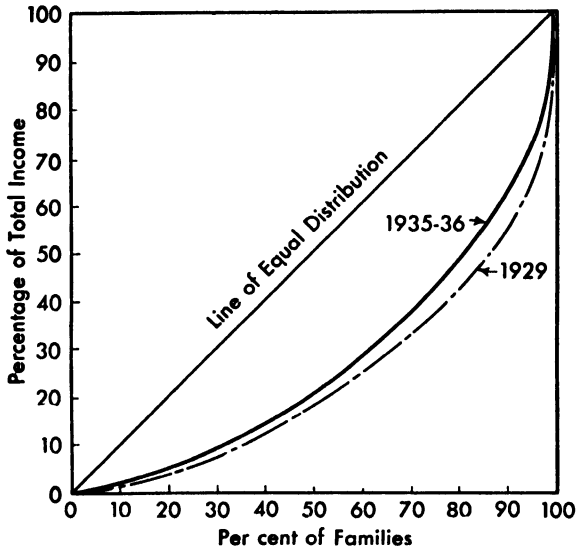


FIG. 4 • 3.—Lorenz curves of income distribution in the United States, 1929, and 1935-1936. The 1935-1936 line lies closer to the diagonal, and hence indicates a more equal distribution of wealth, than the 1929 line.

the families would receive 50 per cent of the income. Thus the straight diagonal line on the Lorenz chart shows what would be true if the distribution were equal. But with an unequal distribution, the poorest 10 per cent of the families must get less than 10 per cent of the total income, and our actual line will fall below the diagonal, as do the lines in Fig. 4 • 3. The more the curve departs from the straight diagonal, the more unequal is the distribution of incomes. Thus in Fig. 4 • 3, the fact that the line for 1929 lies farther from the diagonal than does the line for 1935-1936 is taken to indicate that the 1929 income was less equally distributed than that of 1935-1936. The chart indicates that there is a very considerable departure from equality in the distribution of American incomes.

4·4. Causes of Inequality. To be sure, equality in distribution is hardly what one would expect. After all, men differ in height, weight, and in keenness of vision and in strength and intelligence. Differences among men are commonplace, and one might naturally anticipate differences in income also. It is common to divide the causes of such differences into two groups: heredity and environment, or nature and nurture. In the case of differences in income, we may divide them into those differences which depend on the men themselves and those caused by the society in which they live.

Some economists tend to give very great weight to the first of these classes of causes, while others tend to explain inequalities almost exclusively in terms of the latter class. It is impossible to give any categorical and certain statement as to their relative importance, although it is easy to pick individual cases that do seem to fall wholly within the one class or the other. Surely it would be difficult to rule out either class of causes entirely.

If we look first at the differences in income that are the result of differences in the income earners themselves, we see that there are indeed important differences in income which can be explained in terms of the individual's intelligence or the lack of it, by his strength or lack of it, by his persistence and energy or by their lack. Some men can hit 50 home runs in a season of big-league ball, while most men cannot. Such an ability commands a price on the market; such an ability a decade or two ago enabled a man to earn more than the President of the United States. Sobriety, ingenuity, cooperative spirit, initiative, that strange mixture of qualities which we call "personality," and many other human qualities are more or less desirable, and men who have them can command, on the average, higher incomes than can those who are less gifted. Or, if we put it another way, we can say that, on the average, the higher incomes represent greater contributions to society. The people with these higher incomes produce more, on the average, than do the people with low incomes, of whatever it is that society wants produced —whether wheat or automobiles or home runs. Although anyone can think of individual exceptions—of poor men who have made great social contributions and of rich men who have made no social contribution at all—nevertheless, as long as we stick to averages, our statement is probably a safe one.

On the other hand, those who emphasize the importance of social institutions as causes of differences in income point to such institutions as inheritance, and to the existence of monopolistic privileges, as the basis for great fortunes. They would argue that the son of a successful and wealthy automobile manufacturer is more likely to become a corporation executive with high income than is the equally gifted son of a bootblack; or that a man who obtains, by whatever means, a monopoly over a commodity that consumers want, has much more opportunity than does the average person of

equal gifts to profit. Some large incomes arise because their recipient was lucky, or because he had a wealthy father, or because he married the boss's daughter, or because he profited at the expense of the public.

4·5. Is Equality of Distribution Desirable? In the past few pages our effort has been to describe in a general way the facts without any judgment as to what is desirable or undesirable. Yet, the English language and our democratic predilections being what they are, it is likely that most students as they have read have assumed implicitly that inequality in the distribution of incomes is undesirable, and that our statements of existing inequality are an indictment of the existing situation. Such an assumption should, at least, be challenged, and one who holds such a belief should hold it after some contemplation rather than by intuition.

We might start with the statement that, even if we wanted complete equality of incomes, it would be well-nigh impossible to get it. Reference to Sec. 4·1 will remind the student that, even if we did get complete equality of money incomes, real incomes would still vary; and if we got equality of real incomes, we would have important differences in psychic income. If equality of income were an acceptable ideal, it would certainly be an unattainable ideal.

But, except as a basis for argument, most people would not be willing to admit that they would want equality of income even if it were possible. They would argue that the industrious, capable, thrifty, energetic worker should be allowed more income than his shiftless, unimaginative, wasteful, lazy fellow. It is important not to overlook the part that income plays as an incentive to work and, if the lazy and the industrious share alike in income, there is no great incentive to work hard.

We are getting here out of the field of science and over into the field of ethics or morals (see Secs. 2·1 to 2·5). As is commonly true in this latter field, we find that we can set up pretty convincing arguments in support of very different principles. One man will say that it seems fair to him to divide the incomes in proportion to men's productivity, giving the most to those who contribute the most and the least to those who contribute the least. Another man will argue that it seems fair to him to give incomes to men in proportion to their needs, giving much to those who need much and little to those who need little. Either principle sounds good; we almost instinctively accept it and agree with it. Yet a moment's reflection will show that the two principles are very different. If we allot the incomes on one of these two bases, we certainly shall not have alloted them on the other.

At this early stage in our study we shall have to leave this question of the desirability of equal distribution of incomes hanging fire, after having warned the student against too prompt acceptance of any ready-made formula

in such matters. But possibly we should look ahead enough to say now, even before we have had the opportunity to discuss the reasons for it (see Sec. 5 • 11), that most economists seem to agree on two statements:

1. There is no reason to suppose that complete equality in the distribution of incomes would be desirable, but

2. Almost certainly it would be socially advantageous to distribute incomes somewhat more equally than we now do.

4 • 6. Methods of Reducing Inequality. Every society has methods of reducing inequality in income. In the first place, we should point out that there is not so great a disparity in real income as there is in money income, since, if one family has an income of $1000 and another one of $10,000, the latter family will not consume ten times as much as the former. The family with the $1000 income will probably spend the entire income for goods and services—for real income. The family with an income of $10,000 will save perhaps a quarter or a third, and convert only about 70 per cent into goods and services.[1] In addition, there is reason to suppose that psychic incomes are more equally distributed than are real incomes, since a family that consumes ten times as much in the way of goods and services as another family almost certainly doesn't, on the average, get ten times as much pleasure from the consumption.[2] However, despite these ameliorating circumstances, societies do take various steps to reduce the inequality in incomes. Heavy taxation of large incomes and of inheritances, with the proceeds of the taxes being used largely for benefits to low-income groups in society, is a common method. Free public education has been another. One of the purposes of such education is to permit the children of the poor, who would not be able to afford educational opportunities otherwise, to train themselves for more highly paid positions; thus giving more competition for these highly paid posts and thereby reducing incomes at that level, while reducing the competition for low-paid jobs and thus raising somewhat the incomes at the lower levels. Various programs of social insurance and charitable relief, and measures for the control or elimination of monopoly, tend in the same direction. Indirectly many measures designed to improve health or to reduce the number of physically and mentally unfit serve the same purpose.

4 • 7. Inequality Is Universal. So far we have contented ourselves with describing the situation in the United States. As was pointed out in Sec. 4 • 3, there is no simple, easy, reliable way for measuring inequality, no good way

[1] For example, we find that American families with incomes of $1000 in 1935-1936 spent an average of about $1030 and were running slight deficits; families with incomes of $10,000 were spending an average of about $6750. These figures are approximations, based on data from p. 364 of the 1942 "Statistical Abstract of the United States." Even if the figures in this footnote are not exactly correct, they show the nature of the actual situation.

[2] This is the result of a phenomenon which we shall call "diminishing utility" when we discuss it in due course. See Sec. 5 • 8.

by which to compare income distributions of the United States with those of other countries. Moreover, income data are notoriously inaccurate and difficult to compare. Hence we shall have to content ourselves here with the general statement that studies in many other countries, similar to those which we have used in describing the situation in the United States, have always shown without exception a very considerable departure from equality in distribution of incomes. These studies tend to indicate that the average income in the United States is one of the highest in the world,[1] and in addition that American incomes are less unequally divided than those of most other countries. One well-known American economist has said,[2]

> It is now well known that the United States have become by far the richest nation, with the widest dissemination of general comfort and well-being that ever existed. The average wages, real wages in purchasing power, were in 1930, as computed by the International Labor Office, at least one-half again as much as those of the next leading nation, Great Britain, near to twice the average real wages of Germany, and between two and three times those of Italy.
>
> Actual wages in this country are at least ten times as high as the average wage in the better class of factories in India and perhaps also of China.

More recently a leading English economic journal[3] commenting on the situation from 1938 to 1941, said, "On the average . . . , however, it seems probable that the average British income lies somewhere between three quarters and seven eighths of the American—80 per cent might be a fair average figure to take for *prima-facie* comparisons."

Every statistical study has shown, in every country and at every time, a situation basically like that depicted in Figs. 4·1 to 4·3. In every country, there are great differences in income. In every country, the great mass of income earners are bunched closely together near the extreme lower end of the income range, with the number of incomes getting smaller and smaller as income increases until there are, at the top, a few fabulously large ones. In fact, the similarities in distribution of incomes from country to country and from time to time have been so great as to lead some mathematical economists to formulate "laws" of income distribution by means of which they believe that the distribution of incomes for any country can be described. These laws are altogether too mathematical in nature for treatment here; moreover, their validity is seriously disputed. However, the student with mathematical interests may wish to read the short note on

[1] See Table 30·3.

[2] Mr. Carl Snyder, for many years with the Federal Reserve Bank of New York, on p. 106 of the "Report of the Third Annual Research Conference on Economics and Statistics," Cowles Commission for Research in Economics, 1937.

[3] *The Economist*, vol. 143, no. 5167, p. 300, Sept. 5, 1942.

one of these laws which appears as Note 1 of the Mathematical Appendix.[1] Students without these interests will have what information they need on the subject if they remember that the situation which we have described in the United States is universal in kind and that, in degree, the United States seems to differ from other countries primarily, if at all, in having somewhat higher average incomes and somewhat greater degree of equality in their distribution.

4·8. Who Gets the Income? Students of economics are interested to know not only how big the incomes are, but who gets them. We shall give major attention to this problem in Part VI, Chapters 30 to 33; but we should make a few comments on it now. Table 4·2 shows the percentage distribution of income in the United States for alternate years from 1930 to 1942. Although such procedure is very approximate and open to justifiable criticism, we may nevertheless consider that, roughly, items 1, 2, 3, and 5 represent income in return for labor, while items 4, 6, and 7 represent incomes for the use of property. The proprietors of item 5 include thousands of farmers, small shopkeepers, etc., most of whom are actively working in their own enterprises. If we add together the "labor" items and the "property" items, making due allowance for the inaccuracies involved, we get Table 4·3, which indicates that perhaps the income of the United States is divided, roughly, 85 per cent to labor and 15 per cent to property. Although there are variations from year to year, there is no well-defined trend evident in the short period covered. Another source gives us estimates on the same subject which are very different. We find English estimates[2] for 1939 and for 1941 as in Table 4·4. Again, if we think of the land rents and the incomes from interest and profits as property income and the remaining item as labor income, we get the data of Table 4·5, which have been put into percentage terms for comparability with Table 4·3. These figures seem to indicate that about a third of the income goes to property, and two-thirds to current human effort. Other classifications would certainly yield still other figures, and it is not our purpose to argue over the discrepancies and the differences in definition. It is our intent merely to point out that a large share of social income does go to men in return for their current effort, as compared with the share that some men get for their property.

4·9. The Poor Always Ye Have with You. Economic history records centuries of slow, intermittent, but certain progress which has enabled men to better their lot, until today even the poor have many things to consume which the wealthiest men could not have had a thousand years ago. Yet our study in the earlier sections of this chapter has indicated that great numbers of men are still pitifully poor. This is in part due, to be sure, to the

[1] See p. 883.

[2] *The Economist*, vol. 143, no. 5167, p. 300, Sept. 5, 1942.

Table 4·2. Percentage Distribution of National Income, United States, 1930-1942*

	1930	1932	1934	1936	1938	1940	1942
1. Wages of employees in private industry	61.9	65.2	56.8	53.0	54.8	54.0	55.7
2. Wages of government employees	7.4	12.5	9.3	8.5	9.3	8.7	11.3
3. Relief and other supplementary income	0.7	1.5	3.8	4.8	6.1	4.7	2.8
4. Net income of corporations	2.5	−9.0	1.0	5.8	2.7	7.4	6.3
5. Net income of proprietors	14.6	12.0	15.2	16.8	15.7	15.6	16.8
6. Interest	8.7	14.0	10.5	7.9	8.0	6.6	4.5
7. Rents and royalties	4.2	3.8	3.4	3 2	3.4	3.0	2.6

* Adapted from data in Table 422, p. 384, "Statistical Abstract of the United States, 1943," Item 2 excludes subsistence of the armed forces. Note that the figure for corporate income in 1932 is negative, representing a net loss.

Table 4·3. Percentage Distribution of National Income in the United States, 1930-1942
Data derived from Table 4·2

	1930	1932	1934	1936	1938	1940	1942
Labor income	84.6	91.2	85.1	83.1	85.9	83.0	86.6
Property income	15.4	8.8	14.9	16.9	14.1	17.0	13.4

Table 4·4. Distribution of Income by Classes, United States and United Kingdom 1939 and 1941

Type of income	United States, billions of dollars		United Kingdom, millions of pounds	
	1939	1941	1939	1941
Rent of land	3.5	4.1	373	379
Interest and profits	22.0	31.5	1,351	1,722
Labor	46.8	63.1	2,871	4,237

Table 4·5. Percentage Distribution of Income, United States and United Kingdom 1939 and 1941. Based on data of Table 4·4

Type of income	United States		United Kingdom	
	1939	1941	1939	1941
Labor	64.8	63.9	61.6	66.8
Property	35.2	36.1	38.4	33.2

fact that poverty is a relative rather than an absolute thing. As we saw in Sec. 1·8, men's wants are not fixed and, when we satisfy one set of wants, others spring into being. If we give the "poor" everything they want, they discover new wants.

Poverty is largely a state of mind. When one has a strong and insistent feeling of privation, he is poor. In some crude and uncivilized society the "wealthiest" man of the tribe may have poor and irregular meals and little or nothing to wear; yet, if he knows of nothing better and if all around him he sees others who are in even poorer state, he feels rich rather than poor. If, in our society, a man has a cheap car and a low-priced radio and cotton stockings for his wife and daughters and a home in the less fashionable district, he feels poor rather than rich, despite the fact that he has much more to consume than does the rich tribesman of the preceding sentence.

A man's real income—the goods and services that he actually has to consume—we call his "level of living." If the people of the United States have, on the average, more things to consume than do the people of India, we say that the American level of living is higher than that in India. But in any society people set up goals of consumption. They recognize certain patterns of consumption as desirable. They set up standards for themselves. These desired patterns of consumption we call "standards of living," in contrast with the levels of living just described. It is important for the economist to remember that people differ in their standards of living as well as in their levels of living. It is quite possible that people, who do not live as we do, do not want to live as we do. Their level of living is different, and so is their standard of living. A man's standard of living is not the best level that he can imagine but is a matter of custom and experience and familiarity. We can all imagine things that we would like to have, but we do not seriously miss their absence. These things are not parts of our standard of living. Anything becomes a part of our standard of living when it becomes such a part of our habitual consumption pattern that we would not be happy without it —that we would experience a real sense of lack and frustration if we were to be deprived of it. Our great grandparents did not have electric lights or running water in the house or radios or automobiles. They were not part of their level of living; but neither were they part of their standard of living, and they did not feel the lack of them. Today many a family would feel a distinct sense of loss if they could not have these things. Today they are part of the standard of living of thousands of people.

Poverty exists when people have levels of living that are decidedly lower than their standards of living. Since our standards are themselves built, at least in part, on our knowledge of what other people consume, it is evident that inequality in the distribution of wealth and income fosters poverty. The

people with the smaller incomes see their contemporaries with larger incomes consuming various commodities which the small incomes will not afford. Had no wealthier people ever consumed these things, no poorer people would have craved them. Comparison of our own consumption with that of wealthier people leads us to set a standard of living higher than our level of living and brings us discontent. It is in this sense that inequality is in itself a cause of poverty, and it is in this sense that we understand the Biblical statement that "the poor always ye have with you." The level of living has increased almost unbelievably in the past few centuries, but the continued inequality in the distribution of the good things of life continues the problems of poverty. The problems that arise from inequality in the distribution of wealth and income are some of the most pressing, important, and difficult of the problems that every society faces. Inequality gives rise to a majority of the current applied problems of economics which we shall discuss from time to time in later chapters of this text.

4 • 10. Engel's Laws. One final thing we should note in our preliminary discussion of income: As men's incomes rise, there seem to be regular and predictable changes in the consumption pattern. One might imagine on a priori grounds that, as incomes increase, different peoples would use their increased income in different ways. Possibly the people of one nation would use their new resources to buy better housing, those of another nation to buy better medical care, while those of another nation would dress better. To be sure, it does tend to be true that people with higher incomes buy more of all the good things of life than do people with lower incomes; but if one's income is doubled he does not double his purchases of everything. He increases the purchases of some things tenfold, he doubles the purchases of other things, he increases only slightly the purchases of some things, and he actually reduces somewhat or cuts out entirely the purchases of others (things which he did not really consider "good" at all but which he consumed as lesser evils while he could not afford better).

Some of the changes that actually take place are depicted in Table 4 • 6. As we glance down any column in the table, we see what happens, on the average, to the corresponding per capita expenditure as the income rises. The table shows us, for example, that as incomes rise the percentage of the per capita expenses for food decreases steadily and sharply while the percentage of the per capita expenses for clothing rises slowly. We see, for example, that after families get up to about the $3000 income level, their per capita expenditures for housing exceed those for food, but that at lower income levels this situation is reversed. At high income levels the per capita expenditure on automobiles becomes almost as high as that on food.

These facts have been understood in a hazy and general sort of way

for a long time, but they were first seriously studied in a quantitative way by Ernst Engel, a German statistician, who published his findings on the subject in 1855. Engel noticed that poor people spend a larger percentage of their income on the necessities of life than do wealthy people, and he published data somewhat like those of Table 4·6, showing the changes in the consumption pattern at various income levels in Saxony. Although Engel himself seems to have stated but two of them formally (the first and last), nevertheless his followers have come to state four of Engel's Laws, as follows:

Table 4·6. Percentage Distribution of Expenditures at Different Income Levels, United States, 1935-1936*

	Percentage of income spent for					
	Food	Housing	Clothing	Automobiles	Medical	Other
Under $500	44.5	31.9	7.6	2.1	3.8	10.1
$ 500–$ 750	42.4	29.9	9.2	3.0	3.6	11.9
750– 1,000	40.3	29.2	9.5	4.2	3.7	13.1
1,000– 1,250	37.8	29.3	9.6	5.7	3.9	13.7
1,250– 1,500	36.2	28.6	10.0	6.5	4.1	14.6
1,500– 1,750	34.5	28.5	10.1	7.6	4.5	14.8
1,750– 2,000	32.8	28.8	10.2	8.6	4.5	15.1
2,000– 2,500	31.3	28.4	10.8	9.7	4.6	15.2
2,500– 3,000	29.8	28.7	11.2	10.2	4.7	15.4
3,000– 4,000	28.1	29.3	11.7	10.3	4.9	15.7
4,000– 5,000	25.8	29.6	12.6	11.3	4.9	15.8
5,000– 10,000	23.2	30.6	12.6	11.5	5.7	16.4
10,000– 15,000	19.7	32.2	13 5	11.1	4.2	19.3
15,000– 20,000	19.2	29.6	13 6	10.4	5.0	22.2
20,000 and over	15.2	33 2	14 3	12.1	6.1	19.1

* Adapted from data in "Statistical Abstract of the United States, 1942," Table 386, p 365 Figures are percentage expenditures per consumer unit The housing item above combines items for housing and household operation in the original.

As the family income increases,

1. The percentage of the income that is spent for food decreases.

2. The percentage of the income that is spent for clothing stays approximately fixed.

3. The percentage of the income that is spent for rent, fuel, and light stays approximately fixed.

4. The percentage of the income that is spent for sundries, such as education, health, amusement, savings, and luxuries, increases.

More recent studies do not seem to bear out the second and possibly not the third laws (the ones that Engel does not seem himself to have enunciated), since other studies seem to conform with the tendency of Table 4·6,

to show the clothing expense slowly increasing as income rises, while the housing expense either stays constant or falls slightly.

The facts seem to be that the demand for food has the first call on the family's resources and, if those resources are meager, a large proportion of them has to go for food. But the capacity of the human stomach is limited, and as income increases the expenditure for food does not increase proportionately. To be sure, the absolute expenditure for food does rise as the family eats more expensive kinds of food, but the increase in income is faster than the increase in food expenditure. Doubling the income increases, but does not double, the food expense. The cost of housing about keeps pace with the increasing income, but perhaps not quite. The other expenses, however, increase even more rapidly than the income. Thus we can state Engel's Laws in their modern and more exact form something like this: *As we move from any given income level to a higher level, we find definite and predictable changes in the average pattern of consumption, with the percentage spent for food diminishing rather rapidly, the percentage expenditure for housing staying constant or diminishing very slowly, and the percentage expenditure for all other major items of the budget increasing more or less rapidly.*

It is sometimes said that the items classed as luxuries show the greatest percentage increases, and the more luxurious they are the greater the percentage increase; but it is quite likely that this is confusing cause and effect. It might be better to say that we call things luxuries because they are used only to a minor degree by the poor and in increasing degree by the more well to do, rather than to say that they are used more by the well to do because they are luxuries.

Questions

1. Which do you think would be largest, a man's money income, his real income, or his psychic income? Does the question make any sense?

2. After inspecting Table 4 • 1, what would you say was a typical income for American families?

3. The President of the United States receives a salary of $75,000 a year. Does this give him the largest income in the United States? About where does it place him in the distribution of incomes?

4. Do you think that the largest incomes in the United States are mainly the result of environmental or of hereditary differences in men, as we have used these terms? Explain.

5. List some reasons why equality in the distribution of incomes would be desirable, and some reasons why it would be undesirable. In making such a list, are one's decisions in the field of science, art, or religion?

6. Why would it be difficult to get complete equality in money incomes, even if we wanted to?

7. The text states that we would get very different results if we distributed incomes in accordance with productivity than we would if we distributed in accordance with need. What groups of people would get little on the first basis and much on the second? What groups would get much on the first basis and little on the second?

8. What are some of the inaccuracies involved in the assumptions that lie behind the making of Table 4 • 3 from the data of Table 4 • 2?

9. The text says that "poverty is a state of mind." What is meant by the statement? Is poverty a very real thing, and would one of the families listed in Table 4 • 1, with an income of under $500 per year, have a right to object to the suggestion that its poverty is "imaginary"?

10. Distinguish level of living and standard of living.

11. Are there any physical or physiological laws that underly and explain Engel's Laws? What are they?

5. UTILITY

5 · 1. Nature of Utility. The word "utility" is given various meanings in common speech, but to the economist *utility is the ability to satisfy human wants.* Some commodities satisfy wants because of their taste, some because of their smell, some because of their nutritive value, some because of their warmth, etc.; but, if a commodity has any characteristics whatever that enable it to satisfy human wants, we say that it has utility, and we call the commodity a "good." *Goods are those things which have utility*—which have characteristics that enable them to satisfy wants.

It should be clearly understood that, when we call a commodity a "good" and say that it has "utility," we are not implying at all that it is good for people or that it is socially desirable. There is no moral issue involved. The blackjack with which the criminal intends to fell his victim is a good and has utility, if it satisfies someone's want, and presumably it does satisfy the want of the criminal. Narcotics that are ruining the life and health of a drug addict are goods and have utility if the addict wants them. When we say that something has utility or that it is a good, we do not express either approval or disapproval; we merely state the fact that someone wants it.

5 · 2. Utility Is Always Subjective. If we comprehend clearly all the implications of this definition of utility, we shall see that utility is not an inherent attribute of goods but resides as much in the user as in the good itself. A given commodity may have no utility whatever today; yet tomorrow a change in style or a new scientific discovery may make it very useful indeed. Thus men once knew no use for rubber. It was a useless curiosity. Then they discovered that pencil marks could be rubbed out with it (which is, of course, where it got the name "rubber") and at once it became useful. There was no physical or chemical change in the rubber itself, but a change in the people. More recent discoveries have increased the utility of rubber time after time. But the ability to satisfy men's wants depends as much on men as on

the rubber. Utility, the ability to satisfy human wants, depends as much on the human beings as on the commodities that surround them. The utility of pig iron is not a fixed thing, like its specific gravity or its tensile strength or its hardness, but varies as people vary their wants and their understanding. Utility is subjective, residing as much in people as in commodities, and differing for different people. One man likes turnips while another man dislikes them. The latter's dislike does not mean that turnips have no utility. It is not necessary for everyone to like a commodity in order for it to have utility. As long as the commodity satisfies the wants of anyone, it is said to have utility and is classified as a good.

5·3. Kinds of Goods. We shall find it useful for our later study to classify goods into three classes, according to their relative scarcity. All goods, of course, have utility. But if a good is so plentiful that everyone has all that he wants, so that no user would find his satisfaction increased by any in-increase in the quantity of the good that he had, we say that the good is a "free" good. For example, air satisfies very important wants, and we would give up almost anything rather than get along without it; yet under ordinary circumstances men have enough air so that they would not benefit at all by an increased quantity. The chances are that you would not be willing to give up anything of value for an additional cubic foot of air. Under such circumstances air is a free good.

But most desirable things are scarce; there is not enough to go around. Although perhaps some favored people have all that they can possibly use, most people find that it would be to their advantage to have more. In such a case when goods are both useful and scarce, we call them "economic goods." As their name implies, these are the goods that give rise to our major economic problems. They are the goods on which our attention will be concentrated for the greater part of the time as we study economics.

Finally there are some things which satisfy men's wants but which are overabundant. They exist in such large quantities that it would be to our advantage to reduce the amount. Thus, although water satisfies men's wants, there are circumstances where there is too much of it, as in time of flood or when a farmer finds it necessary to drain a field. Any commodity that is useful but overabundant, so that men prefer to have the total amount reduced, we call "illth." This is a strange and new word, but it is evident that it is intended to contrast with the word "wealth," the words "well" and "ill" corresponding in some tenuous indirect way to wealth and illth.

We see, then, that all goods have utility and satisfy men's wants. If they are overabundant, we call them "illth"; if scarce, we call them "economic goods." If their quantity is just about right so that we get no advantage or disadvantage from small changes in their quantity, we call them "free goods."

5·4. Goods Classified by Use. We have just classified goods into three classes according to their relative scarcity: free goods, economic goods, and illth. But for some purposes we shall prefer a different classification, based on the use to which the goods are put.

Some things yield satisfaction to men directly, as does bread when it yields nutritive value and a pleasant taste, or as does coal when it gives off warmth for the dwelling. But many other things that men want yield no direct satisfaction. For example, one does not get satisfaction directly from an egg beater, but from the omelet that is made with it; nor from a waffle iron, but from the waffles that are made with it; nor from a textile factory or a loom in it, but from the cloth that is made on the loom in the factory.

Those things which yield a direct satisfaction are called "consumption goods"; those things which yield no direct satisfaction, but are useful in producing other economic goods, are called "production goods" or "capital goods" or "intermediate goods" or often just "capital." Thus an egg beater or a waffle iron or a screw driver or a railroad locomotive is capital. A ham sandwich or a living-room rug or a picture on the living-room wall is a consumption good.

Since this classification depends on the use to which goods are put and since different people often use the same good for different purposes, we can see that a commodity that one man correctly calls capital may correctly be called a consumption good by another man. One man plays his flute for pleasure while another plays in an orchestra for income. To the former the flute is a consumption good; to the latter it is capital. Fuel oil is a consumption good to the man who uses it to heat his home; but it is capital to the man who uses it to run his factory. The distinction between production goods and consumption goods is based on the use to which the goods are put. If they satisfy wants directly, they are consumption goods; if they are not used to satisfy wants themselves but to produce other want-satisfying goods, they are production goods.

5·5. Relationship between Amount of Good and Amount of Utility. Since goods satisfy men's wants, one would expect that more goods would yield more want satisfaction. In fact, one might imagine that the amount of utility would vary in direct proportion to the amount of the good, so that doubling its amount would double the utility or cutting its amount in half would cut the utility in half. Yet in practice the relationship is not quite so simple. Although it does seem to be true ordinarily that increases in the quantity of goods increase their utility, nevertheless the utility seems to increase more and more slowly as we increase the size of the stock of goods, as illustrated in Fig. 5·1. Each unit added to the stock seems to increase the

utility by a smaller amount, so that the curve rises rapidly at first, but more and more slowly as the quantity of the commodity gets larger.

5·6. Numerical Illustration of Utility. No one has ever been able to measure utility. We have no way of giving a definite value of even our own satisfactions, and far less of the satisfactions of others. I may know that I like French-fried potatoes better than mashed potatoes, but I am unable to say exactly how much I like either of them, or to state that I like French-fried potatoes just 4.62 times as much as I like mashed potatoes. We have no unit in which to measure satisfactions—no foot or pound or kilowatt-hour that is applicable. But the economist does know something about quantitative variations in utility. He can tell some things about such quantitative vari-

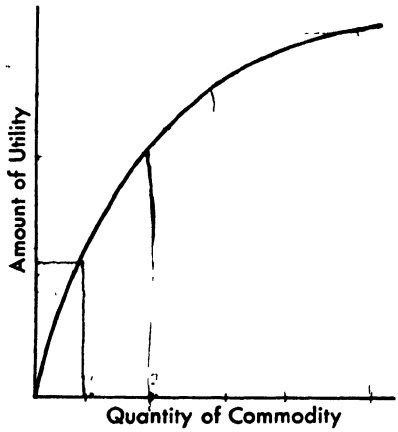


FIG. 5 • 1.—Changes in total utility accompanying changes in the amount of a commodity.

ations by watching the behavior of consumers, as we shall see shortly. And in order to illustrate the kind of thing that he knows does happen, he often makes use of numerical illustrations that are entirely hypothetical. In these fictitious problems he uses as his measure of utility the util, so that in his imagination *a util is a unit of utility.* If we say that one commodity gives us 5 utils of satisfaction and a second commodity gives us 15 utils, we are merely stating in numerical terms that we get three times the satisfaction from the second commodity that we get from the first. And, although such a fact could never be ascertained from actual measurement, we shall see that such pretended measurements may help us to understand the nature of actual situations.

Suppose, now, that we turn the chart of Fig. 5 • 1 into numerical terms. We might say that, if we vary the stock of this commodity which you have, the amount of satisfaction that you will get will vary as follows:

No. of units of the commodity	Amount of utility, utils	No. of units of the commodity	Amount of utility, utils
1	50	6	225
2	95	7	245
3	135	8	260
4	170	9	270
5	200	10	275

We note that the total amount of utility increases as we increase the amount of the good; but we see also that the increases in utility get smaller and smaller each time we add another unit of the commodity. Increasing our stock by one unit increases our total utility by 45 utils if we have but one unit of the goods to start with; but by only 10 utils if we have 8 units of the good to start with.

5·7. Marginal Utility. These changes in utility will be of a good deal of importance to us, and we shall be able to save time later on if we give them a special name so that we can refer to them easily; we shall call them "marginal utilities." We define them thus: *The marginal utility of any stock of goods is the amount by which the total utility would be diminished if we lost one unit of the stock.* Thus the marginal utility of 24 units of a commodity is the difference between the utility of 24 units and that of 23 units. The marginal utility of N units of a commodity is found by subtracting the utility of $(N\text{-}1)$ units from the utility of N units. Thus, if we wish to know the marginal utility of a stock of 7 units of the commodity in our numerical table, we note that 7 units yield 245 utils, while 6 units yield 225 utils. The difference, or 20 utils, is the marginal utility of 7 units.[1]

If we use this new definition of marginal utility, we can add a new column to the table, showing for each quantity of the good the corresponding marginal utility. This will give us Table 5·1.

Table 5·1. Illustration of Marginal Utility

No. of units of the commodity	Total utility, utils	Marginal utility, utils	No. of units of the commodity	Total utility, utils	Marginal utility, utils
1	50	50	6	225	25
2	95	45	7	245	20
3	135	40	8	260	15
4	170	35	9	270	10
5	200	30	10	275	5

[1] The student of the calculus should here refer to Note 2 of the Mathematical Appendix, on p. 885. The student who has not studied the calculus will find no advantage in such reference.

5·8. Law of Diminishing Utility. Although the figures of Table 5·1 are arbitrary and hypothetical, nevertheless they do illustrate a situation of a kind that does exist in practice. We know from observation that it is common for the marginal utility of any good to fall as the quantity of the good rises. In fact, it would be a very unusual case indeed were this not to be true. This tendency is called the "Law of Diminishing Utility" and can be stated as follows: *The marginal utility of any good depends on the size of the stock and varies inversely with the size of the stock.* If the student prefers simpler language, we may say that the larger the stock of any good, the smaller its marginal utility.

This law is so important that we should take time to illustrate it. It is important for the student to understand what the law means in practice, rather than to memorize the wording. Fortunately, illustration is easy, since the law works itself out all around us. It is a rather common thing for one to refuse to take his friend's last cigarette, even though he might take a whole package without compunction if his friend had a carton. He realizes intuitively that the utility of a cigarette depends on how many there are in the stock. The loss of a cigarette will be important if it is the only one; it will be negligible if there are hundreds of them. To take another illustration, under ordinary circumstances if one upsets a glass of water at the dinner table he is sorry to have made a mess, but he does not worry about the loss of the water. There is lots more. Yet for one who was lost on a desert the loss of the same half pint of water would be a serious affair. In the one case the loss of the water would bring but a minor reduction in utility; in the other case the reduction would be a major one; and, the student should note carefully, the difference is not a difference in the water. The loss is not small in the one case because the water is dirty or lukewarm or otherwise undesirable, and large in the other case because the water is fresh and clear and potable. The water is the same in both cases. The difference in marginal utility is the result solely of the difference in the total stock of water available. When we refer to diminishing utility, we are referring only to those differences which arise as a result of differences in the size of the stock. Moreover, we should emphasize that the law of diminishing utility is in reality a law of diminishing *marginal* utility. Additions to a stock of goods do ordinarily increase the total utility.

5·9. What the Law of Diminishing Utility Does Not Mean. Probably it is bad pedagogical practice to talk in terms of negatives—to spend time and effort telling what is *not* true. But it is so easy to fall into mistaken impressions about the Law of Diminishing Utility that it may be profitable to emphasize some of the things that the law does not say. (1) As we have already noted, this law does not treat of differences that arise from differences in quality of the goods; they are all assumed to be alike. (2) The law

does not refer to changes in total utility, but to changes in marginal utility. (3) The law does not tell what happens with the passage of time. It does not say that marginal utility falls or rises as time goes by. It says nothing about time at all. This law tells what is true at a given instant of time, and failure to note this leads many students into error. They say, for example, that a dope fiend gets to like his dope more and more as he uses more of it; that he experiences increasing rather than diminishing utility. As a time phenomenon, this may be true. But the Law of Diminishing Utility says that *at any given time,* the man's tastes being what they then are, he will lose less utility with the loss of a unit of his dope if he has a big stock on hand than if he has a small amount. Neither does the law pretend to compare the satisfactions of different consumers. It does not say that one person gets less satisfaction from a good than another. The law is individualistic and personal. As far as we know, it describes the facts for any individual; but it does not compare him with other individuals.

(4) Finally, marginal utility is not the utility of any particular unit of stock, as contrasted with the utility of other units. Students sometimes fall into the error of saying that each unit added to a stock adds less utility than the one before it, and that the marginal utility of a stock of goods is the utility of the last unit added. Or they may say that the marginal utility is the utility that would be lost if we gave up the last unit that was added. Now marginal utility has nothing whatever to do with the "last" unit or with any other particular unit. If we have a stock of 7 units of the good in question, the marginal utility is the amount of satisfaction that we would lose if we lost *any one* of them. Usually we would not even know which was the last one. The units are all alike and, if we have 10 cigarettes, we do not care which one is lost. If any one of the 10 is lost, we shall have 9 left, and the marginal utility is the difference between the utility of the 10 and the utility of any 9. The marginal utility is not the utility of a particular unit; it is a difference between two total utilities.

5·10. Evidence Substantiating the Law of Diminishing Utility. We have already given several examples of the operation of the Law of Diminishing Utility. We can each of us recognize its operation by introspection. We know, when we reflect on the matter, that the loss of a unit of a good is of less importance to us when it is one of a large stock than when it is one of a small stock. But we can also get evidence by noticing the behavior of other people. To be sure, they cannot measure the changes in their own marginal utility, nor can we measure it for them. But we can see, for example, that a man buys a pair of shoes for $8. Perhaps he buys a second pair. But before long he has enough and stops buying. Additional pairs of shoes are not worth $8 to him. The additional, unbought shoes are just like the ones he already has. There is no difference in the shoes. Yet the first two pairs

were worth $8 and the third pair was not. When we see a man buy one or more units of a good and then stop, leaving unpurchased other units that he could buy (we assume that his stock of money is not completely exhausted when he buys the second pair), we conclude that the addition of a third pair would increase his satisfaction less than the addition of the second pair. The behavior of consumers everywhere bears out the Law of Diminishing Utility.

5·11. Diminishing Utility of Money. As far as we know, all things that men desire are subject to diminishing marginal utility. This seems to be true of money as of everything else. The loss of a ten-dollar bill may well be a matter of indifference to a multimillionaire; yet the ordinary college student would find that such a loss made a material difference to him. When one's resources are low, he may decide not to buy a milkshake for himself; when he has just cashed a check from home, he may treat all his friends. This is not because he has increased his desire for milkshakes, but because he now finds that he can part with money more easily than before. His stock of money is larger, and its marginal utility is smaller.

In Sec. 4·5 we made the statement that most economists agree that it would be socially advantageous to distribute incomes somewhat more equally than we do. We can now see one of the reasons for the statement. Suppose we were to take $500 from the income of a family with an income of a million dollars a year. The chances are that they would never realize the loss; it would make little or no difference to them. If, however, we were to give this $500 to a family with an income of $1000 a year, it would make a great difference to them. They would receive a great increase in satisfaction, and their manner of living would be transformed. Thus the transfer of $500 could be made with almost no loss to the one family and with a great gain to the other. There would be a net gain to society under such circumstances. To be sure, we must not assume that all families get exactly the same satisfaction from income, and thus we cannot be sure that complete equality of incomes would give us the greatest possible total satisfaction. We have already emphasized the fact that diminishing marginal utility is a personal thing and not a comparison between persons. Here we have been comparing two different families and saying in effect that the family with a large income will have a smaller marginal utility of money than the family with a low income. Such comparisons are, in general, dangerous. We believe that it is true that, when a given family has a large income, the marginal utility of money will be smaller than when the same family has a small income, but we are not justified in making similar statements about different families, one with larger income than the other, except in such extreme cases as the one that we have taken here for purposes of illustration. Presumably every-

one will agree in such an extreme case as this one that a transfer of $500 would bring a net increase in satisfaction.

Of course, before we decide whether such transfers would be desirable or not, we must remember that income performs other functions than that of bringing in want satisfaction. It is also to some extent an incentive to production. It is possible that one might oppose a transfer of income or wealth, even though he agreed that it would increase total satisfaction, because he thought that the transfer would have an undesirable effect on production. If we take too much income from the industrious and give it to the shiftless in order to equalize their incomes and increase satisfaction, we may reduce the incentive for the industrious people to work, thus reducing the total production and the satisfaction where we intended to increase it.[1] Hence we have not intended to assert here that complete equality of income would be desirable, but merely to reiterate and explain further our earlier point, namely, that some approach to greater equality of income than we now have would surely be socially desirable.

5·12. Negative Marginal Utility. We have said that marginal utility of any stock of goods continues to grow smaller as the stock of goods grows larger. We have illustrated this numerically with Table 5·1. This table stopped with a stock of 10 units of the good. Suppose, now, that we were to continue the table to larger stocks. What would be the total utility of 11 units? Although we cannot be exact, we can tell something about it. We are not entirely in the dark. We see that the marginal utility of 10 units was 5 utils, and the Law of Diminishing Utility tells us that the marginal utility of 11 units will be less than 5 utils—possibly 4 utils, possibly 2.5 utils, or possibly some other figure smaller than 5. If, say, the marginal utility of 11 units is 2 utils, then the total utility must be 277 utils; if the marginal utility of 12 units is 1 util, then the total utility of the 12 units is 278 utils. Thus we can continue the table if we know the marginal utilities. Given the total utilities, we find the marginal utilities by subtraction; given the marginal utilities, we find the total utilities by addition.

Let us now extend Table 5·1 by adding a few more entries at the bottom to get Table 5·2. We see that the marginal utility has continued to fall, in conformity with the Law of Diminishing Utility. It has dropped from 5 utils to 2, from 2 utils to 1, and from 1 util to 0. A stock of 13 units gave no more satisfaction than a stock of 12. If we have a stock of 13 units and lose one, we are just as well off as before. Under such circumstances the commodity has a marginal utility of zero. If, now, we keep on adding to the stock of goods, we finally get so many that they become troublesome—they

[1] This sentence seems to imply that the industrious are always people of high income and that the shiftless are always poor. Such an implication is, of course, unwarranted.

get in the way and are a nuisance. Now additions to the stock reduce our total utility, so that the marginal utility has become negative. Such things, too, happen in practice. We buy one radio and find it useful. We buy a second and find that we have some use for that, too. Perhaps three radios are even more useful than two; but surely there is a limit. What can we do with 25 radios, or 250?[1] They get in the way, block up the hallways, and gather dust under the bed. We could easily get so many radios that they would become a nuisance. We would pay, if necessary, to get rid of some of them. They would have negative marginal utility.

Table 5·2. Illustration of Marginal Utility

No. of units of the commodity	Total utility, utils	Marginal utility, utils	No. of units of the commodity	Total utility, utils	Marginal utility, utils
1	50	50	9	270	10
2	95	45	10	275	5
3	135	40	11	277	2
4	170	35	12	278	1
5	200	30	13	278	0
6	225	25	14	276	−2
7	245	20	15	270	−6
8	260	15			

It is now possible for us to redefine the terms "economic goods," "free goods," and "illth" in terms of the concept of marginal utility. If we keep the same meaning that we gave to these terms in Sec. 5·3, we now see that

1. *Economic goods are those things which have positive marginal utility.*
2. *Free goods are those things whose marginal utility is zero.*
3. *Illth consists of those things which have negative marginal utility.*

Whether a given commodity is a free good, an economic good, or illth depends, then, on the quantity of it. If the stock is so small that people would like to have more, it is an economic good. If the quantity available is just about right so that there is no particular advantage in having more nor any real loss in having less, the commodity is a free good. If the commodity does satisfy wants but the stock is too large, so that we would gladly reduce the stock, then the commodity is illth.

5·13. Law of Satiety. The Law of the Variety of Human Wants told us (see page 7) that there is no limit to the number of different commodities

[1] To be sure, we could sell the spare ones that we did not want, but we are asking how much more satisfaction we would personally get from 25 radios than from 24. We are not talking here about goods for sale, but about the use of goods in want satisfaction. And that is just the point—when you have 24 radios you have *no more use* for radios.

that men want. Now we see that the want for any one commodity can be satisfied if we increase the quantity sufficiently. It would be possible to produce so much wheat that no one wanted any more. It would be even easier to produce so many meerschaum pipes or so many piccolos that no one wanted any more. Although we do not, under ordinary circumstances, increase the quantity of economic goods until they are free goods, it would be possible to do so if we wished. The drawback would be that, if we were to produce enough wheat so that no one wanted any more, we would have to put into growing wheat, the manpower and land and equipment that could better be used for producing other things. When we increase the quantity of a good, we always decrease its marginal utility, but in practice we do not usually reduce it to zero. It is well to remember, however, in connection with the Law of Variety, which we studied in Chap. 1, a companion principle, which is sometimes called the "Law of Satiety": *The desire for any good can be completely satisfied by some finite quantity of that good.*

5·14. Summary. We should remind ourselves, then, that as we increase the stock of economic goods we increase their total utility but diminish their marginal utility. We cannot actually measure the changes, but we can sense them in ourselves and see that other people act as they would act if they, too, were subject to what we call the Law of Diminishing Utility. To illustrate the kind of thing that happens, we pretend that we can measure satisfactions in an imaginary unit called the "util," and with this unit we show the way in which total utility and marginal utility do vary. The units and the examples are imaginary; the variations in utility are real.

Questions

1. Make a list of ten economic goods. How many free goods can you list? List some things that are illth.
2. "Utility is always subjective." Explain.
3. We define free goods in terms of want satisfaction and scarcity. Are free goods also free in the sense of being costless?
4. Give illustrations, in addition to those of the text, showing that a given commodity may sometimes be capital and sometimes a consumption good.
5. What is the advantage of using such an imaginary thing as the util, since in practice we cannot really measure the satisfaction of wants at all?
6. The table at the top of the next page shows on the lower line the amount of total utility that a man gets from each of the numbers of doughnuts listed on the line above. What is the marginal utility of 3 doughnuts? Of 5 doughnuts? What would be the marginal utility of 6 doughnuts?

Number of doughnuts	1	2	3	4	5
Total utility, utils	30	50	60	67	71

7. The accompanying table shows on the lower line the amount of marginal utility that a man gets from each of the numbers of apples listed on the line above. What is the total utility of 5 apples? Of 3 apples? What is the most accurate thing to be said about the total utility of 6 apples? Does the table illustrate the law of diminishing utility?

Number of apples	1	2	3	4	5
Marginal utility, utils	18	13	9	7	3

8. In addition to those in the text, give illustrations of the operation of the Law of Diminishing Utility in practice.

9. A student, asked to give an example of diminishing marginal utility, says that when a man eats more and more strawberry shortcake he becomes less and less hungry, until finally his appetite is completely satisfied, and he might even eat so much that shortcake would become distasteful to him. This student has the right general idea, but his illustration is not quite perfect. What is the fault with it?

10. In discussing the advantage of transferring funds from a wealthy to a poor family in Sec. 5 • 11, the text apologizes for speaking of diminishing utility in comparing two families. What is the difficulty involved?

11. If we bar the possibility of disposing of goods to other people, do you think that most economic goods would finally become illth as their quantity increased? How many automobiles do you think you could own before they became illth under these circumstances? How many loaves of bread? How many copies of the morning newspaper?

6. DEMAND

6·1. Results of Diminishing Utility. In Chap. 5 we became acquainted with the idea of diminishing marginal utility. Suppose, now, we see what some of the results are of living in a world where men are subject to diminishing marginal utility.

First, we shall go back to the data that we studied in Table 5·2, repeating the same data now as Table 6·1 for ease of reference. Suppose we ask, how many units of the commodity will be purchased? Will the man whose desire

Table 6·1. Illustration of Marginal Utility

No. of units of the commodity	Total utility, utils	Marginal utility, utils	No. of units of the commodity	Total utility, utils	Marginal utility, utils
1	50	50	9	270	10
2	95	45	10	275	5
3	135	40	11	277	2
4	170	35	12	278	1
5	200	30	13	278	0
6	225	25	14	276	−2
7	245	20	15	270	−6
8	260	15			

for the commodity is depicted in the table buy 1 unit or none, or will he buy all 15? The answer is that we do not have enough information to say. In order to tell how many units of a commodity a man will buy, we have to know, not only how much he likes the commodity, but also how much money he has and how much he likes money. Is he a wastrel and a spendthrift, or is he close and stingy and penny-pinching? These things will also influence his decision to buy. Table 6·1 shows us how much he likes the

commodity—the utility of the commodity to him. We also need to know the utility of money to him.[1]

Suppose, then, that we say that this man sets such a store on his money that each dollar has 7 utils. We ask again, "How many units of the commodity will he buy?" Again we find that we do not have enough information, for we know that a man's purchases are determined, not only by his desire for the good and by his evaluation of money, but by the price of the good. So we have to ask in this case how much the good costs. Let us assume that it costs \$1 per unit—that its price [2] is \$1. Now we are in a position to say definitely that this man will buy 9 units of the commodity. Let us summarize our given facts and then see how we reach this conclusion.

Given:

1. The desire for the good, shown in Table 6 • 1.
2. The desire for money: Each dollar has 7 utils.
3. The price of the good: \$1 per unit.

Conclusion: 9 units will be purchased.

In reaching this conclusion we start with the realization that each time the man buys a unit of the commodity he has to give up \$1 (that is, the price), and these dollars are worth to him 7 utils. It costs him 7 utils to get a unit of the good. If he has 5 units of the good and contemplates buying another, he sees that his total utility will rise from 200 to 225 utils. He will gain 25 utils by spending 7. It will pay him to buy this sixth unit. If he has 8 units and is considering the purchase of a ninth, he sees that he can now gain 10 utils by spending 7. But if he has 9 units and thinks of buying a tenth, he will gain but 5 utils by spending 7. He would be foolish to buy a tenth unit. This is how we reach the conclusion that he will buy nine.

Let us try this system again, changing our suppositions slightly. Let us now take these things as given.

Given:

1. The data of Table 6 • 1.
2. Each dollar has 12 utils.
3. The price of the commodity is \$3.

Conclusion: He will now buy only 3 units.

[1] Of course, one may say that money has no utility and that it does not satisfy any wants. To be sure, it is a production good rather than a consumption good (see Sec. 5 • 4). We want money, not for itself, but for what we can get with it; just as we want a steam shovel, not for itself, but for what we can do with it. The utility of money is really merely a reflection of the utility of the goods that we can buy with it. When we speak of the utility of money, here or elsewhere, it must be understood that we refer really to the utility of the goods that it will purchase.

[2] Hereafter we shall follow the practice, common with economists, of using the term "price" to mean "price per unit." If, for example, 5 apples cost 25 cents, the price of apples is 5 cents. In any problems that follow, if we use the term "price," it is to be understood that it is on a "per unit" basis.

Again our reasoning follows the same pattern. To buy a unit of the commodity, our man must give up \$3 which are worth 12 utils each. Thus each unit of the commodity will cost him 36 utils. If he buys a third unit when he already has two, he will get 40 additional utils by giving up 36. This will be worth while. But if he now thinks of buying a fourth unit, he finds that it will bring in but 35 more utils and will cost him 36. This will not be advantageous.

6·2. Purchase Depends on a Comparison of Utilities. We see that the decision to buy goods depends partly on the utility of the goods and partly on the utility of the money. It involves a comparison of the utilities. To be sure, the comparison is not so formal as we have made it appear here. The

Table 6·2. Problem in Diminishing Utility

Facts about the commodity			Facts about money		
No. of units on hand	Total utility, utils	Marginal utility, utils	No. of dollars on hand	Total utility, utils	Marginal utility, utils
1	78	78	1	38	38
2	141	63	2	71	33
3	192	51	3	97	26
4	235	43	4	120	23
5	271	36	5	137	17
6	298	27	6	147	10
7	315	17	7	153	6
8	323	8	8	156	3
9	326	3	9	157	1

housewife does not go to the store, look at the price tags, and then carry out an arithmetical computation in terms of utils. Although the process as we have described it is overformalized, nevertheless something of the kind does actually occur. People do not buy goods haphazardly. Their purchases are by no means entirely nonrational. Insofar as they are able to do so, they buy goods only when they think they are getting their money's worth. If they think the goods are worth more than the money, they buy; if they think that the money is worth more than the goods, they refrain.

In our illustrative numerical examples we have so far taken the utility of money as fixed. We have said, for example, "Let us assume that each dollar has 12 utils." But money, like anything else, has diminishing utility.[1]

[1] Again we speak by ellipsis. The utility is in the goods purchased with the money rather than in the money. See the footnote on p. 70. If we have a large amount of money, some of the goods purchased will be goods of low utility. Some dollars will go for things that bring us little satisfaction. If our stock of money is small, every dollar of it must go for important things—for things of high utility. It is in this sense that we speak here, as we did in Sec. 5·11, of the diminishing utility of money.

Probably in most cases our purchases are so small as to make little difference in the marginal utility of money. If we buy a newspaper for 5 cents, there has been such a small change relatively in our stock of money that we are not far wrong if we consider the marginal utility of money as unchanged. It is in this sense that we have said heretofore, "Assume that money has 12 utils of the dollar," or something of the kind. But our problems work out in the same general way even if the marginal utility of money varies with the quantity of money in our possession. Let us illustrate with the data of Table 6·2.

This table shows that both the commodity and the money are subject to diminishing marginal utility. Let us suppose, now, that we are given the following:

Given:

1. The utility of the good from Table 6·2.
2. The utility of money from Table 6·2.
3. The price of the good is $1.
4. The man has $8 and none of the commodity.

Conclusion: He will buy 6 units of the good.

Our reasoning runs as follows: If he buys 1 unit of the good, he gets 78 utils that he did not have before; but he spends $1, reducing his monetary stock from $8 to $7, and thus sacrificing 3 utils. He gets 78 utils in return for 3, which is advantageous. If he buys a second unit of the commodity, his utility is increased from 78 to 141, a gain of 63 utils. But he now has cut his monetary stock from $7 to $6, reducing the utility of his money from 153 to 147 utils. He loses 6 utils to get 63. Similarly, we could show that the addition of a third unit would bring in 51 more utils at a sacrifice of 10 utils; the purchase of a fourth unit would bring in 43 with a loss of 17; the fifth unit would add 36 and subtract 23; and the sixth unit would add 27 and subtract 26. It is worth while to buy each of these 6 units. But the seventh unit would bring in only 17 more utils, and by this time the monetary stock is so low that each dollar is very valuable—is worth 33 utils, in fact. It is not worth while to give 33 utils to get 17, and the seventh unit of the commodity is not purchased. This is the basis for our conclusion that the man will buy 6 units of the commodity.

While we have the example of Table 6·2 before us, let us ask how many units would be bought if the price of the commodity were $2 instead of $1. We need make only one minor adjustment to our method. Our problem now is

Given:

1. The data of Table 6·2.
2. Price of the commodity is $2.
3. The man has $8 and none of the commodity.

Conclusion: He will buy 3 units of the commodity.

This time, if he buys a unit of the commodity, he will get 78 utils which he did not have before; but he will spend \$2, reducing his monetary stock from \$8 to \$6 and the utility of his money from 156 to 147 utils. He will thus have to sacrifice 9 utils to get 78. In buying a second unit he will get 63 additional utils but will reduce his monetary stock from \$6 to \$4, losing 27 utils. He gets 63 utils by giving up 27. When he buys the third unit, he gives up 49 utils to get 51, which is still to his advantage. But, if he bought a fourth unit, he would spend his last two dollars, giving up the 71 utils which they represent, and this fourth unit of the good would bring in only 43 utils. Therefore he does not buy the fourth and stops with the purchase of 3 units.

6 · 3. The Economic Man. Our illustrations have made our buyer appear to be a cool, calculating mathematician, who appraises every purchase carefully and accurately in terms of the gain and loss involved. Such a human calculating machine is sometimes referred to as "the economic man," and the economist is derided for pretending that in real life people act that way. Let us hasten again to say, then, that our examples have been overformal and overmathematical. Yet, let us again insist that they do show the kind of thing that does happen very approximately in practice. If it is unreal to pretend that men go through such a process of calculation, it is even more unreal to pretend that their purchases are entirely unplanned and random. When the housewife walks bravely past the store window that displays that chic \$5 hat, it is not because she does not want the hat nor because she does not have \$5; it is because she knows that she wants, even more, to spend the \$5 on something else. When the college student foregoes the purchase of that snappy \$2.50 necktie, it is not because he dislikes the necktie nor possibly even that he does not have the \$2.50. He merely realizes that, if he spends the money on the necktie, he cannot afford to take in next Saturday's dance, and of the two he prefers the latter. Again we point out that purchase always involves comparison of utilities: the utility of the goods and the utility of the money, or, more accurately, the utility of the goods and the utility of other things that the money would buy. If the goods are worth more than the money, he buys. If they are worth less than the money, he does not buy. To be sure, men are not always good judges and often make the mistake of buying things that later turn out to give little or no satisfaction. They find that they did not get their money's worth. But, insofar as they are able, they try to avoid such a situation. Although the men of real life are perhaps not the util counters we have depicted, nevertheless the illustrations do show the kinds of forces that are at work in real life.

What happens is that, if the marginal utility of the good is high enough and that of money is low enough, the man starts buying. As he continues to buy, the marginal utility of the good falls and that of money rises. As long as the marginal utility of the good exceeds that of money, he continues

to buy; but when the marginal utility of the good falls as low as that of money, he stops. The buyer tends to make the marginal utility of the good equal that of money. If his purchases were infinitesimal in size, perhaps he could make them exactly equal. As it is, he makes them approximately equal, getting them as nearly so as he can without reducing the marginal utility of the good below that of money. What it all comes down to is that *he tries to get his money's worth.*

6·4. The Decision to Buy Is Marginal. If we turn back to our last example, we find (see page 72.)

Given:

1. The data of Table 6·2.
2. Price of commodity is $2.
3. The man has $8 and none of the commodity.

Conclusion: He will buy 3 units of the commodity.

We might now reexamine the case thus, referring again to the data of Table 6·2. Why does the man not buy 4 units of the commodity? If he bought 4 units, they would give him a total of 235 utils of satisfaction. To get them he would give up his $8 and so would lose a total of 156 utils of satisfaction. Is it not wise for him to give up 156 utils to get 235? Why does he stop with 3 units?

The answer is that the man *would* buy 4 units of the commodity *if he had to.* If he had to take 4 units or none, he would take all 4, since he would thus get more satisfaction than he gave up, but in real life one is seldom faced with the alternative of taking all or none. Usually a price is set on the commodity, and we take as little or as much as we like at that price. In our case we have assumed that the price is $2 per unit, and he can take as many units as he can pay for. Since he has $8 of resources, he can buy anywhere from no units to 4, and he has freedom of choice in the matter within those limits. Now under these circumstances the important thing to understand is that he will tend to make his decision on the basis of the marginal utility rather than on the basis of the total utility. As one of America's leading economists has put it in an effort to reduce the matter to everyday terms, people make their economic decisions on the basis of "the difference it makes." We often hear the expression, "What difference does it make?" and we should remember that marginal utility is a difference—the difference in total utility.

As we say, if the man is given a choice of 4 units or none, he takes 4 units at $2 each; if he is given freedom to take as many or as few as he wants at $2 each, he takes three. Why? Because the *difference* between 3 and 4 units is only 43 utils, and the *difference* between having $2 left and having no money at all left is 71 utils. Having 3 units of the good, a fourth is not worth the money and, if he can take 3 units without 4, he is better off.

6·5. Consumer's Surplus. Perhaps we can see this in another way. Suppose that our man does buy 3 units at $2 each, as we have suggested. He can then reckon up his accounts thus,

Satisfaction gained from 3 units of the commodity . . .	192 utils
Satisfaction lost by expenditure of money	85 utils*
Net gain from transaction	107 utils

* To buy 3 units when he had $8, this man spent $6 and had $2 left. The $8 gave him 156 utils of satisfaction, the $2 give him 71 utils of satisfaction. He has given up 85 utils, as above. We would get the same answer if we added up the marginal utilities of dollars number 8, 7, 6, 5, 4, and 3, which are the $6 he spent These marginal utilities are 3 + 6 + 10 + 17 + 23 + 26 = 85, the amount of satisfaction given up.

This difference we call the "consumer's surplus," which we define as follows: *The consumer's surplus on any transaction is the difference between the total utility received and the total utility expended.* It is the difference between what he gives and what he gets on the transaction. Now let us repeat our former question: "Why does he not buy 4 units, since they would cost him $8 with utility of 156 utils and he would get 4 units of the commodity with utility of 235 utils?" Let us compute the consumer's surplus again:

Satisfaction gained from 4 units of the commodity	235 utils
Satisfaction lost by spending $8	156 utils
Consumer's surplus	79 utils

He would be better off to buy 4 than to buy none. He would have a consumer's surplus of 79 utils. But he gets an even greater consumer's surplus, namely 107 utils, by stopping when he has bought 3 units. He is better off to buy 4 units than to buy none; but he is even better off to buy 3 units than to buy 4. We have said before that decisions to buy are made on the basis of the comparison of utilities. Now let us stress again that *decisions to buy are made on the basis of a comparison of marginal utilities in those cases where the buyer is left free to buy any quantity of the commodity that he wishes.*

6·6. Effect of Changes in the Price. In Sec. 6·2 we took certain things as given and found how many units of the commodity the man would buy. We found these results:

Given:

1. The data of Table 6·2.
2. The man has $8 and none of the commodity.
3. Price of the commodity is $1.

Conclusion: He buys 6 units of the good.

Given:

1. Same as above.
2. Same as above.
3. Price of the commodity is $2.

Conclusion: He buys 3 units of the good.

If, now, we leave the first two given factors the same as before, but change the price of the commodity to \$3, we discover that the man will buy only 1 unit. If we get the price up to \$6, he buys none at all. We can summarize these results in Table 6·3.

Table 6·3. Effect of Changing a Commodity's Price Based on Table 6·2

Price per unit, dollars	Number of units purchased	Price per unit, dollars	Number of units purchased
1	6	4	1
2	3	5	1
3	1	6	0

Let us now go back to Table 6·1 and try the same sort of thing. We discovered there that if each dollar had 7 utils, the man would buy 9 units at a price of \$1. Using the same data and the same value of money but varying the price, we find that he will buy 8 units when the price is \$2; he will buy 6 units when the price is \$3; he will buy 5 units when the price is \$4, etc. We summarize in Table 6·4.

Table 6·4. Effect of Changing Price in Table 6·1

Price per unit, dollars	Number of units purchased	Price per unit, dollars	Number of units purchased
1	9	5	3 or 4*
2	8	6	2
3	6	7	1
4	5	8	0

* When the price is \$5, we cannot be absolutely certain how many units will be purchased. The third unit brings in 40 utils and, at a price of \$5, costs 35 utils. Certainly the man will buy the third But the fourth unit brings in 35 utils and costs 35. It is a matter of indifference to the buyer whether he buys unit 4 or not. We know he will buy 3, and he may buy 4. Therefore we give our answer as "3 or 4." If the student will compute the consumer's surpluses, he will find that they are exactly the same on 3 or 4 units—30 utils of surplus in either case

6·7. Individual Demand. We call a table like Table 6·3 or Table 6·4 the "individual demand schedule" for the commodity. The table shows the quantities that would be purchased at various prices when the individual's tastes for this good and for other goods including money, are given. We have kept his taste for the goods fixed, because we have based all our computations on the same utility table. For example, all the computations of the demand schedule of Table 6·3 are based on Table 6·2, in which the buyer's attitude toward the commodity stays always the same, as given in the first three

columns of Table 6·2. The last three columns of Table 6·2 show the man's attitude toward money or, as we have pointed out before, the reflection of his attitude toward other goods. As long as Table 6·2 stays the same, we keep the buyer's taste for the good and his attitude toward money and toward other goods constant. We then change the price, and what do we find? We find that *the quantity of the commodity which a buyer stands ready to purchase varies inversely with the price of the commodity.* This is called the "Law of Demand." We may restate it in simpler terms: The higher the price of a commodity, the smaller the quantity that a buyer will be willing to buy. Either of these statements must be qualified with the usual *ceteris paribus* assumption—that the buyer's taste for the good and for other goods stays the same, and that prices of other goods and the buyer's attitude toward money stays the same. These things were in given in our example. Naturally we might raise the price of a commodity and have a buyer willing to buy even more at the high price than he had bought before, *if* at the same time he has changed his attitude toward the goods or toward money—if he has come to like the goods better or if the marginal utility of money has declined.

Let us also note that the Law of Demand depends on the Law of Diminishing Utility. Our original tables, Tables 6·1 and 6·2, were drawn up in accordance with the Law of Diminishing Utility. They exhibit diminishing marginal utility. If marginal utility increased as one's stock of goods increased, he might go on buying more and more without limit until his money was gone. But as it is, there is a combination of two tendencies that puts a halt to his purchases: as he buys more, the marginal utility of the good falls; as he buys more, the marginal utility of the money rises. When the two become equal, he stops buying.

And why does the Law of Demand work? Why is it that at higher prices he buys smaller quantities? It is because at higher prices he sooner reaches the point where the marginal utility of the goods equals the utility of the money that he has to give up. Suppose he always buys until the marginal utility of the goods has fallen to the point where it equals the utility of the money given up. If the price is low, he gives up little money for each purchase, and the small amount of money has little utility. It takes a long time to bring the marginal utility of the good so low. But if the price of the good is high, he gives up more money for each purchase, and this larger amount of money has more utility. It does not take so long to bring the marginal utility of the good down to equal this larger utility of money. He buys fewer units of the good at the high than at the low price.

We could, perhaps, have started with the Law of Demand, skipping our work on utility. We could have said, as a matter of observation, that at

high prices people buy less than they do at low prices. But our approach has enabled us to see several things:

1. The Law of Demand depends on the utility of the goods and, if our taste for the goods changes, our demand for them will change.

2. Our demand for the goods will also change if we change our attitude toward other goods or toward money.

3. Whenever there is diminishing utility (and as far as we know it is common to all goods), the demand must be negative; at higher prices, the quantity taken will be smaller.

In other words, we not only see empirically that the Law of Demand is true, but we see why it is true and what things need to be changed to bring about changes in demand.

6·8. Market Demand. We have seen that each individual tends to reduce his purchases of any commodity when the price rises. If we look back at Table 6·3, we see that some price increases bring reductions in purchases while others do not. Thus, when the price rose from $2 to $3, purchases dropped from 3 units to 1; when the price went up again to $4, the purchases stayed at 1 unit. It was not, in this example, until the price rose to $6, that the amount purchased was again reduced.

This, too, is the kind of thing that actually happens in practice. If the price of sugar rises from 9 to 10 cents a pound, the housewife may continue to buy just as much as before. Similarly, when the price goes up to 11 cents, she may make no reduction in her purchases. But sooner or later, as the price goes higher and higher, she will start to economize on sugar and reduce her purchases. Our individual demand schedule shows the kind of thing that happens.

Of course, different consumers have different tastes for sugar, and they have different amounts of money and different tastes for other goods. Table 6·3 shows what happens for some particular consumer as the price changes, and this table is, in turn, based on Table 6·2, showing the utility facts for the same consumer. Another consumer would have a different utility table (similar to Table 6·2, but with somewhat different figures), and consequently a different individual demand (similar to Table 6·3, but with somewhat different figures.) The individual demands of different consumers will vary, but they will all be alike in one respect: as the price rises, the consumer will buy either the same amount as before or a smaller amount. Rising price will never encourage him to increase his purchases. Thus, when the price rises from $3 to $4, there is no change in purchases in Table 6·3. This particular consumer made no adjustment to the price change. But if we consider all the consumers on the market, some of them will be reducing their purchases and some will be retaining the same quantity of purchase

when the price rises from $3 to $4. Every price rise will be expected to reduce someone's consumption if there are many buyers on the market, and no buyer will be expected to increase his consumption as a result of the rising price. Thus, if we consider the market as a whole, we may anticipate what does happen in fact: that rising prices will reduce the quantity purchased. The Law of Demand is valid for the market as a whole just as it is for the individual. We can make up a market demand schedule similar to our individual demand schedule. Where the latter shows the quantities that some individual is willing to buy at various prices, the former will show the quantities that will be bought by all consumers on the market at various prices. If, at a price of $3, A buys 3 units and B buys 5 units, etc., we can add all the individual purchases to get the quantity purchased on the market as a whole. The market demand schedule is the sum of many individual demand

Table 6·5. Demand for a Commodity in the Town of Gurleyville

Price per unit, cents	Quantity buyers will take	Price per unit, cents	Quantity buyers will take
30	382	70	211
35	343	75	201
40	313	80	192
45	288	85	185
50	267	90	177
55	250	95	170
60	235	100	164
65	222		

schedules. We can illustrate the market demand schedule with data such as those of Table 6 · 5. Here we see that every increase in price brings some decrease in purchases.

6 · 9. Demand Curves. We can show the demand situation on the Gurleyville market either by means of a market demand schedule, such as that of Table 6 · 5, or graphically, as in Fig. 6 · 1. If we transfer the data of Table 6 · 5 to ordinary coordinate axes, showing prices on the vertical and quantities on the horizontal scale,[1] we get those points of Fig. 6 · 1 which are

[1] There is a dispute among economists as to which axis should be used for price and which for quantity. The matter may seem to be one of indifference, but there are conventions in accordance with which such things should be decided. Unfortunately, the early economists did not follow the usual conventions of mathematics and put price, the independent variable, on the axis of ordinates rather than on the axis of abscissas. We follow here the usual custom of economists, despite the fact that it has some handicaps. In showing demand curves and later in showing supply curves, we shall put price on the vertical and quantity on the horizontal axis.

shown within the little circles. There are 15 such points, one to correspond with each of the 15 prices of Table 6 • 5. We have connected these 15 points with a smooth line, to indicate the probable quantities that would be taken at intermediate prices. Thus Table 6 • 5 tells us that consumers in Gurleyville will buy 343 units when the price is 35 cents, or 313 units when the price is 40 cents. The table tells us nothing about the amounts that would be purchased at intermediate prices, such as 37 or 38 cents. In drawing the curve of Fig. 6 • 1 we assume that the quantities sold at 37 or 38 cents will lie along the same smooth curve that passes through the points in the circles. Thus we get what the economist calls a "demand curve," which is one

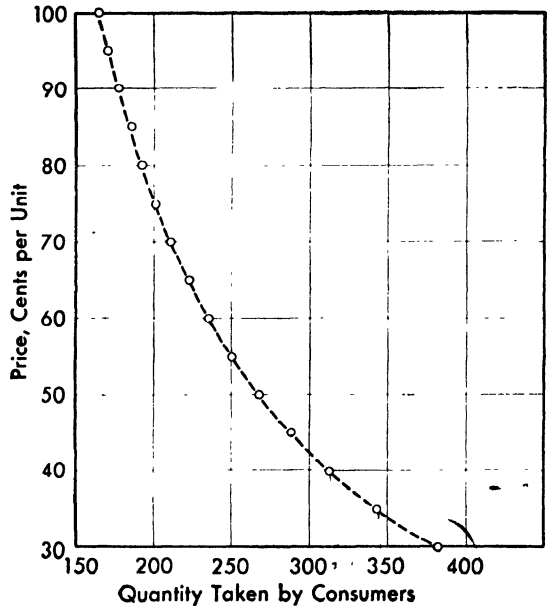


FIG. 6 • 1.—Demand for a commodity in Gurleyville. Based on demand schedule of Table 6 • 5.

showing the quantities that consumers would be willing to buy at all possible prices. The demand schedule is a table showing quantities that would be taken at certain arbitrary prices. The demand curve shows graphically the quantities that would be taken at all prices—the same general kind of thing that is shown by the demand schedule, but in more detail and with greater completeness. An individual demand curve would show graphically the amounts taken by some individual at various prices. A market demand curve shows graphically the amounts taken by all buyers on the market at various prices. Each market demand curve is dependent on the individual demand curves of the people represented in the market, just as a market demand schedule is the sum of individual demand schedules.

6 • 10. Demand. The term "demand" is used in popular speech to mean many different things. Some people use it as roughly synonymous with the word "desire." Some mean by it the quantity of a good that people buy on the market. Some mean the general willingness to buy the good at various prices.

The economist differentiates demand and desire sharply. To be sure, demand depends on desire, just as our Table 6 • 4 depended on Table 6 • 1. But we must remember that Table 6 • 4 depends on the utility of money as well as on the utility of the good. A man may have very great desire for a good and still not buy it, because he has not enough money. He places such a high value on money that he does not buy the good, despite his desire for it. Another man, desiring the commodity no more, does buy it because he has a good deal of money, and the marginal utility of the money is low. The economist's concept of demand couples the idea of desire with the idea of purchasing power. It depends on the utility of the good and also on the utility of money.

The economist notices also that there is no one single quantity of the good that buyers are willing to take. The quantity that will be bought varies with and depends on the price. Thus we get, not one quantity, but a whole series of quantities corresponding to various prices, as we see in Table 6 • 5 or in Fig. 6 • 1. This relationship between price and quantity is what the economist has in mind when he talks about demand. He thinks not of any particular quantity or price, but of the whole relationship shown by the demand curve. Thus his definition would say that *the demand for any commodity is the relationship between the price and the quantity that will be purchased at the price;* the demand is not a single quantity, but is all the quantities that would be taken at all possible prices. It is represented by a demand schedule or the demand curve. It does not represent quantities that consumers would like to have, but quantities that they actually stand ready to buy. I might like to have a new house, yet I may not be able to afford it. Demand couples desire with purchasing power.

6 • 11. Factors That Change Demand. At the end of Sec. 6 • 9 we noted that the market demand for any commodity is merely the sum of individual demands. Therefore anything that will alter individual demands will change market demand. But we have already noted, in Sec. 6 • 1 and 6 • 2, the things that will change individual demand. We saw, for example, that Table 6 • 3, showing an individual demand, was derived from Table 6 • 2 by selecting various prices and seeing how many units it would be advantageous for our buyer to purchase. The results depended on the facts given in Table 6 • 2; namely, on the desire for the goods and on the desire for money and other goods—on the buyer's desire for the commodity in question and on the alternative uses for his money. If we change these facts of Table 6 • 2, we get an

entirely new demand schedule. Thus we can say that anything that changes people's tastes for a commodity, or their desire for money, will change their demands for the commodity; and their desire for money will depend on their desires for other goods and on the prices of the other goods. Thus we see that the demand for potatoes will change if we change

1. Desires for potatoes.

2. Desires for other goods, especially for other similar goods, such as carbohydrate foods.

3. The prices of other goods, especially for other similar goods, such as carbohydrate foods.

4. The desire for money. The marginal utility of money will vary with our stock of money and, if we fall heir to an unexpected fortune, the marginal utility of money will fall and affect our demand for commodities.

One thing, however, we should note well: *The demand for a commodity is not changed by changing the price of the commodity.* The careless, superficial student is apt to fall into the error of stating that, when we raise the price of a commodity, we decrease the demand for it. As we have defined demand, this is not so. Let us look back to Table 6 • 5. If we raise the price of this commodity from 50 to 55 cents, we reduce the quantity that buyers are willing to take from 267 to 250 units. There is a reduction in the quantity purchased, but no reduction in the demand. We have already emphasized that the demand is not any particular quantity, but is the whole relationship represented by Table 6 • 5. Changing the demand means changing the table. Changing the price merely shifts us to a new position in the same, unaltered table. Raising the price of a commodity reduces the quantity taken but does not change the demand. The demand already includes all possible prices, as will be noted by rereading the fifth sentence from the end of Sec. 6 • 10.

6 • 12. How a Change in Demand Would Be Shown. Figure 6 • 1 shows the demand for a commodity. If we change the price from 90 to 40 cents, we do not change the curve but merely slide down to a lower spot on it. Changing the demand means changing the curve itself. Suppose, now, that something does happen to change the curve. Possibly the commodity becomes more stylish, so that more people want it. (This would represent a change in the desire for the commodity.) Perhaps substitute goods become more expensive, so that consumers shift to this commodity. (This would represent a change in the prices of other goods.) Possibly consumers' incomes are increased as a period of prosperity displaces a former depression, and as a result of the higher incomes people change their purchases. (This would represent a change in the desire for money.) The change in demand may come about as the result of any of the factors enumerated in Sec. 6 • 11. If our demand increases, consumers will buy more than they used to at any particular price. The whole curve is shifted upward and to the right. In

Fig. 6·2 we repeat the demand curve of Fig. 6·1 as the solid line, and then we show with a broken line a new demand curve farther up and to the right. If the demand curve shifts from the solid to the broken curve, we say that demand has increased. This happened, not because the price changed, but because of some one or more of the factors mentioned in Sec. 6·11. If the demand curve, on the other hand, shifts downward to the left, from the position of the broken curve to that of the solid curve, we say that demand has decreased. People buy less than they used to at any price. Perhaps the good has gone out of style, or good substitutes have appeared for it, or incomes

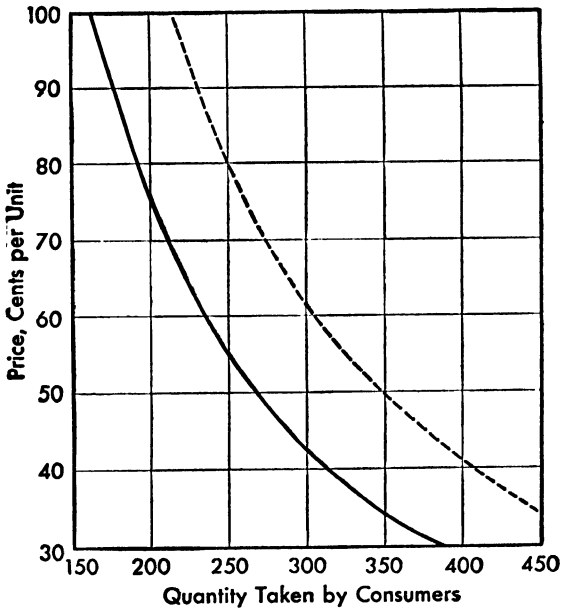


FIG. 6·2.—Illustrating a change in demand. The dotted curve shows a larger demand than the solid curve. Data are from Table 6·6.

have dropped, etc. Thus a change in demand would be shown by a shift of the entire demand curve.

Of course, it could be shown likewise by a change in the demand schedule. Thus the data of Table 6·6 correspond with the two curves of Fig. 6·2. The first two columns repeat the demand schedule of Table 6·5, while the last two columns are a demand schedule which corresponds with the broken curve in Fig. 6·2. When we compare the two demand schedules, we see that the right-hand schedule in the table represents a situation in which consumers are buying more at each price than they bought in the left-hand schedule of the table. For example, when the price of the commodity is 60 cents, buyers take 235 units in the left-hand schedule, but they take 305 units

in the right-hand schedule. Therefore, we say that the demand has increased in shifting from the left-hand to the right-hand schedule of Table 6·6, just as we said that demand had increased in moving from the solid to the broken curve in Fig. 6·2. A change in the demand means a shift to a new demand curve. It means a change in the figures in the table. When we change the price of a commodity, we move to a new position on the same curve, or to a new point in the column of the same demand schedule. Therefore, we say that changes in price do not change the demand but merely the quantity taken.

Table 6·6. Showing Two Contrasting Demands for the Same Commodity (see Fig. 6·2)

Smaller demand		Larger demand	
Price per unit, cents	Quantity buyers will take	Price per unit, cents	Quantity buyers will take
30	382	30	496
35	343	35	445
40	313	40	406
45	288	45	374
50	267	50	347
55	250	55	325
60	235	60	305
65	222	65	288
70	211	70	274
75	201	75	261
80	192	80	250
85	185	85	240
90	177	90	230
95	170	95	221
100	164	100	213

6·13. Demand Is Instantaneous. We saw in Sec. 6·11 that there are many things that can change the demand or a commodity. In fact, demands for various goods do change very often, every time consumers change their tastes or their incomes. The demand for a commodity is not an innate, inherent, permanent characteristic of the commodity. It is one thing today and another thing tomorrow. If we were to make a careful statistical study of the market for parsnips and, as a result, make up an accurate demand schedule for them, it might well be wrong by the time we got it made up because, while we were analyzing our statistics, the demand might have changed. The demand for a commodity cannot be computed once and for all and set down in the reference books for later use. The demand tells us the facts *at a given*

time and *on a given market.* Thus, we might find the demand for parsnips in Gurleyville in 1947. The demand in New York City in 1947 might be different. The demand in Gurleyville in 1948 might be different. Any demand schedule or demand curve tells the facts for some given market at some given instant of time and should not be assumed to depict accurately the situation in other localities or at other periods. Surely the demand for bustles is different today from that of 1890.

6·14. Summary. The demand for any commodity is a list of the quantities that would be bought by consumers at all possible prices. This demand depends on the attitudes of the buyers toward the good, and toward other things that they could buy with their money. Since economic goods are subject to the Law of Diminishing Utility, the quantities that people are willing to buy vary inversely with the price. Changes in taste and changes in income effect changes in demand, but changing the price of a commodity does not change the demand for it. The facts on demand are shown either by a table of figures, called a "demand schedule," or by a chart, called a "demand curve." These may show the facts of demand for an individual, or they may be schedules and curves showing the facts for an entire market.

Questions

1. Distinguish demand and desire.

2. What kinds of facts do we need to know before we can determine how many units of a commodity a man will buy?

3. Suppose you are given the data of Table 6·1 and are told that each dollar represents 9 utils and the price of the commodity is $2. How many units will a man buy? Take the same suppositions except that the price is $1. How many will he buy?

4. Suppose that the man in question 3 falls heir to a large fortune. What will be the effect on the marginal utility of his money?

5. In line with question 4, suppose that the man of Table 6·1 values his money at 3 utils to the dollar and the price is $2. How many units will he buy? Compare the result with the result at a $2 price in question 3. What was the effect of his enlarged fortune?

6. The demand schedule of Table 6·3 is computed from the data of Table 6·2. Suppose, now, that the man whose desires are depicted in Table 6·2 doubles his desire for the commodity and gets just twice the satisfaction from any number of units that he did before. This will double the figures in the second and third columns of Table 6·2, leaving all other figures unchanged. Compute a new demand schedule similar to Table 6·3, showing his demand under the new conditions. What is the effect on his demand of increasing his desire for the good?

7. Table 6·2 shows a man's desire for goods and his attitude toward money. Sup-

pose, now, that he loses his job and is out of work. Presumably he will have to be more careful of money than before, and its marginal utility will rise. If his attitude toward the commodity stays the same, as shown in the first three columns of Table 6•2, but the utility of his money doubles, thus doubling each figure in the last two columns of the table, what will his individual demand be like? Carry out the computations and compare with Table 6•3. What is the effect on his demand of his changed evaluation of money?

8. When we say that we cannot afford something, we do not always mean that we have no money to buy it. What do we mean?

9. Suppose you are
Given:
1. The data of Table 6•2.
2. Price of the commodity is $3.
3. The man has $9 and none of the commodity.
   a. How many units will he buy?
   b. What will be his consumer's surplus?

10. On cross-sectional paper draw the demand on the Gurleyville market as it would be if each of the quantities of Table 6•5 were exactly cut in half; that is, you are to suppose that the demand has decreased by that amount, and draw the new curve.

7. ELASTICITY OF DEMAND

7·1. Response to Price Change. In Chap. 6 we described the nature of demand. We learned that under ordinary circumstances the quantity of any commodity that buyers are willing to take depends on the price at which the commodity is offered for sale. When the price is lowered, consumers increase their purchases; when the price is raised, they reduce their purchases.

Although this is true in most actual cases, we find that commodities differ greatly in the degree to which consumers adjust to price changes. In the case of some commodities, even a small increase in price brings a sharp reduction in consumer purchases; in the case of others, even rather large price changes bring but small reductions in consumer purchases. The difference is one of degree. But this difference in the sensitivity of quantity to changes in price is of key importance in many economic problems, and we shall therefore study it more fully and describe it somewhat more precisely.

Maybe it will help to start with a simple numerical example. Suppose that, when the price of corn meal is 10 cents per pound, the consumers in Gurleyville are willing to buy 100 pounds per month; but at a price of 11 cents per pound they tend to shift to other grains and buy but 75 pounds of corn meal per month. We can summarize these facts thus,

Price, cents	Quantity, pounds
10	100
11	75

We see that the figures illustrate the Law of Demand, since the rising price brings a reduction in the quantity purchased.

Suppose that the consumers in Gurleyville are willing to buy 100 pounds of salt per month when the price is 10 cents per pound; if the price rises to

15 cents per pound, they reduce their purchases to 95 pounds per month. Again we summarize, thus,

Price, cents	Quantity, pounds
10	100
15	95

These figures also illustrate the Law of Demand, with consumers reducing their consumption as the price rises. The change in corn-meal consumption and the change in salt consumption are both of the same general nature. In either case a rising price brings reduction in purchases. But, if we compare the two cases, we see that the consumers make a much greater response to price change in the case of corn meal than they do in the case of salt. In either case, when the price is 10 cents, consumers buy 100 pounds, But even a small one-cent increase in the price of corn meal reduced consumption 25 per cent; while a change of five times as much in the price of salt reduced consumption but 5 per cent. Consumers are much more sensitive to changes in the price of corn meal than in the price of salt. We would say that the demand for corn meal, in our hypothetical case, is much more elastic than is the demand for salt. When we say that a demand is elastic, we mean that quantities purchased vary greatly when the price changes; but if price changes bring only minor adjustments in the quantity purchased, we say that the demand is inelastic.

7·2. Measuring Elasticity. We could, of course, leave the matter there. We could easily compare two such cases as those given above and see which was more elastic. But we would be unable to tell whether a single case, taken alone, was elastic or inelastic. Even if we were comparing two cases, they would not be likely to start off with both commodities selling exactly the same quantity at the same price, as was true in our illustration where we sold 100 pounds of salt and 100 pounds of corn meal when their prices were 10 cents.

We can easily tell that one day is cooler than another; but perhaps both days should be classed as cool, or both as warm. What one man calls a cool day may seem warm to another. Similarly, one man might call a given demand elastic, and another call it inelastic. Unless we define the concept more specifically, our results will be matters of subjective interpretation. The economist finds it advantageous to define the terms more precisely, so that there can be no question whether given data show an elastic or an inelastic demand. For this purpose he uses what at first glance looks like a difficult algebraic formula; but it turns out, when applied to actual data, to be very simple. The formula is

$$\eta = \left(\frac{\text{diff. in } Q}{\text{sum of } Q}\right)\left(\frac{\text{sum of } P}{\text{diff. in } P}\right)$$

In this formula, η is the Greek letter eta, corresponding with the ordinary letter e. It represents the elasticity of demand. The phrase "diff. in Q" is an abbreviation for "the difference in the quantities" (always for this purpose taken as positive). Similarly "diff. in P" means "the difference in the prices," again taken as positive. The phrases "sum of Q" and "sum of P" represent, respectively, "the sum of the quantities" and "the sum of the prices."

The formula will seem less terrifying if we apply it to the corn-meal example just studied. In this example, the two prices were 10 and 11 cents; so the sum of the prices is 21 cents and the difference between the prices is 1 cent. The two quantities were 100 and 75 pounds; so the sum of the quantities was 175 pounds and their difference was 25 pounds. If we substitute these values for sums and differences in our formula, we get

$$\eta = \left(\frac{25}{175}\right)\left(\frac{21}{1}\right) = \left(\frac{1}{7}\right)(21) = 3$$

Thus we get a numerical measure of the elasticity, and we say that in this example the demand for corn meal has an elasticity of 3. The student will find it good practice to compute similarly the elasticity of the demand for salt from the data of the example on page 88. In that case the sum of the quantities was 195 and their difference was 5; the sum of the prices was 25 and their difference was 5. Application of the formula will show an elasticity of 5/39, or approximately 0.128, as the student should verify for himself.

While we are at it, let us take one further example for practice. Table 6·5 shows a demand schedule. We note that, when the price falls from 65 to 60 cents, the quantity purchased rises from 222 to 235 units. What is the elasticity of demand? The sum of the quantities this time is 457 and the difference is 13. The sum of the prices is 125 and the difference is 5. Applying our formula, we get

$$\eta = \left(\frac{13}{457}\right)\left(\frac{125}{5}\right) = \frac{325}{457} = 0.71$$

7·3. Elasticity of Demand Defined. We have defined elasticity of demand by means of an algebraic equation and have worked out three numerical examples of it. Our results were 3.00, 0.128, and 0.71. These examples illustrate cases of both elastic and inelastic demand. We are now ready to define these two terms precisely by saying that whenever η, as we defined it in the preceding section, has a value greater than 1, the demand is elastic; but whenever the value of η turns out to be less than 1, the demand is inelastic. Thus in our hypothetical examples, the demand for corn meal was elastic, since η was 3, and 3 is greater than 1. In the other two cases the demands

were inelastic, since in each case η was smaller than 1. In the salt example, for instance, η was 0.128, so the demand was inelastic.

To be sure, we might find a case where the value of η was neither greater or smaller than 1, but exactly equal to it. For example, let us suppose that Gurleyville consumers buy 90 baskets of tomatoes per day when the price is \$1 per basket; but when the price falls to 75 cents per basket, they increase their purchase to 120 baskets a day. Application of our formula shows that

$$\eta = \left(\frac{30}{210}\right)\left(\frac{175}{25}\right) = \left(\frac{1}{7}\right)(7) = 1$$

In this case the demand is neither elastic nor inelastic, but lies on the border between. We call this "unit elasticity" or "unitary elasticity" of demand.

We can now summarize our definitions in numerical terms, as follows:

If $\eta > 1$, the demand is elastic.
If $\eta < 1$, the demand is inelastic.
If $\eta = 1$, there is unit elasticity of demand.

No longer do we have any difference of opinion as to whether a given demand is elastic or inelastic. The decision is not a matter of subjective interpretation, but one of simple arithmetic.

7·4. Relationship of Elasticity to Consumer Expenditures. Let us go back for the moment to our illustrative cases of corn meal and salt. In the former case we assumed that when the price was 10 cents consumers bought 100 pounds, but when the price was 11 cents they bought only 75 pounds. Let us notice now what happened to total consumer expenditures as the price rose and the consumption fell. We put the figures in convenient tabular form, as follows:

Price per unit, cents	Quantity sold, pounds	Total consumer outlay
10	100	\$10.00
11	75	8.25

Making the same sort of summary for salt we find

Price per unit, cents	Quantity sold, pounds	Total consumer outlay
10	100	\$10.00
15	95	14.25

Finally let us compare these with the case of the tomatoes in the preceding section.

Price per unit	Quantity sold, baskets	Total consumer outlay
$1.00	90	$90
0.75	120	90

The figures on consumer expenditures given in the last column show very great differences. When we raise the price of the product, consumers find that they are spending less for corn meal, spending more for salt, and spending just the same for tomatoes. It would be easy enough to show mathematically that these results are associated with the elasticity of demand,[1] and that we can make the following general statements:

When we raise the price of a commodity,

1. If the demand is elastic, consumers will make a smaller total expenditure than they did before.

2. If the demand is inelastic, consumers will make a larger total expenditure than they did before.

3. If there is unit elasticity of demand, the total expenditure will remain unchanged.

In our examples, the demand for corn meal was elastic, since $\eta = 3$; and we find that raising the price by 1 cent reduced consumer expenditures by $1.75. The demand for salt was inelastic, since $\eta = 0.128$; and we find that raising the price by 5 cents increased consumer expenditures by $4.25. The elasticity of the demand for tomatoes was unitary, since $\eta = 1$; and we find that, when the price rose 25 cents, consumers spent exactly the same amount as before.

Summarizing the few things we have been over so far we can say

1. Elastic demand:
   a. $\eta > 1$.
   b. Consumer expenditures vary inversely with price.
2. Inelastic demand:
   a. $\eta < 1$.
   b. Consumer expenditures vary directly with price.
3. Unit elasticity of demand:
   a. $\eta = 1$.
   b. Consumer expenditures stay constant as price changes.

[1] See Mathematical Appendix, Note 3, p. 887.

7 · 5. Elasticity of Demand Interpreted. It is important that the student should not get to think that the simple arithmetical operations through which we have been going are the important things about elasticity. To be sure, the number work is elementary, requiring nothing beyond addition, subtraction, multiplication, and division. But the danger always is that the student will get wrapped up (or snarled up) in the arithmetic and fail to see the economics. Therefore let us interpret our results in economic terms.

First, let us remember what the concept of elasticity is all about. We are discussing the responsiveness of consumer purchases to changes in price. When we say that the demand for a commodity is elastic, we mean that consumers make relatively great adjustments in their consumption when the price changes. A small rise in price leads them to make relatively large reductions in consumption; a small reduction in price leads them to increase their consumption considerably. When, on the other hand, we say that the demand for a commodity is inelastic, we mean that consumers make only slight adjustments in their consumption when price changes. The Law of Demand tells us that rising prices will reduce consumption but, if the demand is inelastic, the reduction will be a small one.

It is hard to compare price changes and quantity changes for different commodities. If the price of automobiles rises $10, we think of the change as insignificant; if the price of bread rises 10 cents a loaf, we think of it as a major price change. The 10 cent rise is a big one and the $10 rise a small one. To get comparability, economists convert all price changes to percentage terms. They then see that the $10 rise in the price of the automobile represented a rise of perhaps 1 or 2 per cent, while the 10 cent rise in the price of bread represented a rise of perhaps 80 or 100 per cent. Similarly when we consider the changes in consumption that arise as a result of the price change, we should put them into percentage terms. We do not care so much whether the consumption dropped 500 or 5000 units. We want to know whether it dropped 3 or 13 per cent.

In order to get complete comparability of different cases, we see what happens to consumer purchases when prices change 1 per cent. If we raised the price of corn meal 1 per cent and found that consumption dropped 10 per cent, we would know that the demand was elastic. The change in consumption was greater than the change in price (when both of them are expressed in percentage terms). If we found that a 1 per cent rise in the price of salt reduced consumption only one-third of 1 per cent, we would know that the demand was inelastic. The change in consumption was smaller than the change in price. If a rise of 1 per cent in the price of tomatoes reduced the consumption exactly 1 per cent, so that the reduction in consumption

and the increase in price were exactly equal, we would say that there was unitary elasticity of demand.[1]

Keeping this in mind, then, let us interpret the results of our earlier problems. We found that the elasticity of the demand for salt was 0.128. This means that, if we make a 1 per cent change in price, the quantity purchased will be changed but 0.128 per cent, a little more than an eighth of 1 per cent. The demand is inelastic. Consumers keep on using almost the same amount of salt regardless of price. If the price rises, they reduce their consumption only a little. If the price falls, they do not put much more salt on their potatoes but continue to use about the same amount. In another hypothetical example, we found that the elasticity of the demand for corn meal was 3. This means that if we make a 1 per cent change in the price of corn meal, the quantity purchased will be changed by 3 per cent. The demand is elastic. Consumers make rather large changes in consumption as the price changes. Finally, in our example of tomatoes, we found that there was unitary elasticity of demand. This indicates that, if the price of tomatoes rises 1 per cent, the quantity purchased will fall 1 per cent. The rising price and the falling consumption exactly counterbalance each other, so that the consumers' expenditures stay the same.

7·6. Factors Affecting Elasticity of Demand. Several things influence the elasticity of the demand for any product. In the first place, is the product a necessity or a luxury? If a commodity is a necessity of life, people have to use it even when the price is high, and they do not increase the use greatly when the price is low. If, on the other hand, it is a luxury, consumers can get along without it when prices are high, and therefore they tend to reduce their purchases considerably when prices are rising and to increase their purchases when prices are falling. It is easy to see that the demand for necessities will tend to be inelastic, and the demand for luxuries will tend to be elastic. Perhaps we would be justified in putting the thing the other way around and in saying that we determine what things are luxuries and what things are necessities from the elasticity of the demand for them. If we see that people insist on having a given commodity even when it can be had only at a high cost—that they pay the high price rather than go without the commodity, and give up other things—then we say that the commodity is a necessity. In Sec. 1·5 we said that it is hard to list things that are necessities in the sense that

[1] The formula that we have been using for elasticity is an approximation formula for finite differences. It would give the exact numerical result only if the change in price were extremely small; but it gives a good approximation for ordinary cases as long as one remembers not to apply it where there are really major price changes. As long as the price changes only a small percentage, the method will be reliable enough for our purposes. In practice it is usually the only method that can be used. The student who has studied the calculus will understand this better if he reads Note 3 of the Mathematical Appendix on page 887.

life cannot be supported without them. We really mean by necessities those things which men give up very reluctantly. If, on the other hand, we find people consuming something that means so little to them that they give it up at once when the price rises, we call it a luxury.[1] Surely it is easy to see that the idea of elasticity of demand is closely allied to the idea of necessities and luxuries.

Also, we would expect that the elasticity of the demand for a commodity would depend to some extent on the number of substitutes available, and on how good substitutes they were. If there are no good substitutes for a commodity, then people tend to keep on using it even when the price rises; but if there are several substitutes which are almost as good, then when the price goes up people tend to shift to substitutes. In this latter case the rising price drives them quickly to the substitutes, and there is a sharp drop in the consumption of the commodity in question. The demand tends to be elastic. Where there are few or poor substitutes, the demand tends to be inelastic.

We see that the elasticity of the demand for a commodity is not a permanent characteristic of that commodity, like its specific gravity or its tensile strength. The elasticity of demand changes as the number and effectiveness of substitutes change. We saw some time ago (see Sec. 6 • 11) that there are many things that may change the demand for a commodity. Any of them is likely to change the elasticity of the demand. Hence we could repeat here all the factors listed in Sec. 6 • 11 as forces that may change the elasticity of demand.

It has been suggested that the demands of wealthy people are less elastic than the demands of poor people. The supposition is that, if the price of a commodity rises, wealthy people can still buy as much, or almost as much, as before, because they have free funds. Wealthy people are not spending all their income but are saving some. They can meet the higher price by reducing their savings and continuing to buy as much or almost as much of the commodity as before. Poor people, on the contrary, have no such surplus to "take up the slack." They are already spending all they have (see footnote on page 48). If the price rises on one of the commodities that they buy, they must reduce their consumption somewhere, either on that commodity or on some other. Therefore, perhaps it would be reasonable to assume on a priori grounds that demands of wealthy people are less elastic than those of poorer people.

These suggestions as to factors affecting the elasticity of demand should give us some further understanding of the idea of elasticity. If, for example, we think of a product like meat, we ask ourselves, "Is meat a necessity of life?" No, not completely; yet as commodities run, meat would be classed with those things which are relatively necessary. "Are there good substitutes

[1] See Sec. 4 • 10.

for it?" To be sure, there are some substitutes. Some people never eat meat. But most men think that the substitutes are not very satisfactory. We would rather anticipate, even in advance of actual measurement, that there might be a rather inelastic demand for meat. If, however, we narrow the field and ask about the elasticity of the demand for beef, we realize that there are better substitutes. If beef prices rise, one can shift to pork or lamb or turkey. We would expect that the demand for beef would be more elastic than the demand for meat. Carrying the same point one step further, what of the demand for a particular cut of beef, such as sirloin steak? Now we have even more and even better substitutes. We can substitute not only pork and lamb and turkey, but also the other cuts of beef. The demand for sirloin steak should be more elastic than the demand for beef. And how about the demand for Swift's Premium sirloin steak? Now we have narrowed our field so much that there are many substitutes which, as the consumer sees it, are just as good. If the price of Swift's Premium sirloin rises, he can shift to Armour's or to some other brand. In a field where there are many competing brands, the demand for any one of them can be expected to be very elastic indeed. Thus we have moved in this illustration from a broad category, meat, where we thought the demand might be fairly inelastic, to a narrow category, Swift's Premium sirloin steak, where we would expect the demand to be very elastic indeed.

7·7. Importance of Elasticity of Demand. We must not end this chapter without stressing again the economic rather than the arithmetical aspect of elasticity. We take it up because of its importance in many diverse sorts of economic problems. Some of these things we have mentioned already. For example, we have noted that elasticity is important to consumers. It determines the effect on consumer expenditures of rising prices. Rising prices are always harmful to consumers, but the more elastic the demand the less harmful they are. If the demand is elastic and the price rises, consumers spend less money on the good and can offset their loss somewhat by spending more on other goods; if the demand is inelastic and the price rises, consumers spend more on the good, and not only get less of the good for their money, but have to reduce the purchases of other things too. Surely consumers have to be interested in the elasticity of demand.

Elasticity is also of importance at many other points in the economic sphere. We shall discover later (see page 848) that tax policy depends to a considerable extent on elasticity of demand. We shall note that the problem of monopoly and competition is fundamentally a problem of elasticity of demand (see page 294). We shall discover that the methods of regulating public utilities depend largely on the elasticity of demand (see page 606). We shall learn that the effect of collective bargaining depends largely on the elasticity of demand (see page 771). This concept of elasticity is of basic

importance to an understanding of economic problems, and it is of utmost importance that the student get a good understanding of it at the start, so that he can understand the problems that are to come. Most students want to stick, in economics, to what is "practical," and to skip the theory. They overlook the fact that elasticity of demand, like most theory, is the most practical part of the field and that it is impossible to understand the applied problems save against a background of theory.

7·8. Cross-elasticity. We have been studying, in this and the preceding chapter, the relationship between the price of a commodity and the quantity of it that is purchased. We have assumed that the prices of all other commodities stayed unchanged. But we know that the quantity of a commodity that is sold may depend not only on the price of that commodity but on the prices of other commodities. Thus, if we change the price of butter, it may change the consumption of oleomargarine. If we change the price of beef, it may change the consumption of mutton. If we change the price of gasoline, it may affect the consumption of tires. Thus we have various interrelationships of commodities, which we can best study by the method that we have been using in studying elasticity of demand. In each case we are changing a price and seeing what happens to a quantity purchased. In the case of elasticity of demand, we change the price of beef and see what happens to the quantity of beef purchased. In the case of cross-elasticity, we change the price of beef and see what happens to the quantity of lamb purchased.

A moment's reflection will show us the kind of thing that we can expect. If we raise the price of beef, we shall tend to shift consumption from beef to its substitutes and thus increase the consumption of mutton. The price of beef and the quantity of mutton purchased will tend to vary directly, one rising when the other rises. On the other hand, if we raise the price of gasoline, people will tend to use their automobiles less and consequently will not wear out tires so fast. The increased price of gasoline will tend to reduce the consumption of tires. An increased price of golf clubs might reduce the consumption of golf balls by causing people to shift to other forms of amusement and exercise. The price of gasoline and the quantity of tires purchased will tend to vary inversely.

In the first type of case, where we raise the price of one commodity and increase the consumption of the other, we call the commodities "substitutes" or "competing goods." In the second type of case, where we raise the price of one commodity and decrease the consumption of the other, we call the commodities "complementary goods." And we measure the degree of competition or complementarity by means of coefficients of cross-elasticity which are exactly parallel to the coefficient of elasticity of demand described in Sec. 7·2. The only difference is that there the prices and the quantities involved all referred to the same commodity. We studied cases, for example, where

the price of corn meal rose and consumers altered their consumption of it. Prices and quantities both referred to corn meal. But in our new case the prices refer to one commodity and the quantities to another. We study cases, for example, where the price of beef rises and consumers alter their consumption of mutton. The prices are prices of beef; the quantities are quantities of mutton. In cases of cross-elasticity there are always two commodities involved. Let us call them A and B. Then we can rewrite our elasticity formula from Sec. 7·2 as follows:

$$\text{Cross-elasticity} = \left(\frac{\text{diff. in } Q_A}{\text{sum of } Q_A}\right)\left(\frac{\text{sum of } P_B}{\text{diff. in } P_B}\right)$$

The symbol Q_A means quantity of commodity A; P_B means price of commodity B. Otherwise the elements of the formula mean exactly the same thing that they did in the case of elasticity of demand.

Let us take a numerical example. Suppose that, when beef is 50 cents a pound, consumers in Gurleyville buy 80 pounds of mutton a week; when beef is 55 cents a pound, they buy 120 pounds of mutton a week. What is the coefficient of cross-elasticity? The quantities are 80 and 120 pounds. The sum of the quantities is 200 and their difference is 40. The prices are 50 and 55 cents. The sum of the prices is $1.05 and their difference is 5 cents. Substituting these in our equation we find,

$$\text{Cross-elasticity} = \left(\frac{40}{200}\right)\left(\frac{105}{5}\right) = 4.2$$

Since the rise in price was accompanied by an increase in consumption, we call the answer positive. If we found that the cross-elasticity of beef and mutton was 4.2; that of beef and pork 3.1; and that of beef and horse meat was zero, we would say that mutton was a better substitute for beef than pork was, and that horse meat was no substitute at all. In comparing commodities, the larger the coefficient of cross-elasticity, the better the substitute —the more effective the substitute.

Let us now try our other type of case. Suppose that when the price of gasoline rises from 15 to 20 cents a gallon, the consumption of tires in Gurleyville drops from 80 to 60 tires a month. Our prices are 15 and 20 cents. The sum of the prices is 35 cents, and their difference is 5 cents. The quantities are 80 and 60. The sum of the quantities is 140 and their difference is 20. Our formula gives us the following information:

$$\text{Cross-elasticity} = \left(\frac{20}{140}\right)\left(\frac{35}{5}\right) = 1$$

But in this case, the rising price of commodity A was accompanied by falling

consumption of *B*. The relationship was inverse, or negative. Therefore instead of giving our answer as 1, we attach a negative sign to it and call it -1. When we see a negative coefficient of cross-elasticity, we know that the commodities are complementary goods—goods that are usually used in combination. And the larger the coefficient, the closer is the relationship of complementarity. Thus, if we find a coefficient of cross-elasticity of -1 between gasoline and tires, and a coefficient of cross-elasticity of -0.04 between gasoline and admissions to big-league ball games, we would know that in each case that the pair of goods was complementary. Gasoline and tires are used together, and gasoline and big-league ball games are used together (possibly because people drive their cars to the game). But the relationship is very much closer between gasoline and tires than between gasoline and ball games. In fact, there is almost no relationship in the latter case, since the coefficient of cross-elasticity is almost zero; but there is quite a close relationship between gasoline and tires, since the coefficient differs considerably from zero.

Let us summarize this matter of cross-elasticity:

1. Compute the coefficient of cross-elasticity by formula.
2. Attach either a positive or a negative sign to the coefficient, as follows:
   a. If prices and quantities increase or decrease together (move in the same direction), attach a positive sign to the coefficient.
   b. If prices and quantities move in opposite directions, one increasing and the other decreasing, attach a minus sign to the coefficient.
3. If the coefficient is positive, the two goods are substitutes or competing goods. If the coefficient is negative, the two goods are complementary goods, which are used in combination.
4. In either case, the greater the absolute size of the coefficient (the farther it lies from zero), the more pronounced is the relationship. If the coefficient is zero, the goods are what we call "independent" goods.

7·9. Practical Difficulties in Measuring Elasticity of Demand. So far we have been trying to get a clear understanding of the meaning of elasticity; and to that end we have taken a number of simple numerical examples. In practice, it is extremely difficult to measure the elasticity of demand for any commodity. For example, we used data showing prices and quantities of corn meal. We said that when the price of corn meal was 10 cents a pound, consumers bought 100 pounds but, when the price rose to 11 cents a pound, they reduced their consumption to 75 pounds. In practice, how would we get such data? We might, of course, study the market to see what happened. We might notice, for example, that in 1940 the average price of potatoes in the United States was 54 cents and at this price the amount sold was 378

million bushels. Then, in 1941, we find the facts again and learn that the average price was 70 cents and the amount sold was 358 million bushels.[1] Why not use these as our price and quantity figures, computing the elasticity of demand as $\eta = 0.21$? Right away we see the difficulties. Perhaps the potatoes of 1941 brought a higher price because they were of higher quality or because consumers had higher incomes. In other words, quite possibly the demand changed between 1940 and 1941. The demand for a commodity is a temporary, transitory, evanescent thing, depending on people's tastes, their incomes, and the prices of other goods (see Sec. 6·11). If we could get two points on the same demand curve, we could measure the elasticity of the curve. Probably in this case we have points on two different curves—one a point on the demand curve for potatoes in 1940, and the other a point on the 1941 curve. If these curves are the same, we can find the elasticity of either one from our figures. But we have no way of being sure that the demand is the same and, since there are so many factors that can change it, the chances are pretty good that the demand has changed. In that case we have points on two different curves, and our figures cannot tell us the elasticity of either.

Demand is a peculiar thing. It is, as we have seen, not a single quantity, but all the quantities that would be taken at all possible prices. One of these particular pairs of prices and quantities exists—the actual price that there is on the market, and the quantity that is actually taken at that price. The rest is nonexistent. The rest of it is quantities that would have been taken if the price had been different. What quantity would have been taken if tastes and incomes and prices of other goods had been the same, but the price of this good had been different? We never can know. One point on the demand curve exists, and the rest is what would have existed if things had been different! This makes a peculiarly difficult thing to measure.

This is not the place to discuss the methods that astute and resourceful economists and statisticians have concocted for dealing with this problem. Enough here to note that the problem is by no means so simple as our fictitious examples have made it appear. All sorts of corrections have to be made to the data used in actual practice, and strangely enough some of the data used are very different in form from those which we have used here, being much more like those used by Engel in his studies of consumer expenditures (see Sec. 4·10). As these are problems of advanced economic theory and of statistics, we shall have to pass them over for treatment in more advanced courses. Suffice it to say that economists have succeeded in making what they think are fairly reliable measures of the elasticity of the demand for many different commodities. Also let us stress again the fact that

[1] These data are from the "Statistical Abstract of the United States, 1942," Table 738, p. 775.

no one can understand the major applied problems of economics without at least understanding the concept of elasticity, whether he ever succeeds in measuring it accurately or not. For our purposes we shall continue to assume that we have the necessary data, as we have done in the cases of corn meal and salt and tomatoes, and we shall let the advanced worker in the field worry about how the data are to be obtained.

We run into another difficulty which we have purposely neglected so far so as not to complicate our problem while we were getting an understanding of it. That is the fact that a given demand curve for a given commodity may have many different elasticities at different points on the curve. Suppose, for example, that we were given the demand curve shown in the first two columns of Table 7·1.

Table 7·1. Illustrating Variation in Elasticity at Different Parts of the Same Demand Schedule

Price per unit, cents	Units taken by consumers	Elasticity at this region
100	330	
		0 28
90	340	
		0.49
80	360	
		0.60
70	390	
		1.08
60	460	
		1.36
50	590	
		1.77
40	880	

The table illustrates the Law of Demand, with higher prices bringing reductions in the quantity taken by consumers. Let us consider the first two pairs of figures—the pair of price and quantity at the top of the table and the next pair immediately below them. We see that when the price fell from $1 to 90 cents, the quantity increased from 330 to 340 units. If we compute the elasticity of the demand by our usual formula, we find that $\eta = 0.28$. We enter this in the third column on the line between the two pairs on which we computed it. Now let us drop down to the second and third rows of figures in the table. We see that, when we drop the price of the commodity from 90 to 80 cents, consumers increase their purchases from 340 to 360 units. Using these figures as the basis, we compute $\eta = 0.49$, which we set in the third column on the line between these two pairs of figures. Similarly we

compute the elasticity of demand between each of the successive pairs of prices and quantities, setting our results at the appropriate spot in the third column. What do we find? What is the elasticity of the demand for this product? There is not one elasticity, but six of them. As a matter of fact, there are doubtless a limitless number of them. We have computed the six that correspond to the six pairs of prices and quantities given in the demand schedule but, if we took others of the possible prices and the quantities corresponding to them, we would doubtless find still other elasticities.

It is possible (although certainly not likely) that a demand curve should have the same elasticity at all points. In such a case we would say that there was "constant elasticity." And of course, we could then say that the elasticity of demand was 0.3 or 2.1 or whatever this constant value was. But, when the elasticity varies from point to point along the curve, our problem is more difficult. In such a case we would either have to say that the elasticity was a certain amount *at a given part of the curve,* naming the prices and quantities involved; or we might pick out some average price and quantity and give the elasticity at that region of the curve. Thus we could say, with regard to the data of Table 7 • 1, "The demand for this commodity has an elasticity of 0.60 when the price is in the neighborhood of 70 to 80 cents." If we hear someone say, "The elasticity of demand for milk is 0.45," we shall assume that he means that this is true at some price that he thinks of as normal or average.

Sometimes we distinguish our ideas of constant elasticity from what we call "stable elasticity." The former, as we have seen, means that the elasticity is the same at all parts of the demand curve. The latter means that the elasticity stays the same from time to time. If we were to study demand curves for milk and were to find year after year that the elasticity was the same at given prices, we would say that the elasticity was stable. If, for example, we got each year a table similar to Table 7 • 1, with different elasticities at the different prices but always with the same elasticities as those given in the third column, we would say that the elasticity was stable even though it was not constant. In all cases where the student is asked to measure elasticity, he should take it for granted that he is to find the elasticity in the region of the curve for which the prices and quantities are furnished.

7 • 10. Graphic Representation of Elasticity. When we say that the demand for a given commodity is very elastic, we mean that even a small price change brings a large change in the quantity that consumers will buy. Such a situation would be shown, of course, by a demand curve like that labeled D_1 in Fig. 7 • 1. This curve has a very gentle slope, so that even a slight drop in its height brings a considerable lateral movement across the diagram. But vertical distances represent price changes, and horizontal distances represent

changes in quantity, so that a curve with gentle slope such as this one represents an elastic demand. On the other hand, the curve labeled D_2 in Fig. 7 • 1 is very steep. Even a very considerable vertical drop on the curve brings but a slight lateral movement across the diagram. Large price changes are associated with small changes in quantity, and the demand is inelastic. If, as usual, we show prices on the vertical and quantities on the horizontal axis, steep curves will represent inelastic demands and gentle curves will represent elastic demands. To be sure, this rule is not exact for two reasons. First, we can alter the slope of a curve by altering the scales along our axes. If we reduce the horizontal scale, the curve will become steeper, even though exactly the same data are plotted. There is no change in the elasticity, merely a change in the graphics. It is also true that elasticity is not a simple question of changes

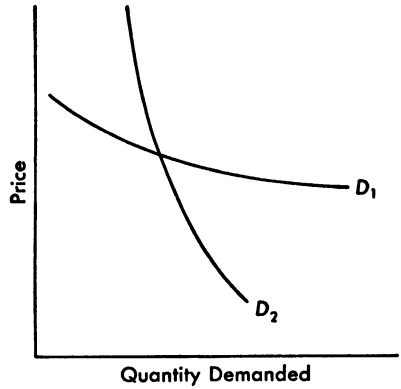


FIG. 7 • 1.—Two demand curves which differ in elasticity. At the point where the two curves cross, curve D_1 is more elastic than curve D_2.

in prices and the corresponding changes in quantities demanded. It is a matter of *percentage* changes in price and the corresponding *percentage* changes in quantity. After the scales of our diagram are given, the slope of a given curve is determined, but the curve will have very different elasticities at different points if slope is the same at different points. The reason for this is easy to understand. Suppose we take two cases in which there is exactly the same change in price and in quantity. In the first case, when the price is \$10, the quantity sold is 100 units; when the price rises to \$10.10, the quantity drops to 90 units. In the second case, when the price is \$1, the quantity sold is 1000 units; when the price rises to \$1.10, the quantity drops to 990 units. In each case the change in price is 10 cents and the change in quantity is 10 units. The changes in price and the changes in quantity are the same, and the two curves will show the same slope on ordinary graph paper. But, if the student will work out the elasticities of demand for the two

cases, he will discover that they are very different indeed. In the first case $\eta = 10.6$, and in the second case $\eta = 0.105$. The first demand is very elastic, and the second is very inelastic. The reason, of course, is that a change of 10 cents in price is *proportionately* a much smaller change when the price is \$10 than when it is \$1; and a change of 10 units in the quantity is *proportionately* a much smaller change when based on a quantity of 1000 units than on 100 units. Equal elasticities exist when given *proportionate* price changes bring equal *proportionate* changes in quantity.

However, despite the fact that the slope of a curve on a given diagram does not show us the elasticity directly, it is worth while to remember that

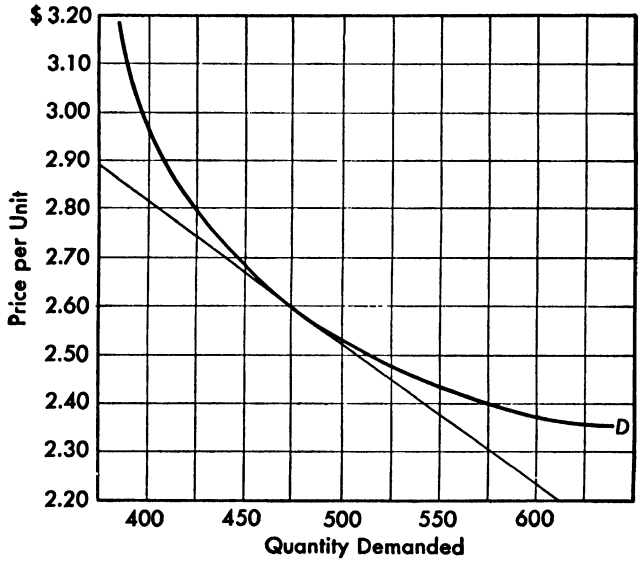


FIG. 7 • 2.—Graphic determination of elasticity of demand.

if, at given prices and quantities, two curves differ in slope as do those of Fig. 7 • 1 at the point where they cross, then the steeper curve is the less elastic.

We can find the approximate elasticity of the demand for a commodity from the demand curve by graphic means. Suppose that the demand for a commodity is shown by the curve labeled *D* in Fig. 7 • 2. We note that at a price of \$2.60 consumers are willing to buy 475 units of the commodity, and we wish to know the elasticity of demand at this point on the curve. (Since the elasticity is usually different at different parts of the curve, we must of course specify the point at which we are determining the elasticity.) First we draw a tangent to the demand curve, which is the straight line barely touching the curve at the point where we wish to measure the elasticity. We

note that this tangent cuts the vertical axis at a point that represents a price of $2.90, or at a distance above the base line of 70 cents. The tangent cuts the horizontal axis at a quantity of 610 units, or at a distance from the origin of 235 units (since 375 from 610 leaves 235). Now, if we divide the horizontal quantity distance (235) by the vertical price distance (70) and if we divide the price at the point of tangency (260) by the quantity at the point of tangency (475), we get two quotients which, when multiplied together, give the elasticity of demand. To make a general rule for finding the elasticity at any point on the curve, we find four things:

A. The price corresponding to the given point on the curve;

B. The quantity corresponding to the given point on the curve;

C. The difference between the quantity where the tangent cuts the base line and the quantity at the left-hand side of the diagram;

D. The difference between the price where the tangent cuts the vertical axis and the price shown at the bottom of the vertical scale.

We then find the elasticity by the following formula:

$$\eta = \frac{C}{D}\frac{A}{B}$$

where the letters represent the four things enumerated above. In our problem these are

A is 260 cents,
B is 475 units,
C is 610 — 375 units, or 235 units,
D is 290 — 220, or 70 cents.

Hence the elasticity of demand at this point on the curve is

$$\eta = \frac{235}{70}\frac{260}{475} = 1.8$$

Since it is difficult to draw the tangent accurately unless the diagram is on a very large scale, the results obtained by this method will be only approximate, but it is surprising in practice to see how closely the results obtained by a number of different students on such a problem approximate one another.

7 · 11. Summary. The Law of Demand tells us that rising prices tend to reduce the quantity that consumers buy. In some cases the reduction is great and in some cases small. The degree of response is called "elasticity of demand." If the consumers make great changes in the quantity of consumption when the price changes, we say that the demand is elastic; if they make only minor changes in their consumption in response to a price change, we say that the demand is inelastic. There are simple arithmetical methods for measuring the degree of elasticity if we can get two prices and their cor-

responding quantities. In practice such basic data are hard to get, and the problems met in getting them are outside the scope of an elementary text. A knowledge of elasticity and its effect is necessary if we are to understand many applied problems of economics. We need to know both what forces affect the elasticity of demand and how the elasticity of demand in turn affects the problems that we shall study. We find that the methods that we develop for studying elasticity of demand can also be used, with minor modification, to determine whether pairs of commodities are competing goods, complementary goods, or independent goods, and the degree of the relationship, if any, between them. This raises the problem of what economists call "cross-elasticity."

Questions

1. Describe the concept of elasticity of demand in simple and nonmathematical terms as you might describe it to a businessman who had not studied any of the technical terms but who had a good intellect.

2. We are told[1] that, when the price of corn is cut to 82 per cent of normal, the volume sold is 20 per cent above normal. What is the elasticity of the demand?

3. We are also told[2] that the price necessary to cut consumption of each of the following crops to 80 per cent of normal is the following:

	Price necessary above normal, per cent
Potatoes	29
Wheat	25
Oats	23
Hogs	21
Apples	17

Compute the elasticity of demand for each crop. In each case, is the demand elastic or inelastic?

4. Tell in simple terms how the effect of changing prices on consumers' total expenditures is dependent on the elasticity of demand.

5. On purely a priori grounds, which would you expect to have the more elastic demand, cucumbers or wheat bread?

6. Why are elasticities of demand measured in percentage terms rather than in absolute terms?

7. In our purely fictitious examples we found that the elasticity of the demand for corn meal was 3 while that for salt was 0.128. Suppose that a study of the two commodities really showed that the two elasticities were 3 and 0.128, would you expect the former to be salt and the latter corn meal, or vice versa? Why?

[1] *Farm Economics*, No. 142, October, 1944, p. 3618, Cornell University, Ithaca, N. Y.

[2] *Ibid.*, p. 3619.

8. What do you think would happen to the elasticity of the demand for beef if someone discovered a perfectly practicable method for extracting a good protein food from the air? Why?

9. Suppose we measure the elasticity of demand for wheat in general. Then we measure the elasticity of the demand for the wheat produced by a particular farmer in McPherson, Kans. How will the two elasticities compare? Why?

10. When the price of commodity *A* rises from 25 to 30 cents, consumers increase their use of commodity *B* from 200 to 250 units. What is the coefficient of cross-elasticity? How do you interpret the result?

11. When the price of commodity *X* rises from 80 to 90 cents, consumers reduce their purchases of commodity *Y* from 4000 to 3950 units. What is the coefficient of cross-elasticity? How do you interpret it?

12. Define competing goods; complementary goods. Give an example of each in addition to those given in the text.

13. Define constant elasticity; stable elasticity.

14. On page 100 we computed the elasticities at various points along the demand schedule of Table 7 • 1. In Table 6 • 5, appears another demand schedule. Compute the elasticities at the different parts of this schedule. Is the demand constant or approximately so? What else would you have to know to decide whether or not the demand was stable?

15. Suppose that consumers increase their consumption of pork from 175 to 200 pounds a week in a given town when the price drops from 55 to 53 cents a pound. What is the elasticity of demand? Is it elastic or inelastic?

16. In a given market consumers spend $5000 per week on eggs. A tax then raises the price of eggs, and as a result consumer expenditures rise to $5200 per week. Do you think that the consumers were using more eggs or fewer eggs after the tax than before? Do you think that the demand for eggs was elastic or inelastic? Explain.

17. A housewife says, "The high price of milk is very hard on poor families. After all, I have just so much money that I can spend on milk. If the price is high, I have to reduce the amount I buy. If the price is cut in half, I can buy twice as many quarts with my milk budget." If she tells the truth, is her demand for milk elastic or inelastic? Explain. Is this a case of individual demand or market demand? If all housewives are like her, what can we say about the elasticity of the demand for milk in general?

18. Actual figures seem to show[1] that when the price of cotton is at half the normal level, consumers buy about 78 per cent more than normal; when the price is twice

[1] The data for this question are read from a graph in John D. Black, "Parity, Parity, Parity," p. 198, Harvard Committee on Research in the Social Sciences, Cambridge, 1942.

normal, they buy about 56 per cent of normal. What is the elasticity of demand? Is the demand elastic or inelastic?

19. Figures for the years 1920-1939 indicate[1] that, when the price of wheat is 58 per cent above average, consumers buy only 56 per cent as much as usual; when the price drops to 89 per cent of normal, they increase their purchases to 148 per cent of normal. What is the elasticity of the demand for wheat? Is the demand for wheat more or less elastic than the demand for cotton (see question 18)?

[1] These figures are from the same source as those of the preceding question.

8. CONTROL OF DEMAND

8 • 1. Consumer Ignorance. In the field of production, men soon become expert; they learn the "tricks of the trade." The apple raiser learns about sprays and insecticides, storage and marketing; the shoemaker learns about leathers and lasts, preservatives and methods of stitching; the tailor learns about fabrics and weaves, sewing machines and styles, and so it goes. Each of them becomes acquainted with a good deal of technical jargon and information that is outside the experience of most other people.

The usual consumer is by no means an expert. He buys dozens or hundreds of different things, and he cannot become expert on them all. He would have to be an electrical engineer to understand his radio, and a skilled musician and a cabinetmaker in the bargain. He would have to be a chemist when he bought beefsteak and a metallurgist and skilled appraiser of gems when he bought his fiancée's diamond ring. He would have to be a carpenter and builder when he bought or rented a house, as well as a legal expert and a judge of plumbing, wiring, lighting, etc. The field of consumption is so wide and varied that no one masters more than a very small part of it. And there seems to be little likelihood that this situation can be changed as long as men continue to want so many different things. It has been said, for example, that in 1941 the stores of one large five-and-ten-cent chain carried an average of over 4600 different items. What consumer lives long enough to become expert on even 10 per cent of them?

This ignorance on the part of consumers raises many problems, only part of which are of an economic nature. Not the least of these is that the consumer may, in his ignorance, spend some part of his precious and limited resources for goods which will not really satisfy his wants and which at worst may even be positively deleterious to him. In fact, his ignorance may be so great that he buys and continues to use goods that are harmful without knowing enough to realize that they are doing him harm. Nor should we overlook the fact that the consumer may be taught to want goods that a more enlightened mind would have realized were unnecessary. In such a case

the goods that he buys do satisfy his wants, but the wants are those which he should not have had. Here we get out of the field of science and into ethics and morals with a vengeance! (See Secs. 2·1 to 2·5.)

8·2. Groups Interested in the Control of Demand. We have seen (Sec. 6·11) that there are many factors that may change demand and also many agencies that would like to see these demands changed. In the first place, consumers presumably have some interest (often, to be sure, not an overt nor a conscious one) in controlling their own demands for products. If they can be taught to want some things and to do without others, their lot will be easier if they can be taught to want those things which are easy to get, and to shun those which can be obtained only with great expenditure of time and effort. If they are to have food, it is better for them to learn to like those foods which are conveniently available in their own locality, rather than those which have to be grown at great expense and transported at high cost from distant parts of the world. The nutritive value may be identical, and possibly the consumer can be taught to want one as easily as the other. In such a case he has some interest in the result.

Consumers have usually taken relatively little interest in their demands. In fact, their demands appear to them as given things, beyond their control. And to the extent that wants are products of society (see Sec. 1·4) rather than products of individual thought and reflection, consumers are right in concluding that there is nothing they can do personally about their wants.

There are two other groups of people, at least, who are interested in the consumer's wants and who would like to manipulate and change them. These are the government and people with goods to sell. Their interests are, to be sure, very different. Every government is interested in the consumption habits of its citizens, and every seller is interested in the desires of every prospective buyer. We shall turn our attention first to some of the activities of government in the field of demand and then to some of the activities of producers and sellers of goods.

8·3. Sumptuary Legislation. Sumptuary legislation is legislation that is intended to alter consumption habits—to change the goods and services that people consume. It may be that it is hoped to increase this consumption or to decrease it, or possibly even to prohibit it altogether; but, unless the law in question is intended to change consumption, it is not sumptuary legislation.

It will be seen from this definition that sumptuary legislation is not easy to identify with certainty. We have not said that it is legislation that *does* influence consumption, but legislation that is *intended* to influence consumption. It is a matter of purpose rather than of accomplishment. As soon as we think of it, we note that a great deal of legislation does affect consumption indirectly even though it is not passed with that aim. Thus, a law that lays a tax on gasoline may have as its sole purpose the raising of revenue for the

support of the government, but it will ordinarily raise the price of gasoline and reduce the consumption of it. We would not regard it as sumptuary legislation. If we were to include all legislation that does affect consumption to any extent, directly or indirectly, we would have to include a large share of all legislation.

Sumptuary legislation is based on the assumption, not by any means always tenable, that the legislator knows more than the consumer, that he is a better judge of what the consumer should consume than the consumer is himself. To be sure, we have just been emphasizing the ignorance of consumers, and we shall have occasion to emphasize it again. It is natural for us to suggest that the lawmakers should get advice from the best minds in each field and, on the basis of that advice, regulate consumption. Do the nutrition experts know more about what the average consumer should consume than he does or can know for himself?

There are several answers to this question. One is that lawmakers are as likely to get their advice from interested and biased experts as from disinterested and altruistic ones. Experts are hard to recognize, and it is especially difficult to recognize one in a field of which a man is himself largely ignorant. Another answer is that the consumer wants more from his food than nutrition, and the scientist's appraisal of any consumption good is only a partial one. The fact that unpolished rice is more nutritious than polished rice is no indication that people should eat the unpolished variety nor that they should abstain from polished rice. If they like polished rice better and prefer to get the nutritive elements elsewhere, they may well think it an imposition for their government to force them to eat unpolished rice. The difficulty is that the scientist often thinks of consumption in scientific terms, while as a matter of fact consumption is spiritual and artistic in addition. It is presumptuous of any "expert" to assume that he knows better what consumers "ought" to want than they know for themselves. But we can get a better understanding of the place and functions of sumptuary legislation if we examine its history and its present purposes. To those matters we now turn our attention.

8·4. Historical Changes in Sumptuary Legislation. The past century or two have seen tremendous reductions in the amount of open and evident sumptuary legislation. Although we continue to have a good deal of it, it is usually now more subtle and indirect, supposedly directed at other objectives. But two or three centuries ago sumptuary legislation was open and common, detailed and voluminous. Many of the everyday activities of consumers were controlled by law. Thus, the laws of England once specified just what clothes might be worn by men of various professions and vocations, and by the members of their families; how many meals they might eat a day, with how many courses to a meal; how often they might eat meat or drink wine;

etc. We would have a comparable situation today, without any exaggeration, if the law specified that automobile mechanics were to eat oatmeal and popovers for Monday breakfast and that their wives were to wear only black shoes; that they were to put 75 cents in the church collection plate on Sunday and subscribe to no newspaper with headlines over 2 inches high. In fact, these old laws went into minute detail in regulating the lives of various kinds of artisans, professional men, nobles, and the other classes that made up the society of that day.

This old sumptuary legislation seems to have been intended in part to maintain and emphasize class distinctions, which were beginning to disappear. It was intended to maintain the *status quo* in the face of threatened change; in part, its avowed purpose was to check extravagance. But the lawmakers of the day thought it extravagant for a family of "low estate" to consume those things which, as they saw it, should be reserved for the nobly born. Without question, this legislation was also used sometimes to prevent the introduction of new goods. Thus most of the motives that lay behind this older sumptuary legislation were those which are unpopular today, especially in the democratic countries.

In fact, sumptuary legislation has always been unpopular and difficult to enforce, especially in the United States. Its lack of popularity here has been the result in part, perhaps, of the fact that it still carries the stigma of aristocratic snobbishness, and in part of the historical accident that this country grew out of a revolution that had as one of its excuses such sumptuary legislation as a tax on tea. Modern sumptuary legislation is, in fact, very different in purpose from that of the late Middle Ages. It is based less on a desire to maintain the *status quo* and more on a desire to protect the consumer against the results of his own ignorance and weakness. It recognizes the fact that life has become too complex for any consumer to master, and that the effects of errors in the field of consumption may be so far-reaching and of such a serious nature that such errors cannot be allowed. Today's legislation in this field has shifted from the old maxim of *caveat emptor* to the newer admonition, *caveat venditor*.

8·5. Sumptuary Legislation in the United States. In every modern country there are national laws and local regulations almost without number that influence consumption. In the United States, the Federal government passes laws regulating the consumption of articles that pass in interstate and foreign commerce; the 48 states pass laws regulating the purchase and sale of commodities within their own borders; and each city or town passes ordinances or bylaws or regulations designed to protect the health and safety of its citizens by influencing consumption. We shall deal here with only a few of the more important of the Federal regulations, since the activities of the states and municipalities differ greatly from one jurisdiction to another.

Most modern sumptuary legislation takes one or another of five forms: (1) the prohibition of the manufacture, sale, or consumption of certain products, (2) the requiring of licenses or permits to engage in certain occupations or businesses, thus allowing state inspection and control, (3) state regulation of the production or sale of products, (4) government monopolies, or (5) taxation. Although many authorities put a good deal of stress on taxation as a type of sumptuary legislation and although it could perhaps be effective in this field, nevertheless the examples usually given are not sumptuary legislation at all as we have defined it. For example, the protective tariff does raise prices of foreign goods and thus shift consumption from those goods to commodities of domestic origin. But the purpose of tariffs is either the raising of revenue or the protection of producers, and their effect on consumption is incidental to these ends. No one is trying to help the consumer with the tariff, and it cannot properly be classed with sumptuary legislation. The same thing may be said of a large proportion of the so-called "sumptuary" laws that operate through taxes. Moreover, government monopolies as a method of influencing consumption have been rather uncommon in the United States. Perhaps the postal service is such a monopoly, since it is often suggested that its purpose is to increase the amount of intercommunication within the United States. Some countries have set up liquor monopolies in an effort to overcome some of the evils of that trade. But we can pass over both taxation and government monopoly as methods of influencing consumption in the United States on the ground that they are relatively unimportant. This leaves us the other three types of legislation, all of which are fairly common. To be sure, most sumptuary laws in the United States combine two or more of these forms of legislation, so that it is often difficult to classify a given statute or government agency. The student will notice that most of the activities fall into one or another of these three remaining fields.

8·6. Food and Drug Administration. One of the earliest of the nationwide attempts to protect consumers against their own ignorance and weakness was the Pure Food and Drug Act, passed originally in 1906 as a part of President Theodore Roosevelt's "reform" of big business. The act has been amended over and over again, and the responsibility for administering the act, which long rested with the Department of Agriculture, has now been given to the Federal Security Agency. Major changes were made in the underlying laws in 1938, effective July 1, 1940. There is no particular point in tracing here all the stormy history of the original act. We shall rather describe the present law and the activities that are carried on under it.

The Food and Drug Administration is now responsible for the enforcement of five statutes intended to ensure the purity and honesty of food, drugs, cosmetics, and curative devices. These statutes are the Food, Drug, and Cosmetic Act; the Tea Act; the Import Milk Act; the Filled Milk Act;

and the Caustic Poison Act. These acts are intended to secure freedom from adulteration and to require the truthful labeling of certain commodities. In order to carry out the provisions of the acts the Food and Drug Administration was employing, at the start of the war in 1941, a field force of some 600 persons and a Washington staff of about 300. These figures include chemists, bacteriologists, physicians, veterinarians, pharmacologists, and many other specialists in addition to the necessary clerical and administrative workers.[1] Their major activities can be visualized from the names of the major divisions of the agency: a Bacteriological Division, a Cosmetic Division, a Drug Division, a Food Division, an Interstate Division, a Microanalytical Division, a Division of Pharmacology, a Division of State Cooperation, and a Vitamin Division.

The Food, Drug, and Cosmetic Act as it stands at present has both positive and negative provisions with respect to labeling. The negative provisions prohibit false or misleading statements on the labels of foods, drugs, cosmetics, or curative devices, on the package or container, or on written or printed matter accompanying them. The positive provisions state that certain facts must be stated on the label. It is not enough that the label be true; it must also be complete. Even if the label and all accompanying printed matter are acceptable, the container itself must not be so formed or filled as to mislead the buyer. The administrators of the act are also empowered to stipulate that the information describing the goods must be in such form and terms that an ordinary consumer would be likely to see it and able to understand it. Foods that do not conform to some legally established standard must reveal on their label a full statement of ingredients. It is under this provision that the consumer finds, for example, on the wrapper of his loaf of bread some such statement as, "This bread contains wheat flour, granulated sugar, skim milk solids, table salt, pure shortening, yeast and yeast nutrients, malt syrup, and water. Six ounces of this bread supply at least the following amounts or percentages of the minimum daily requirements of these food elements: 41% Vitamin B-1; 13% Vitamin B-2 (Riboflavin); 3.7 mgs. Niacin; 30% iron." The presence of artificial flavoring, artificial coloring, or chemical preservatives must be stated with a few exceptions. Drugs and curative devices must bear labels that give adequate directions and warnings as to use, dosage, and administration. Those containing habit-forming ingredients must carry plain warning to that effect.

The Food and Drug Administration, however, has jurisdiction over and responsibility for many things in addition to labeling. For example, they check goods against adulteration, manufacture under unsanitary conditions,

[1] The figures given here, as well as much of the other descriptive material, are condensed and adapted from *Miscellaneous Publication* No. 1, Food and Drug Administration, Washington, D. C. 1941.

or use of inferior and harmful materials. They may in some cases set up standards under which goods must be produced under permit. They may require proper tests and experiments as to the safety of new drugs before they are sold to the general public; and in practice they have been carrying out such tests at the rate of several hundred new drugs per year. They may set up standards to be met by foods that enter interstate or foreign commerce. These standards usually specify the ingredients and their proportions within limits. For example, they might state that any product sold as raspberry jam must have at least 45 per cent by weight of raspberries. Otherwise an unscrupulous manufacturer or dealer might sell a concoction in which raspberries were a negligible ingredient. These food standards may specify such things as color, size, texture, quantity, and freedom from specified defects.

8 • 7. Enforcement of the Food and Drug Acts. There are far too many producers and handlers of food, drugs, and cosmetics for any government agency to investigate and inspect all goods that enter interstate or foreign commerce. Consequently the Food and Drug Administration follows a selective plan in which it gives first attention to those cases where there have been complaints or past trouble or in which the public health is at stake. Next in importance come those cases involving adulteration of goods without an urgent health menace. Then come the cases which the Administration lists as "economic cheats," such as deceptive packages and foods mislabeled as to quantity, quality, etc. In cases where violations are found, the authorities may give a warning without prosecution, they may seize the offending goods, or they may prosecute the person or firm responsible for violation of the act. If warning is not considered sufficient, it is obvious that the seizure of the offending goods is much more rapid than prosecution of the violator. Goods can be seized and removed from the market at once, saving consumers from possible serious danger; but a court case may take months. In practice the seizures are much more common than the prosecutions. Products that have been seized may be destroyed, reassigned to some other use, or reworked until they meet the requirements. In cases of repeated violation, the Federal district courts may issue restraining orders denying the channels of interstate commerce to the offending articles.

The Federal Security Administrator publishes from time to time reports known as "Notices of Judgment under the Federal Food, Drug, and Cosmetic Act," which summarize the cases that have been acted on, showing what violations have been found and what action was taken on them. The annual reports of the Food and Drug Administration summarize the work of the year and the most important new developments. Table 8 • 1 summarizes the work for 1944.

8·8. Federal Trade Commission. Although the Federal Trade Commission was established in 1914, court interpretations of the statute that set it up were such that it could not be called a "consumer" agency until 1938. Prior to that time, it had been held that the Commission was empowered to regulate or prohibit unfair practices in commerce only if they affected other businessmen. If a business firm carried on activities that were harmful to consumers, the Commission was powerless to act unless the activities could be shown to hurt business competitors. But in 1938 the Wheeler-Lea Act remedied this difficulty, and the Commission is now enabled to carry on activities that are directly in the field of consumer protection. Even today many of the activities are more nearly in the field of production than in

Table 8·1. Summary of Work of the Food and Drug Administration, Fiscal Year, 1944

	Foods	Drugs and devices	Cosmetics	Caustic poisons	Total
Samples collected	30,147	11,782	436	30	42,395
Criminal prosecutions instituted:					
Violative samples	661	334	7	2	1,004
Actions	182	100	2	2	286
Seizures accomplished:					
Violative samples	2,390	586	48	2	3,026
Actions	1,739	380	22	2	2,143
Import shipments:					
Inspected	12,888	2,265	72	0	15,225
Refused entry	3,722	497	58	0	4,277

the field of consumption, and we shall have to give further attention to the Federal Trade Commission when we discuss, in Chap. 25, the problems of regulation of monopoly. For the time being, we shall confine our remarks to those aspects of the Commission's work which have to do with consumers.

The Federal Trade Commission administers the Federal Trade Commission Act, which prohibits unfair competition and unfair and deceptive acts or practices in commerce; Sec. 2 of the Clayton Act, prohibiting price discrimination and various other sorts of discrimination; the Webb-Pomerene Act concerning monopolistic practices in foreign trade; and the Wool Products Labeling Act of 1939, which is intended to prevent the presence of unrevealed substitutes and mixtures in wool products.

Although the Commission may initiate studies and investigations itself, most of its cases originate through complaints by consumers or competitors. The usual procedure is for the Commission to notify the party complained

against, tell him what the charges are, and ask him for such information as he may care to furnish. Often competitors and consumers are interviewed. The examining attorney then submits a summary of his findings and his recommendations to the Commission. Usually the recommendation is for one of the following courses of action:

1. Closing the case for lack of evidence or because no law has been violated.

2. Disposition of the case by signed agreement of the person complained against that he will cease and desist from the practices in question.

3. Issuance of a formal complaint.

If a formal complaint is issued, the respondent (the person complained against) is given 20 days to file an answer admitting or denying each allegation. If the case is contested, a hearing is held, after which the trial examiner submits to the Commission a report of the evidence. The Commission then studies the case and reaches a decision, sustaining the charges of the complaint, dismissing the complaint, or closing the case without prejudice.

In those cases where the complaint is sustained, the Commission issues what are called "Cease and Desist Orders." As their name implies, these are orders directing the offender to cease and to desist from the acts that have been contrary to law. Such orders may be appealed to the United States Circuit Court of Appeals, but become final and binding unless appealed in 60 days.[1]

Under the amendments to the Federal Trade Commission Act made in 1938 as part of the Wheeler-Lea Act, the Commission is given the duty of preventing false and misleading advertising of food, drugs, curative devices, and cosmetics. In such cases the Commission may proceed in the usual way, which we have just described, or, if the public is seriously endangered, the Commission can apply to the U. S. District Court for a temporary injunction, prohibiting the dissemination of such false advertisements until the Commission has had time to take action in the usual way.

From the start of its work in 1915 to the end of the fiscal year in June, 1944, the Commission has instituted 29,036 inquiries, had issued 5185 complaints, had issued 3606 orders to cease and desist, and had had 348 cases appealed to the courts. Of these appealed cases, 163 were decided in favor of the Commission, 96 for others, and 56 were withdrawn. Many of these cases, of course, were "production" cases rather than "consumption" cases, since, as explained earlier, the Commission was not in any real sense a consumer agency throughout the first 23 years of its existence.

[1] Under the Clayton Act such an order does not become final unless affirmed by the court on petition of either the Commission or the respondent. In the cases of the other laws administered by the Commission, the 60-day provision mentioned above is applicable.

Aside from complaints lodged by consumers, the most important consumer work now done by the Federal Trade Commission is in the fields of labeling and advertising. In the former the Commission attempts to aid intelligent purchasing and to prevent deception by informing consumers what they are to receive for their money, and to protect honest businessmen from dishonest competitors. The work so far has been largely in the textile field, but it has been expanding slowly into other territories as well. In the field of advertising the Commission surveys magazines, newspapers, mail-order catalogues, almanacs, and radio broadcasts. From these surveys it gets data for many of its cases and, in addition, discovers whether or not advertisers are complying with past orders. When false or misleading advertising is discovered, the offending concern is given an opportunity to sign an agreement to cease and desist. Most cases are disposed of in this manner.

8·9. Other Federal Agencies. We have described the work of the Food and Drug Administration and of the Federal Trade Commission in some detail. There are many other Federal Agencies that carry on some activity in the consumer field. The Bureau of Human Nutrition and Home Economics in the U. S. Department of Agriculture has done a great deal of work in the field of food and diets and has also made studies and released information in the field of clothing. The Post Office Department, through its Fraud Section, attempts to protect consumers against false and fraudulent advertising and labeling sent through the mails. The Federal Alcohol Tax Unit in the Treasury Department has jurisdiction over the labeling and advertising of liquors; the Securities and Exchange Commission, over the advertising of corporate securities. Both the Department of Agriculture and the Department of Commerce have been active in setting up grades and standards for commodities and in inspecting goods to see that they conform to those standards. Many of these agencies are primarily interested in producer problems, yet they all carry on important activities that assist consumers in getting better goods for their money or protect them against adulterated or harmful commodities.

8·10. Advertising. In the preceding sections we have been describing activities intended to benefit consumers. We have seen that various agencies attempt to influence the character of the goods available for consumers, or to teach consumers what sorts of goods they should want. Now we turn to other activities, also directed at influencing consumer wants, but from a selfish rather than from an altruistic point of view. The advertiser wants to change demand, not particularly for the benefit of the buyer, but for his own advantage. This does not mean, to be sure, that the buyer gets no advantage from advertising, but merely that whatever advantage he gets is incidental and fortuitous.

Selling activities of all sorts engage large numbers of people and involve the expenditure of tremendous sums of money. Studies by the Federal Trade Commission indicated that, in 1940, the advertising expense in 91 different industries varied from 0.06 cent per dollar of sales to 13.94 cents per dollar of sales. The median advertising cost was 1.1 cents per dollar of sales. Many of the 91 industries included in the study were selling industrial rather than consumer goods, but inspection of the figures shows that generally the heavy advertising expenditures were on consumer goods. This may be more evident if we select those industries which had the highest proportional advertising expense. If we take only those industries that spent 5 cents or more on advertising for every dollar of sales, we find that they included cigars, tobacco products, malt beverages, distilled liquors, soaps and cooking fats, cigarettes, cereal preparations, and drugs and medicines. In contrast, the industry with the lowest expenditure, 0.06 cent per dollar of sales, was the shipbuilding industry.

The importance of the advertising business will be more evident, perhaps, if we describe it in totals rather than in percentages. It is estimated[1] that the total expenditure for advertising in the United States was 2.60 billion dollars in 1929, 1.84 billion in 1942, 2.13 billion in 1943, and 2.27 billion in 1944.

This means that such expenditures amounted to perhaps 1.5 per cent of the national income. If these figures are even approximately correct (and, of course, they are estimates), it means that consumers are spending, on the average, about 1.5 per cent of their incomes for advertising. It should be emphasized that this does not include all selling expense, since personal salesmanship is supposed to be considerably more expensive even than advertising.

The expenditures for advertising are estimated[2] to have been divided as follows:

	Per Cent
Newspapers	28
Radio	18
Direct mail	13
Magazines	13
Trade papers	4
Outdoor advertising	3
Farm papers	1
Miscellaneous	20

The large miscellaneous classification consisted of such things as the administrative costs of advertising agencies, motion-picture advertising, window display advertising and novelties.

[1] Estimates by L. D. H. Weld, in *Printers' Ink*, Mar. 16, 1945, pp. 19-20.

[2] WELD, *op. cit.*

8 • 11. Effect of Advertising on Consumers. When one begins to discuss the effect of advertising on consumers, he finds himself at once in a field that is the subject of great dispute. Some of its critics see nothing but evil in modern advertising; some of its apologists see it as an almost holy calling, beyond ordinary criticism. We might well start our appraisal of the subject by repeating that advertising is intended to benefit the advertiser and that any benefit to the consumer is incidental. Those who talk of advertising as a great education medium, like those who criticize it for its failure in the educational field, fail to comprehend its purpose. Advertising is only incidentally educational and informative. Its major purpose is to change the demand for particular products; specifically, to move the demand curve upward and toward the right (see Sec. 6 • 12) and to decrease the elasticity of the demand. We have already noted the fact (Sec. 7 • 6) that demand tends to be very elastic when we are dealing with narrow classes of commodities. The demand for beef is more elastic than the demand for meat; the demand for sirloin steak is more elastic than the demand for beef, etc. Similarly, and for exactly the same reasons, the demand for the product of a particular producer tends to be very elastic if there are competing products. If there are 20 different manufacturers of automobiles and if the products are even roughly alike, the demand for any particular make or brand will be very elastic. Rising prices for that one brand will influence people to shift in great numbers to the competing brands. Such a shift is, of course, disadvantageous to the producer. He would prefer to be able to raise the price and still retain all, or almost all, of his trade. He would prefer to have the demand inelastic. He would prefer to get the consumer in a state of mind where he would "accept no substitutes." Thus the advertiser is interested both in getting people to use the product who never used it before, and in getting those who use the product to insist on his particular brand. Thus we say that advertising attempts to change demand in two definite ways: to shift the curve upward to the right and to decrease its elasticity.

Some cynic has said that advertising can be defined as a picture of a pretty girl eating something, wearing something, doing something, or enjoying something that the advertiser wants other people to eat, wear, do, or enjoy. If such a picture does raise the demand curve or if it does decrease the elasticity, then such advertising is successful in the sense that it accomplishes its purpose. On the other hand, advertising may be educational and carry a message that is of real advantage to the consumer. It may tell him where he can get goods of higher grade at lower prices. Again, if it raises the demand curve or decreases the elasticity of demand, the advertising has been successful in terms of its own purposes. The advertiser is interested in the effect on the demand curve; but we must be interested also in the effect on the public.

The ignorance of consumers, about which we spoke in Sec. 8 • 1, con-

stitutes at once the main need for advertising and the main danger of advertising. Advertising does, to be sure, furnish information about goods that are for sale, about qualities, prices, and sales outlets. Such information is absolutely necessary in a society as specialized as ours. The complete lack of advertising would bring a breakdown of our system. Most consumers are not personally acquainted with sources of supply of the things that they want. Moreover, advertising introduces new commodities to consumers—things that they did not want because they had never heard of them and never thought of them. In a society such as ours, advertising and the allied activities of personal selling convey information about known goods and introduce new goods to the market. Yet one needs only a casual acquaintance with advertising, as it is carried on, to realize that the information conveyed is by no means impartial, disinterested, and unbiased. Everything advertised is "the best of its kind." One never sees, "Our product is fairly good," or, "Our product is third in its field." Advertising runs to superlatives. It stresses inconsequentials. All this it is able to do because consumers do not know inconsequentials from consequentials.

Some advertising is downright false, but the various agencies mentioned earlier in this chapter, as well as state and private agencies, try to eliminate actual untruths. Yet consumer ignorance plays into the hands of the unscrupulous advertiser by making it possible for him to confine himself to statements that are literally true, yet misleading to the consumer. Thus, we have concerns advertising "health lamps," stressing the fact that their lamps are rich sources of infrared rays, necessary for the support of life. The statements are entirely correct; yet the advertiser is aware that most consumers will not know the difference between infrared rays and ultraviolet rays. They will not realize that a hot-water bottle or a furnace is a good source of infrared rays, which, after all, are nothing but heat rays. Having heard something about ultraviolet rays and their uses, consumers buy the infrared lamp under a misapprehension. Every new discovery that is well publicized, such as vitamins or penicillin, opens up new fields for the exploitation of consumers who do not know enough to evaluate the advertising statements that they read.

Let us repeat: It is the ignorance of consumers that makes advertising indispensable; it is that same ignorance that makes it possible for some advertisers to profit by statements that mislead consumers. It is easy to agree with both sides in the controversy over the value of advertising: that we cannot get along without it, yet that much of our present advertising leads to ill-considered buying. Since there seems little likelihood that we shall overcome consumer ignorance completely, it is likely that advertising will continue and that there will continue to be abuses in the field. It would not be fair to leave this subject without repeating that the situation has been much

improved by the efforts of various agencies in their attempt to get truth in advertising, and that every effort that gives consumers better education helps them to protect themselves against deception. But, unless consumption decisions are to be taken out of the hands of consumers entirely, we may expect that they will continue to make mistakes and that other people will be glad to profit by them.

Questions

1. Do you think that consumers are, in general, more or less expert in judging the things that they buy than they were a century ago? What factors have been responsible for the change?

2. List several examples of present-day sumptuary legislation.

3. "The scientist's appraisal of any consumption good is only a partial appraisal." Explain and illustrate.

4. If the Food and Drug Administration were to find someone selling adulterated food, what could they do about it?

5. What is a cease and desist order? If such an order is issued against our concern, do we have to obey it? Explain.

6. Collect from current magazines and newspapers five advertisements that seem to you to be informative, and five that seem misleading. Explain the nature of the difficulty in each of the latter.

7. Since advertising is sometimes untruthful and even more often misleading, it might be advantageous to set up some sort of government censorship, under which a committee of experts would study all proposed advertising copy and decide which could be published and which not. What would be the advantages of such a system? The disadvantages?

8. Advertising expenditures are said to run to around 2 billion dollars a year in the United States. For many products, the total selling cost (including advertising and personal selling) is said to amount to half of the consumer expenditure. This is far larger than it was a century ago. Why?

PART THREE. Production

9. THE ORGANIZATION OF BUSINESS

9·1. Diversity of Business. In this chapter and in the few chapters that immediately follow, we shall turn our attention to the problems of production—those facing business firms. Before we embark on a discussion of them, however, we shall examine the business firm itself, to see what it is like and how it operates.

The reader will realize from his own experience that such problems differ greatly. At the one extreme we have concerns that are owned and operated entirely by one man, involving an investment of but a few hundred dollars of capital and with an output of a few thousand dollars' worth of product per year. At the other extreme we have gigantic business firms, such as the American Telephone and Telegraph Company, which has assets of around 3 or 4 billion dollars and is the joint property of well over half a million owners. The methods that would be well adapted for the control and operation of a local automobile repair shop would hardly do for the Standard Oil Company of New Jersey or for the General Motors Corporation. The differences in the problems faced by different business firms give rise to variety in the structure and control of those firms. There is no one best form of business organization.

9·2. Individual Proprietorship. Originally a man and his business were one and the same thing. No one knows when man first went into business for himself; the individual proprietorship is older than recorded history. Today in many kinds of economic activity, for example, in agriculture, the individual business is still the major type of organization. As compared with those types of business which have been developed more recently to meet the complex problems of modern business, the individual proprietorship has many and important advantages. It is easy to start and is subject to relatively little regulation by the state. It is simple to understand and to operate and requires relatively little capital. It has unified control, and the profits, if there are any, go directly to the owner. In fact, legally the owner *is* the business. The fact that men have many sorts of activities does not lead

them to think of themselves as many different men. John Jones does not think of himself at one moment as a Presbyterian, at another as a scoutmaster, at another as a father, and at another as a butcher. He is all these things at once, and the law does not separate his activities as a butcher from his activities in the other directions. His business activities are part and parcel of his other activities, and this fact gives rise to some of the most important peculiarities of the individual proprietorship.

Especially, the inseparability of the man and the business leads to some of the major disadvantages of the proprietorship. If the proprietor dies, his business comes to an end. The debts must be paid off and the assets distributed either in accordance with the will of the deceased or in accordance with the law in case he left no will. If the proprietor loses his mind or if he goes bankrupt, the business must also be wound up. When the proprietor personally becomes unable to carry on, the business stops. Perhaps his son or someone else will come into possession of the property and start a new business; but the old business is inseparable from its owner and stops when he does. The individual proprietorship is thus unstable. Many unexpected events may occur, any one of which will terminate the business. At the best, a proprietorship is relatively short-lived, since the length of a man's business life is relatively short. Moreover the individual proprietorship is limited in the amount of capital that it can command. The business has no more capital than that which the owner can himself furnish or which he can personally borrow. This will seldom run to more than a few thousand dollars, which is sufficient for some types of business but far too little for many others.

Another disadvantage of the individual proprietorship, which is associated with the fact that the man and the business are inseparable, is that all his assets are available as security for his creditors. Our butcher may own his home, an automobile, a radio, and a lot at the seashore. If he gets into business difficulties at his market and is unable to pay his bills, his creditors can take over his store, its fixtures, and his other business assets, but they can also take over his house, car, radio, and lot at the shore. It makes no difference that these latter assets were not used in the business. They are the butcher's assets and must be available to his creditors. We say that the individual proprietorship is characterized by *unlimited liability,* meaning that the proprietor runs the risk of losing everything he owns in case the business is unsuccessful.[1] The law does not think of his business assets as

[1] The term "unlimited liability" is very slightly exaggerated. Actually there are some few assets of an individual proprietor that cannot be taken by his creditors. These excluded assets differ in different states according to the local laws, but, for example, the creditors must leave the proprietor a suit of clothes, his toothbrush, the family Bible, and some such enumerated items. The exceptions are, however, few in number, and we are not misleading to any serious extent by calling the liability "unlimited."

distinct from his other assets because the law does not distinguish the butcher from the man.

In the United States today, the individual proprietorship is by far the commonest type of business organization in the field of agriculture. It is also common in the retail merchandising field, and there are many small manufacturing businesses that are so organized. But, because of the limitation of capital, these businesses are relatively small and seem fated to remain so.

9 · 3. Partnerships. One of the early developments of business organization that sought to overcome one of the disadvantages of the individual proprietorship came when two or more men pooled their capital in what we call a "partnership." Aside from the fact that the partnership can combine the assets and credit of two or more men, it is very much like the individual proprietorship. It is, to be sure, a very little harder to start than the individual business, because it is necessary for the partners to agree among themselves on the amounts of their contributions, on the distribution of the profits, etc. Such a partnership agreement is usually, although not always, a formal, written document. Like the proprietorship, it is subject to relatively little special government regulation or taxation, and it may go into any sort of legal business on which the partners agree.

Again, the law looks upon the partners and their business as inseparable. If Jones, Smith, and Brown form a partnership to run a garage, the law sees Jones, Smith, and Brown rather than their garage. The garage building, to be sure, is there for everyone to see; but the law sees it, not as a business, but as the property of these men, personally. This brings us many of the same disadvantages that we found in the individual proprietorship, but in magnified form. For example, if any one of the three owners dies, loses his mind, or becomes bankrupt, the partnership is ended. The business is thus three times as unstable as the individual proprietorship, because there are three lives and personalities to continue if the business is to continue. It is easy to see how this fact limits the possible size of the partnership. There are many businesses in the United States with over a thousand owners; some with several hundred thousand. If it were necessary to wind up these businesses every time one of the owners died, they would be constantly in the courts. The partnership is always limited to a relatively small number of persons for this reason. But worse yet, since the partners are the business, each of them is liable for the debts of the business. If one of the partners in our garage uses bad judgment and invests too much in equipment or materials and if we are then unable to pay the bills, the creditors may take the property of all or any one of us. Perhaps it was Smith's bad judgment that got us into trouble. That makes no difference. Any partner may act for the partnership, and the others will be fully responsible. Smith got us in trouble,

but the creditors can take Brown's house, his automobile, his stocks and bonds, and whatever else he owns with minor exceptions noted under the individual proprietorship. The creditors can collect the entire amount owed from Brown, or they may decide to collect from Brown and Smith, or possibly they will try to collect from all three partners. The important fact is that each and every partner has unlimited liability for the debts of the partnership. This again limits the number of partners that we can get together. No one can afford to go into partnership with other men unless he knows them well, trusts them implicitly, and watches them carefully. It would not be likely that there would be 1000 individuals whom I would trust, nor even 100. Partnerships are almost always relatively small because of the dangers of unlimited liability and because of the instability that comes with numbers; yet they do make it possible to get somewhat more capital than the typical individual proprietor can command.

The personal element in the partnership—the importance of individual trust and confidence—is so great that a man has the right to select his own partners. If Jones, Smith, and Brown are partners, Brown cannot sell out his interest to Williams, since Jones and Smith cannot be forced to assume the partnership relation with Williams. The partnership is characterized by what the lawyers call the right of *delectus personae*. This is one's right, in selecting a person for a position of trust and confidence, to insist upon someone satisfactory to himself. If Brown wishes to withdraw from the partnership, the partnership must be dissolved. A new one could now be formed by agreement between Jones, Smith, and Williams; but it will be a new partnership and not the old one. The withdrawal of any partner terminates the partnership. Transfer of ownership is not only difficult; it is impossible.

Despite their handicaps, partnerships are still very common in some fields, such as banking and the practice of law, or in the retail trade.

9·4. Limited Partnerships. One cannot afford to run the risk of unlimited liability in a partnership unless he is willing to take an active part in the management of the business, keeping himself constantly posted on the condition of the business and on the decisions that are being made. Yet there are many people who would like to invest surplus capital in a business without taking an active part in management if they could escape the threat of unlimited liability. To meet this situation there was developed a variant of the partnership which we call the "limited partnership." As in the general partnership, it is a joint venture in which two or more owners pool their capital in a common purpose. But in the limited partnership one or more of the partners takes no active part in management. He invests his money but leaves the management to the general partners or active partners. Every limited partnership has and must have two kinds of partners: the limited partners and the general partners. There may be one or more limited part-

ners and one or more general partners. But we do not have a limited partnership unless both kinds are included.

The general partners are in every way like those of the ordinary partnership. They are active in management and have unlimited personal liability for the debts of the partnership. The limited partners (sometimes called "silent" or "sleeping" partners) are merely investors in the business. They invest funds and share in the profits if there are any; but they have no voice in the management of the business, and their liability for its debts is limited to the funds that they have contributed. It is important for the student to realize that limited liability does not mean "no liability." The limited partner who has invested $5000 in a clothing store may lose that amount if the business gets into financial difficulty, but the creditors cannot take his home or his other assets.

The limited partnership differs in one other important respect from the ordinary general partnership. The death, bankruptcy, or withdrawal of a limited partner does not terminate the business. This is natural, since the limited partner has contributed his money rather than his ability or personality. He is not active in management, and therefore the other partners do not have to feel that they are liable for his acts. We begin, in the limited partnership, to get a first small degree of separation of the man from his business. The limited partner may die, yet the business may still continue. This was an important precursor of other developments that were soon to appear in the business field.

9·5. Joint Stock Companies. As modern technological developments made it profitable to operate businesses on a larger and larger scale, with more and more capital and employees, it became necessary to associate more and more people together as owners of the business. But the partnership form was unstable, being broken up on the death or incapacity of any partner. It was necessary to find some way by which the ownership of a business could be transferred from one person to another without breaking up the business itself. This was done by the development of what we call the "joint stock company." In most ways the joint stock company was[1] like the partnership. It was organized by an agreement or contract among its owners, called the "articles of association." The members were individually liable without limit for the debts of the business. The owners could carry on any legal type of business without restriction, and there was relatively little special regulation or taxation by the government directed at the joint stock company. In all these respects it was like the partnership.

[1] The past tense is used because, although joint stock companies do exist today, they are relatively unimportant, having been replaced largely by the corporation. Historically they were important in developing new business concepts, and they are included here as a transitional type between the partnership and the corporation rather than for any present importance.

There were several important points of difference. Ownership in the joint stock company was represented by shares of stock, which could be freely bought, sold, and transferred. Thus there was no right of *delectus personae* and no limitation on the transfer of ownership. The owners no longer took an active part in the management of the business but elected a board of managers or a board of directors who had full power to run the business. The business could continue indefinitely, since when an owner of stock died or became incapacitated his stock continued to exist and passed into the hands of someone else.

We can think of the joint stock company, then, as being in most respects similar to the partnership, differing mainly in transferability of ownership and in method of management. These two differences were of tremendous importance, and the joint stock company grew quickly to a position of importance in industry. It was soon displaced, however, by the corporation, which had these same advantages and many more. Today the joint stock company is of minor importance in the United States.

9·6. Corporations. As time passed businessmen tried to combine the advantages of the various sorts of business organization and to eliminate the disadvantages. They sought a type of organization which would combine the ease of transferring ownership of the joint stock company with the limited liability of the limited partnership and which would make possible the accumulation of capital in the large amounts necessary for machine production under the factory system. They hit upon a fairy tale which we call the "corporation," which separates sharply the owners of the business from the business itself. The law pretends that the corporation is a person. If five men come together and form a partnership, they are still five men. But if five men come together and form the Gurleyville Rapid Transit Corporation, the law pretends that there are now six people: the original five plus the corporation. In the famous Dartmouth College case, Chief Justice Marshall of the U. S. Supreme Court said that a corporation is "an artificial being, invisible, intangible, and existing only in the contemplation of the law." In plain English this means that the corporation cannot be seen by anyone but a lawyer, while lawyers look at the corporation and see a person. And this fairy tale, which makes believe that the corporation is somebody, turns out to be the most important single thing in determining the rights and powers and peculiarities of the corporation as a business agency. The corporation is a legal entity, separate and distinct from the people who make it up. The law insists that the General Electric Company is a person—a real person with most of the rights of people.

The legal fiction that the corporation is a person gives the corporation one decided advantage at the outset; namely, the owners have limited liability for its debts. If the Gurleyville Rapid Transit Corporation is poorly managed

and loses money, the creditors cannot sue the stockholders. Why should the stockholders be liable for the debts of someone else? The debts are the debts of the person whom the lawyers call the Gurleyville Rapid Transit Company. He is, or should be, responsible for his own debts. Thus the owners of the corporation enjoy limited liability. They may lose the money that they have invested, up to the full par value of their stock, but beyond that the creditors cannot go.[1] We see again that limited liability does not mean no liability at all. It means that the maximum liability is known in advance—that it is limited to some predetermined amount.

But the fiction of personality brings other advantages. For example, the corporation does not cease to exist because one of its owners dies or becomes bankrupt. The corporation is a person, and the fact that a stockholder dies does not make the corporation die. The corporation is fictitious anyway. We pretended that it was there, and we can still assume that it is there even after all the original owners are dead. Thus corporations get a stability of existence which is of prime importance, especially if we are to have thousands of owners, with some dying almost every day. Some corporations are established for limited periods of time, such as for 50 or 100 years. Some corporations exist *in perpetuity.* At any rate, they are not subject to dissolution from the chance and unexpected happenings that are so annoying in the case of the partnership.

Ownership of the corporation is evidenced by shares of stock.[2] In this way the ownership may be broken down into small pieces which may be readily transferred. We may have a billion-dollar corporation, yet I may be able to buy $10 or $100 worth of ownership in it by buying a share of its stock. This divisibility of ownership makes it possible to bring together the small savings of thousands of different people until we have a business with far more capital than any one man or any few men could accumulate. Since any owner is free to sell his stock at any time, there is very easy transfer of ownership. The corporation is not dissolved by the withdrawal of a member, nor by his death. The principle of *delectus personae,* which is fundamental in the partnership, has disappeared entirely in the corporation.

In these ways the corporation is a great advance over its predecessors. But it has its disadvantages which offset in some measure the advantages just enumerated. In the first place, if the corporation is legally a person, the law must bring him into existence. Any man may start in business as an individual proprietor without legal formality. Two or more men may,

[1] Some corporations, especially financial corporations such as banks and insurance companies, have had "double liability," under which a man who had invested $1000 could lose that and still be assessed another $1000 to pay the debts of the corporation. Even here the liability was limited. Double liability has recently tended to disappear.

[2] See Sec. 9•9 for a description of corporate stock.

by agreement among themselves, start a partnership or a joint stock company. But we cannot start a corporation by ourselves. It is a legal concept[1] and comes into existence only by the intervention of the law. Men wishing to start a corporation have to comply with the provisions of the law of the state where the corporation is to be formed,[2] after which they receive a *charter,* which is the document that actually brings the corporation into existence. The state is the father, the mother, and the attending physician; the charter is the birth certificate. The charter also specifies what sorts of activities the corporation may carry on and, unlike the other types of business that we have described, the corporation is then limited in its activities to those things specified there. If the Gurleyville Rapid Transit Company is chartered to operate subways in Gurleyville and if the managers see an opportunity to make some money by starting a ferry service on the near-by Fenton River, they cannot take on the new venture unless their charter makes provision for it. If a corporation performs any *ultra vires* act (that is, any activity not within the scope of those allowed in its charter), it may be dissolved or held to be a partnership with all the dangers of liability inherent therein.

Not only is the corporation harder to form than other kinds of business and limited in the sphere of its activity, but also it has the disadvantage that it is often subject to various sorts of special regulation and taxation by the government. The regulations may take the form of reports to government agencies giving detailed information about the business and its operations, or they may require various notices to members, the maintenance of books and accounts in accordance with specified rules, or the following of specified procedure for the declaration of dividends. The taxes often include an organization tax which must be paid on issuance of the charter, annual franchise taxes which may be payable in every state in which the corporation does business, property taxes and income taxes, stock transfer taxes, recording fees, etc. Thus corporations often have more onerous legal obligations than do any of the other types of business organization.

9·7. Comparison of Business Organizations. Before we describe further the operation of the corporation, let us review briefly and in outline form the various types of organization that we have described. For the sake of comparability, we shall consider for each type the following things:

a. Ease of formation.
b. Liability of the owner.
c. Ease of getting capital.
d. Stability and durability.

[1] The corporation is another good example of what we called "social institutions" on p. 18.

[2] In unusual circumstances corporations are formed by the Federal government, but practically all business corporations are established by the individual states.

e. Amount of special regulation.
f. Limitation on activities.
g. Ease of transferring ownership.

On this basis we find the following:

1. The individual proprietorship:
   a. Formed without formality or complication.
   b. Owner has unlimited liability.
   c. Capital limited to resources and credit of the single owner. Usually small businesses.
   d. Dissolved by death or incapacity of owner. Relatively short-lived.
   e. Relatively little legal restriction.
   f. Can do any legal business or combination of legal businesses without restriction.
   g. If the proprietor sells out, the original business is not transferred. A new business is started. A buyer must be found who will take over all the assets and liabilities.
2. The partnership:
   a. Formed by agreement between partners. Relatively easy.
   b. Each owner has unlimited liability.
   c. Capital limited to resources and credit of the partners. Usually relatively small businesses.
   d. Very unstable. Dissolved by death, incapacity, or withdrawal of any partner. Relatively short-lived.
   e. Relatively little legal restriction.
   f. Can do any legal business or combination of legal businesses on which partners can agree.
   g. Ownership cannot be transferred. If partner withdraws, partnership is dissolved.
3. The limited partnership:
   a. Formed by agreement between partners. Relatively simple.
   b. Always two classes of owners. General partner or partners have unlimited liability. Limited partner or partners have limited liability.
   c. Capital limited to resources and credit of the partners. Limitation of liability makes it somewhat easier to get contributions than in the case of the general partnership, but these firms are still usually small.
   d. Unstable. Dissolved by death, incapacity, or withdrawal of any general partner. Relatively short-lived.
   e. Relatively little special legal restriction, although the introduction of limitation of liability did require the setting up of some new restrictions.
   f. Can do any legal business or combination of legal businesses on which partners can agree.

g. Ownership of general partners cannot be transferred. If general partner withdraws, the firm is dissolved.

4. The joint stock company:
   a. Formed by written articles of association, but without complicated legal restriction.
   b. All owners have unlimited liability.
   c. Capital rather easy to get because ownership is represented by small, transferrable shares.
   d. Stable and durable. Can last indefinitely. Not broken up by loss of individual members.
   e. Amount of regulation varies in different states but is generally intermediate between the foregoing types or organization and the corporation.
   f. Activities limited only by the articles of association.
   g. Ownership easily transferred by sale of stock.
5. The corporation:
   a. Difficult to form. Much legal red tape. Requires charter from the state.
   b. Owners have limited liability.
   c. Very easy to get capital, since ownership is broken up into small, transferrable shares, and owners do not have unlimited liability.
   d. Very stable and durable. Can last in perpetuity. Not dissolved by loss of individual members.
   e. Many onerous special regulations and taxes.
   f. Activities limited to those enumerated in the charter.
   g. Transfer of ownership very easy by sale of stock.

9·8. Relative Importance of the Corporation. The relative importance of the various types of ownership organization varies in different fields of endeavor. As has already been noted, the individual proprietorship is still common in farming and the retail trade, and partnerships still have a limited field of popularity and usefulness. But in most industries the corporate form is the typical one, and the vast and overwhelming majority of the output comes from corporations. Thus, in manufacturing, 32 per cent of the firms are individual proprietorships and 15 per cent are partnerships,[1] yet between them they hire but 11 per cent of the employees and make but 7 per cent of the output. Similarly in retail merchandising, 77 per cent of the units are individual proprietorships, but they hire only 35 per cent of the workers and do but 39 per cent of the business. The partnerships make up 11 per cent of the retail outlets, yet they hire but 10 per cent of the workers and do but 12 per cent of the business. The figures are perhaps most easily grasped if we separate the corporate businesses from the others and show for each

[1] These figures and those immediately following are for 1939, taken from the 1940 census.

type of trade the importance of corporations based on three measures: the percentage of all the business units that are corporations, the percentage of all the workers in the industry who work for corporations, and the percentage of the output of the industry that is produced by corporations. The data appear in Table 9·1.

Table 9·1. Percentage Importance of Corporations in Various Fields of Production in the United States, 1939*

	Percentage of all business units	Percentage of all employees	Percentage of all output
Manufacturing	52	89	93
Retailing	12	53	47
Mineral industries	45	87	90
Construction	8	44	46

* Data are derived from the 1940 census.

Another authority[1] has estimated the relative importance of the corporation in various industries on the basis of the percentage of the national income produced by the industry that comes from corporations. His estimates are that the importance of the corporation is 96 per cent in mining and quarrying, 92 in manufacturing, 86 in transportation and other public utilities, 63 in

Table 9·2. Proportion of National Income Produced and Volume of Business Done by Corporations and Other Forms of Enterprise, United States, 1937

Industry	Percentage of national income produced	Percentage of business done by	
		Corporations	Others
Manufacturing	24.0	92	8
Government, including work relief wages	13.5	58	42
Trade	12.5	58	42
Services	11.9	30	70
Finance	9.3	84	16
Agriculture	8.9	7	93
Transportation	7.3	89	11
Mining	2.1	96	4
Contract construction	2.1	36	64
Electric light and power and manufacturing gas	1.6	100	0
Communications	1.3	100	0
Miscellaneous	4.2	33	67
Total	98.7	60–65	35–40

[1] BERNHEIM, ALFRED, "Big Business, Its Growth and Place," pp. 17-18, The Twentieth Century Fund, New York, 1937.

trade, 56 in finance (including banking, insurance, real estate, etc.), 33 in the service industries (including the professions, amusement, hotels, etc.), 33 in the construction industry, and 6 in agriculture. Still a third authority [1] gives the data of Table 9 • 2, showing the proportion of the output of various industries produced by corporations, and the proportion of the entire national income produced by each of these industries.

Because these figures are for different years and are computed by somewhat different methods, they do not agree exactly; yet they show vividly the importance of the corporation in the economic life of the United States and indicate in which parts of our productive system it has been most prevalent.[2] Although the corporation itself is fairly old, its wholesale adoption by business has been recent. A century and a half ago there were but 225 corporations in the United States, and fewer than 20 of them were manufacturing or mercantile concerns.[3] It was not until after the Civil War that corporations took on a position of importance. Today their role in economic life is so outstanding that the student of economics finds it necessary to understand something of their structure and functioning. To these matters we now turn our attention.

9 • 9. Stocks. Stock and bonds together are called "corporate securities." A share of stock indicates ownership in a corporation. To whom does the Gurleyville Rapid Transit Company belong? To its stockholders. Originally the "stock" of a corporation referred to its materials, goods, and equipment, and the owners each possessed a share of this stock of goods. To evidence such ownership, the corporation issued certificates showing what part of the total stock of goods belonged to the holder of the certificate. These stock certificates have come to be called "shares of stock," and we shall use the term in this generally accepted sense; but it is well for the student to realize that the relationship between a share of stock and a stock certificate is really about the same as the relationship between an automobile and the receipt that one gets when he leaves his car in a parking lot. The stock certificate is a piece of paper that evidences ownership in actual property of the corporation.

The stock is more important to the corporation than the stockholder. When corporation elections are held, each share of stock gets one vote. When dividends are declared, each share of stock participates alike in the distribution. If the corporation is dissolved, each share of stock receives the same

[1] DEWEY, ANDERSON H., Taxation, Recovery, and Defense, p. 319, *Monograph* No. 20 of the Temporary National Economic Committee's Investigation of Concentration of Economic Power.

[2] See also Sec. 25 • 9.

[3] DEWEY, ANDERSON H., "Our California State Taxes," pp. 230*ff*, Stanford University Press, Stanford University, Calif., 1937.

proportion of the assets.[1] It is shares of stock, and not stockholders, that are treated alike. If you have 10 shares of stock and I have 2, you will have 10 votes to my 2, and you will receive five times as much as I do when dividends are declared.

Although the stockholders of a corporation are its owners, they are not its active managers and have no direct voice in the operation of the business. Their sole function in this regard is their right to vote for directors. The board of directors determines the general policies of the corporation, usually giving major attention to the financial policies. The directors usually do not have the time nor the training, however, to manage the technical details of the business, and for this purpose they appoint the major officers of the corporation—the president, the general manager, etc. These officers then appoint whatever subordinates seem desirable. If we trace our way back through these hierarchies of officials, we might start with the workman at his lathe, whose superior is a foreman, whose superior is a department superintendent, whose superior is the general manager, who was appointed by the president, who was appointed by the directors, who were elected by the stockholders. The stockholder would have no right to give orders to a workman at his lathe nor to any of the officials mentioned. His sole management function is to cast his vote or votes at the annual election of directors.

The directors, as has been said, appoint the officers of the corporation, determine its broad policies, and decide whether or not dividends are to be declared. This latter responsibility is one of great interest to the ordinary stockholder, for he usually has purchased his stock in the hope that the corporation will be profitable and that he will share in the profits in the form of dividends. Dividends may not be legally declared unless the corporation has made profits, either recently or at some time in the past; but, even if the business is profitable, the directors are under no obligation to distribute the profits in the form of dividends. They may decide that there is need for more money in the business, to expand the plant or to buy machinery. In such a case they may reinvest the profits in the business—"plow the profits back into the business," as the saying goes—rather than pay them out to the stockholders. The corporation makes no promise to pay dividends; it is under no obligation so to do. Stock is not a promise to pay, but an evidence of ownership. If the corporation never makes money, there will be no dividends; it is possible that there will be no dividends anyway. Yet, of course, most successful companies do pay them.

[1] A corporation may have several kinds or classes of stock outstanding. In such a case, the different kinds may differ in voting power, dividend rights, etc. These things we shall cover shortly.

Many corporations have but one kind of stock, which in such a case is always what we call "common stock." Other corporations have two or more kinds of stock, some of which is given certain preferences over the common stock and is therefore called "preferred stock." The commonest and most important preference is in the receiving of dividends and can most easily be explained by an example. Suppose that we have what is called a "6 per cent preferred stock." In such a case the management guarantees that it will not pay anything to the common stock in any year unless the preferred stock has been given 6 per cent of its par value.[1] Notice that there is no guarantee that anything will be paid on the preferred stock at all but, if the preferred stock gets nothing, the common stock gets nothing. If this preferred stock gets 3 or 4 per cent, the common stock gets nothing. Unless the preferred stock gets its full preference percentage, which we have taken to be 6 per cent, the common stock receives no dividends at all. To be sure, the preferred stock would usually get no more than 6 per cent under these circumstances. If the corporation was very profitable, the preferred stock might get 6 per cent and the common stock 26. But, in any distribution the preferred stock must be considered first. To offset this advantage, preferred stock often is not given any voting rights. Whether this is true in a particular case or not would depend on the contract contained in the stock certificate. Preferred stock does vote in some cases. But it is perhaps more usual for the voting power to be reserved entirely for the common stock.

It should be noted from the foregoing description that the decision on payment of dividends is made a year at a time. If the corporation "passes" its dividend (decides not to pay it) in any given year, it starts off fresh the next year, and again would have to pay 6 per cent on its 6 per cent preferred stock if it were to pay anything to the common stock. Sometimes, however, a special provision is written into the contract with the preferred stockholders stipulating that, if dividends are passed, the corporation must pay all these earlier unpaid dividends before anything can be paid on the common stock. Thus, if our corporation pays no dividends this year nor next, in the year following it will have to pay 18 per cent on the preferred stock before it can pay anything on the common—this year's 6 per cent and 6 per cent for each of the two previous years when dividends were skipped. A preferred stock that has this provision is called a "cumulative preferred stock," because the unpaid dividends accumulate as a claim against future earnings. On Jan. 1, 1945, there were 62 stock issues, out of all those listed on the New York Stock Exchange, which were in arrears on cumulative dividend provisions. The lengths of time and the total amount of unpaid back dividends are given in Table 9·3. We note, for example, that there

[1] The term "par value" is explained on p. 149.

was one issue that was in arrears for between 5 and 6 years and had accumulated arrears of about 3½ million dollars. Nothing can be paid on the common stock of this corporation until the entire amount of these unpaid back dividends has been paid on the preferred stock, since it is cumulative preferred stock.

In the cases so far described, the preferred stock may receive more than the common stock or it may receive less. If it is a 5 per cent preferred stock with par value of $100, it must receive $5 per share or the common stock will receive nothing. But the preferred might get $3 per share and the common nothing. The preferred might get $5 a share and the common

Table 9·3. Preferred Stock Dividends in Arrears, Jan. 1, 1945, New York Stock Exchange Issues*

Number of years in arrears	Number of issues	Total amount of arrears
Less than 1	4	$893,000
1–2	3	1,633,000
2–3	4	12,069,000
3–4	2	874,000
4–5	4	57,867,000
5–6	1	3,553,000
6–7	7	74,511,000
7–8	6	20,107,000
8–9	1	14,402,000
9–10	2	25,161,000
10–11	6	103,545,000
11–12	5	121,282,000
12 and over	17	548,099,000

* Data from *The Exchange*, vol. VI, no. 2, February, 1945, published by the New York Stock Exchange.

$2. The preferred might get $5 a share and the common $8. But some preferred stock has provisions such that it can receive more than the common stock but not less. In such a case the arrangement would be, if it were a 5 per cent preferred stock, that the corporation would have to pay $5 on the preferred stock or it could pay nothing on the common; but, if the common received more than $5 per share, the preferred would participate in such excess dividends equally with the common stock. If the common stock were to receive $8 a share, the preferred would have to receive $8 a share. Such preferred stock is called "participating preferred stock" because it participates with the common stock in any dividends above the preference rate. Participating preferred stock may receive more than the common stock but

cannot receive less. If we repeat the example from the beginning of this paragraph, with a 5 per cent participating preferred stock with par value of $100, each share must receive $5 or the common stock will receive nothing. The preferred might get $3 a share and the common nothing. The preferred might get $5 a share and the common $2. The preferred might get $8 a share and the common $8. But we could no longer have the situation, as with ordinary nonparticipating preferred stock, where the preferred would get $5 and the common $8. Participating preferred stock may get more than the common stock, and will get more if any dividends are paid and the common gets less than the preference percentage. But the common stock cannot receive more than the participating preferred, since this preferred stock does participate in dividend payments beyond the preference percentage.

Nor does this end the possibilities. We may, and sometimes do, have cumulative participating preferred stock. As one would expect from the name, this combines both the cumulative and the participating features already explained. For example, the Westinghouse Electric and Manufacturing Company has two kinds of stock outstanding: common stock with a par value of $50 a share, and 7 per cent cumulative participating preferred stock with a par value of $50 a share. Since 7 per cent of $50 is $3.50, this corporation must pay $3.50 per share on its preferred stock before it pays anything on common. It must also pay any unpaid dividends of earlier years (because the stock is cumulative) and, if the common stock is to receive more than $3.50 per share, the preferred must receive just as much more along with it (because the stock is participating). We can illustrate this from the dividend record of the company, part of which is shown in Table 9·4. Inspection of this table shows that, whenever the preferred stock received less than its preference amount of $3.50 (as in 1909), the common stock received nothing. This would be true of any preferred stock, whether cumulative, participating, or not. If we look back to 1908, we see that the dividend was first passed on the preferred stock. The following year, in 1909, it would have been necessary to pay $7 on the preferred if anything were to have been paid on the common, but only $2.625 was paid on the preferred. Hence nothing was paid on the common, and there were arrears of $4.375 on the preferred. In 1910, it would have been necessary to pay these arrears plus the $3.50 now due in 1910 if anything were to be paid on common. The actual payment on preferred was but $5.25, leaving arrears of $2.625, so that nothing could be paid on the common stock. In 1911, the company paid $6.125 on the preferred, which paid off all the arrears and also paid the regular $3.50 dividend for 1911. Although nothing was paid on the common stock, the arrears on the preferred stock were now all taken

care of and, since the company has paid at least $3.50 on its preferred stock each year since 1911, there have been no arrears since that time. In the period since 1911 the preferred stock has often received more than the common stock (as, for example, in 1932 to 1935), but in those cases the preferred stock received its full $3.50 so that it was possible to pay something on the common stock. In every case when the common stock received more than $3.50, the preferred stock received the same amount as the common stock. It is evident that the preferred stock described in Table 9·4 is par-

Table 9·4. Dividend Record of the Westinghouse Electric and Manufacturing Company on Common and on 7 per cent Cumulative Participating Preferred Stock, 1906-1943

Year	Dividend on		Year	Dividend on	
	Common	Preferred		Common	Preferred
1906	$5	$5	1925	$4	$4
1907	5	5	1926	4	4
1908	0	0	1927	4	4
1909	0	2.625	1928	4	4
1910	0	5.25	1929	4	4
1911	0	6.125	1930	5	5
1912	1	3.50	1931	3.875	3.875
1913	2	3.50	1932	0.875	3 50
1914	2	3.50	1933	0	3.50
1915	2.25	3.50	1934	0	3.50
1916	3	3.50	1935	1	3.50
1917	3.75	3.75	1936	5.50	5.50
1918	3.50	3.50	1937	6	6
1919	3.75	3 75	1938	2.50	3.50
1920	4	4	1939	3 50	3.50
1921	4	4	1940	4.75	4.75
1922	4	4	1941	5	5
1923	4	4	1942	4	4
1924	4	4	1943	4	4

ticipating, since it receives more than its preference amount whenever the common stock receives more. It is evident that it is cumulative, because of the fact that when its preference dividends were not paid in the period 1908-1909, the company made them up before it paid anything on the common stock.

Corporations sometimes issue more than one kind of preferred stock, such as first preferred stock and second preferred stock, where the first preferred stock must receive its entire preference dividend or nothing can be

paid on the second preferred, and the second preferred must receive its entire preference dividend or nothing can be paid on the common stock. Again, it is possible for either issue, or both, to carry cumulative or participating features, or both.

We have explained the way in which preferred stock is given an advantage in the declaration of dividends. It is also often given preference in the distribution of the assets of the corporation if it goes out of business and if its assets are disposed of. In such a case it would be common for the holders of preferred stock to receive the full par value of their shares before anything was paid to the holders of common stock. This latter condition is called "preference as to assets," while the former, which we have described at length, is called "preference as to dividends."

9 · 10. Bonds. A stockholder, as we have seen, is one of the owners of a corporation; but a bondholder is merely one who has lent money to a corporation—one of its creditors. A bond is, in effect, a form of promissory note, in which the corporation acknowledges the receipt of the borrowed funds, promises to repay it at some time in the future (usually at a rather distant time, such as 25 or 50 years), and promises in the meantime to pay interest on the borrowed money at some definite rate. Thus a bond differs radically from a share of stock. The share of stock does not guarantee to pay anything ever. The bond, on the contrary, promises to pay both principal and interest and, in case either is not paid in accordance with the contract, the bondholders may throw the corporation into bankruptcy. The payment on bonds is called "interest," while the payment on stock is called "dividends." The stockholders own and control the corporation, but the bondholders are not the owners and hence have no vote in the selection of its officers.[1] The stockholders run the risk that the business may be unsuccessful and that they will receive no dividends. The bondholders, to be sure, run some risk also. Even the bonds may not be paid. If they are not, the corporation is thrown into bankruptcy. The corporation could make no payments on any of its stock unless the bondholders were paid whatever was owed them first. Thus bonds are safer than stock.[2] On the other hand, bonds receive a fixed rate of interest, while a successful company may pay far more on its stock than on its bonds.

Bondholders run another sort of risk that is often overlooked by the

[1] It is often true that a provision is included in corporation's bonds which stipulates that, if the corporation defaults in its payments of interest or principal, the control is to be taken over by the bondholders.

[2] This statement might, by itself, be misleading. No investment is completely safe, and the stock of a good company may be much safer than the bond of a weak company. In general, bonds are safer than stocks.

casual investor. If the cost of living rises, bondholders find that they are getting a fixed money income and their real income (see Sec. 4·1) falls. If, on the contrary, the cost of living falls, their fixed money income will buy more and their real income rises. Thus the bondholder runs a risk on changes in the general level of prices. The stockholder does not run this risk to the same extent, because usually, when prices are rising, business profits and dividend rates are also rising, so that the stockholder gets rising money income to offset rising prices, and his real income does not suffer, or at least not to the same extent as does that of the bondholder.

Just as we found that there are several kinds of stocks, so there are several kinds of bonds. We have said that bonds are promissory notes, agreeing to repay borrowed money. When one lends money, it is not uncommon for him to demand security. Thus the pawnbroker lends money and accepts one's watch as security. If the debt is not repaid, the pawnbroker sells the watch to recover what is owed him. The farmer borrows money and mortgages his farm. If he does not repay, the lender forecloses the mortgage, sells the farm, and recovers what is owed him. The major differences between corporate bonds are in the type of security which is pledged and which can be sold by the bondholders in case the payments of interest and principal are not made as promised. On this basis we recognize four types of bonds: mortgage bonds, collateral trust bonds, equipment bonds, and debentures. *Mortgage bonds* are those which have real estate as security. In this way they are like mortgages on farms or homes. If the interest or principal on the bonds is not paid, the bondholders may sell the real estate and collect what is due them. *Collateral trust bonds* have other securities pledged as backing: the stocks or bonds of other corporations, or perhaps government bonds. Thus the Gurleyville Rapid Transit Company might borrow money on an issue of bonds and pledge as backing certain securities that it owned. If payment were not made on the bonds that it issued, then the bondholders could sell the pledged securities to recover their investment. *Equipment bonds,* most common in the railroad business, are issued when a corporation borrows to buy new equipment, and that new equipment is itself put up as security. Thus a railroad might borrow to buy new cars and put them up as security. In the event of default on the bonds, the bondholders could sell the cars to raise the funds that were due them. A *debenture bond,* usually called merely a "debenture," has no specific backing. It is the sort of thing that we have when a college student borrows $5 from his roommate. Under such circumstances the borrower seldom pledges specific property that can be sold in case the debt is not repaid. Usually the borrower does not put up his watch or his best suit as security. Yet it would be true that the roommate could, if worst came to worst, sue for the return

of the funds, and, if the borrower were earning an income, that income could be attached by the roommate. Debentures have no specific backing but, if the bonds are not paid when due, the bondholders can take over the earnings of the company and get their return in that way.

9·11. Priority of Claims against a Corporation's Assets. In the ordinary course of events a corporation pays its various bills as they come due; in case it is not able to pay all its bills, some decision has to be made as to which creditors will be paid. Suppose a corporation has property worth 1 million dollars but has debts of 2 million dollars. If it cannot pay and is thrown into bankruptcy, the assets will be disposed of and arrangements made to pay the creditors. Under these circumstances the creditors will not usually all be dealt with alike, each receiving half the amount owed him. On the contrary, amounts owed to the government in taxes would first be paid in full if there were sufficient funds. After taxes were completely paid, any remaining funds would be used to pay wages owed. If there was still money left, it would be used to pay for raw materials that had been purchased. Of course, bonds that were backed by specific security would be paid out of that security, but debenture bonds would come in for the next share after repayments for raw materials. Then, if there were still sufficient funds, the preferred stock would be paid; last of all, would come the common stock. Taxes, wages, materials, bonds, preferred stock, and common stock would be paid in that order, each receiving the entire amount due if there were sufficient funds before anything at all was paid to the next in line. Thus we see again why it is that bonds are safer than stocks. They get the first share in regular payments, since interest on bonds must be paid before dividends can be paid; and, in case of the dissolution of the corporation and the liquidation of its assets, the bonds must be repaid in full before anything can be paid on the stock. Of course, in our illustrative example, it is possible that the firm with only 1 million dollars' worth of assets to pay 2 million dollars' worth of debts might use up the entire amount available in payment of taxes, wages, and materials, and the bonds might receive nothing. Bonds are not riskless, merely less risky than stocks.

9·12. Control of Corporation. We have already discussed in Sec. 9·9 the organization of the corporation and the way in which its officers are selected. Let us examine the effect of this organization. Suppose we have a company with a capitalization of a million dollars. A company's *capitalization* is the total amount of stocks and bonds of all kinds issued, so this company has stocks and bonds outstanding of a million dollars. Let us suppose that half the capitalization is in bonds and half in stock. The bonds have no vote and hence no control of the corporation's policies. If half of the stock is preferred and half common, we would have $250,000 of each; but again

the preferred stock often has no voting power and hence no control over corporate policies. During the present century a new development has been the occasional issuance of two kinds of common stock: voting and nonvoting. If our corporation is one of those which issues nonvoting common stock and if its common is divided equally between voting and nonvoting common stock, then we have $125,000 of voting common which has all the votes. If an investor can control half of this voting stock, he can control the corporation. Thus, in our fictitious example, an investment of $62,500 in the voting common stock would control the corporation capitalized at a million dollars. Although this case is entirely imaginary, it is by no means extreme. Many modern corporations have the control concentrated as much as this.

Moreover we have to remember that it is often possible to control a corporation without owning a full majority of the voting stock. If the remainder of the stock is scattered among many small owners, the chances are pretty good that the owner of 30 to 40 per cent of the voting stock can in practice control the situation. It is very unlikely that all of the remaining stock will combine against him. Thus a recent study of the 200 largest nonfinancial corporations in the United States led to the conclusion that about 40 of them were controlled by single families, yet in only 8 did the controlling family have a clear majority of the voting stock. In 12 of the companies, control was maintained through ownership of between 30 and 50 per cent of the stock; in about the same number, the control was maintained with the ownership of from 10 to 30 per cent of the voting stock. Finally the investigators concluded that there may have been effective control in 7 cases through ownership of less than 10 per cent of the voting stock.[1] Thus we see that the organization and method of functioning of the corporation are such that it becomes possible for relatively small investments to control relatively large corporations. Many of the investors in the corporation will have no voice in formulating its policies, and others will have no effective voice. Even if we eliminate the bondholders entirely, since they are not owners of the business at all, we find that it is often true that only a small proportion of the stock has voting power, and possibly a minority of the voting stock actually controls the corporation. Thus, we have witnessed with the development of the corporation a marked tendency for separation of ownership and control. The ownership of corporations has become very widely dispersed, until it is estimated that some 8 or 9 million Americans own corporate stock.[2] But most of them own very minor parts

[1] See The Distribution of Ownership in the 200 Largest Nonfinancial Corporations, *T.N.E.C. Monograph* No. 29, p. 104, Investigation of Concentration of Economic Power, 1940.

[2] *Ibid,* p. xvii.

of the outstanding voting stock of their corporations, as can be seen from the data of Table 9·5, which shows the numbers of shareholdings in 1584 corporations whose stocks were listed on a national stock exchange in 1937-1939.

When a corporation has thousands of stockholders, most of them will of necessity own very small fractions of the voting stock. Most of them exercise no management functions at all. They are, and think of themselves as, investors rather than as managers. As ownership has become more and more widely dispersed, control has become relatively more and more highly concentrated.

9·13. The Balance Sheet. Corporation accountants keep track of the complicated and intricate transactions of the corporation, and of its financial

Table 9·5. Numbers of Shareholdings in 1584 Issues of Common Stocks Listed on a National Stock Exchange, United States, 1937-1939*

Number of Shareholdings	Number of Stock Issues
Under 100	18
100–249	74
250–499	191
500–999	305
1000–2499	431
2500–4999	220
5000–9999	147
10,000–24,999	106
25,000–49,999	51
50,000–99,999	26
100,000 and over	15
Total	1584

* Data from Survey of Shareholdings in 1710 Corporations with Securities Listed on a National Securities Exchange, *T.N.E.C. Monograph* No. 30, p. 67, Investigation of Concentration of Economic Power, 1941. Shareholdings are not exactly the same as shareholders but, when we consider single stock issues, as we are doing here, the terms are very nearly synonymous.

position at any time, by means of many sorts of devices. It is not our intent here to make accountants of the readers of this book, but it is necessary for us to become acquainted with a few of the simpler and most useful end products of the accountant's work if we are to visualize the corporation in action.

One of the most widely used of the financial statements of the corporation is the balance sheet, which is merely a list of all the corporation's assets and another list of all its liabilities as of some given moment of time. The *assets* of a business are all those things which the business owns or which are owed to it. The *liabilities* of the business are those things which the business owes to others. The firm's assets and liabilities change from year to year, from day to day, and even from minute to minute. Every time

we buy or sell property or incur debts, we have changed our assets or our liabilities, or both. Therefore a balance sheet always describes the situation as it exists at some specified time.

Let us assume that a group of men come together to form the Veteran Manufacturing Company. Some of the owners of the new business contribute more and some less, but altogether they contribute $800,000 in cash, receiving in return shares of stock in the new corporation. At this point the balance sheet of the corporation will appear as follows:

Veteran Manufacturing Company
Balance Sheet
July 8, 1946

Assets		Liabilities	
Cash	$800,000	Capital Stock	$800,000

At this point the only thing owned by the corporation is the cash that has been subscribed for the stock. Yet, since the corporation is a fictitious entity, all its assets must belong to someone. They are thought of as being owed to the stockholders and hence appear among the liabilities.[1]

If the corporation now buys or builds a manufacturing plant and equips it with machinery, it will have to spend some of its cash. Suppose that its real estate costs $300,000 and the equipment costs $250,000. The balance sheet after these investments have been made would appear thus:

Veteran Manufacturing Company
Balance Sheet
August 1, 1946

Assets		Liabilities	
Cash	$250,000	Capital stock	$800,000
Real estate	300,000		
Machinery and equipment	250,000		
	$800,000		$800,000

We note one peculiarity, which is true of every balance sheet and which, in fact, gives the statement its name: the sum of the assets always equals the sum of the liabilities if we include among the latter the claims of the stockholders, in this case the capital stock. The assets and the liabilities always balance unless an error has been made in computation.

Suppose the corporation now buys $12,000 worth of raw materials.

[1] Many accountants would hold that the capital stock is not really a liability, even though it is always listed on the liability side of the balance sheet. We shall mention this again later when describing the entire owners' equity (p. 148).

It pays cash for $3000 worth of them and gets the other $9000 worth on credit.[1] The balance sheet will now show the following items:

Veteran Manufacturing Company
Balance Sheet
August 15, 1946

Assets		Liabilities	
Cash	$247,000	Capital stock	$800,000
Real estate	300,000	Accounts payable	9,000
Machinery and equipment	250,000		
Raw materials	12,000		
	$809,000		$809,000

It has a new liabilities item, since it now owes $9000 for the raw materials purchased. Since this is a sum owed to other people, it appears among the liabilities. We have now $809,000 of assets and $809,000 of liabilities, but the assets are now owed partly to outsiders ($9000 worth of them) and partly to the stockholders ($800,000 worth of them). If the corporation continues to do business for the next year or so, buying materials, manufacturing them, and selling the finished product at a profit, its balance sheet may come to appear as follows:

Veteran Manufacturing Company
Balance Sheet
December 31, 1947

Assets		Liabilities	
Cash	$ 91,000	Capital Stock	$800,000
Real estate	300,000	Accounts payable	68,000
Machinery and equipment	250,000	Notes payable	22,000
Raw materials	38,000	Undivided profits	20,000
Finished goods	146,000		
Accounts receivable	75,000		
Notes receivable	10,000		
	$910,000		$910,000

Again the sheet balances. The new asset items are self-explanatory. The corporation now has on hand both unmanufactured raw materials and finished products ready for sale. It has sold $75,000 worth of goods on credit and has not yet been paid for them. This amount is owed to the corporation and therefore appears among the assets as accounts receivable. It is also owed $10,000, listed as notes receivable. These are in every way analagous to the accounts

[1] The method by which the corporation would buy goods on credit, and the effects of such credit transactions, are discussed in Chap. 19.

receivable save for the fact that the corporation holds formal promissory notes from its debtors and therefore lists these amounts separately. On the liabilities side there are parallel entries of accounts and notes payable. These are amounts owed by the corporation for goods and materials purchased on credit, $68,000 worth of them being on ordinary credit and $22,000 of them being covered by formal promissory notes. Finally, we see that the corporation has made profits of $20,000 which have not yet been distributed to the stockholders. These are listed as undivided profits and appear among the liabilities because they are owed to the stockholders.

If we may take one further illustrative balance sheet as it might appear still a year later, we might have the following:

Veteran Manufacturing Company
Balance Sheet
December 31, 1948

Assets		Liabilities	
Cash	$ 50,000	Accounts payable	$ 75,000
Accounts receivable	95,000	Notes payable	35,000
Notes receivable	45,000	Taxes due	3,000
Finished goods	175,000	Bonds outstanding	500,000
Raw materials	65,000	Capital stock	800,000
Machinery and equipment	580,000	Surplus	75,000
Real estate	500,000	Undivided profits	22,000
	$1,510,000		$1,510,000

Since the preceding balance sheet, the firm has issued $500,000 of bonds which were used to expand the plant and to buy more equipment. Bonds, as we have learned, are debts of the corporation and appear among its liabilities. It also appears that when the accounts were made up on Dec. 31, 1948, the corporation owed $3,000 of taxes which had not been paid. Being a debt of the corporation, they also appear among the liabilities.

9·14. Owners' Equity. It is the last three items on the liability side to which we should give our attention for the moment. We recall that the $800,000 item represents the original investment of the stockholders. At some time since the corporation was started, it has evidently made $97,000 of profits that have not been distributed. When a profit is first made and before the corporation has disposed of it, it is listed as "undivided profits." There are then two things that the directors may decide to do: (1) They may declare dividends and pay them, in which case the undivided profits disappear from the liability side and a corresponding amount of cash disappears from the asset side. (2) They may decide to retain these profits in the business, as we noted in Sec. 9·9. In such a case, if they decide not to

pay the profits out as dividends but to retain them in the business, they call them "surplus" instead of "undivided profits." Both surplus and undivided profits are amounts earned with the stockholders' money. The difference between them is merely that the directors have not yet decided what to do with the undivided profits, but they have voted to reinvest the surplus in the business. The $75,000, now listed as surplus, once appeared as undivided profits, but the designation was changed by vote of the directors.

Although these last three items—capital stock, surplus, and undivided profits—are always listed with the liabilities, they differ greatly from the others listed there. In all the other cases amounts are owed which must be paid sooner or later. But it is never necessary for the corporation to pay the capital stock, the surplus, or the undivided profits items. We note that our corporation has $1,510,000 of assets. Of these, $613,000 are owed to other people and firms as taxes, as payment for raw materials, in return for money borrowed, etc. But this leaves $897,000 worth of assets over and above all debts to outsiders—$897,000 worth of assets that belong to the owners of the business clear of debt. This we call the "net worth" of the business, or the "owners' equity." The net worth, or owners' equity, is always the sum of the capital stock, the surplus, and the undivided profits. It is the set of items that makes the balance sheet balance. If the firm were to lose $5000, we would deduct it from the undivided profits. If they were to lose $25,000, we would wipe out the entire undivided profits and have to reduce the surplus. Continued losses might even impair the capital stock. Our company had owners' equity of $897,000, which means that they could lose that amount and still be able to pay off all the outside creditors in full. If they were to lose more than this, they would be *insolvent.* A company is insolvent when its assets are insufficient to meet the obligations to others than the owners.

The student may at first wonder how it is possible for our corporation to have $75,000 of surplus when it has only $50,000 of cash altogether. This would reflect a common misconception. Surplus is not cash. In fact, no item on the liability side corresponds to any particular item on the assets side, or vice versa. The corporation has $1,510,000 worth of assets, but not by any means $1,510,000 of cash. The liability side of the balance sheet shows the amounts of the claims against the assets, but no particular asset offsets any particular liability. The liability items are not cash, but claims to property. The $50,000 listed among the assets is all the cash that the corporation has.

9·15. Values of Stock. Corporations issuing stock often print on their stock certificates an arbitrary figure called the "par value." Commonly, although by no means uniformly, the par value is $100. In such a case the stock certificate would bear the statement, "Par value $100." This figure is, as we

say, arbitrary. It does not show necessarily what the stock is worth or what it ever will be worth. It is merely a convenient method for figuring how many votes a stockholder gets, or how much dividends are due him. Thus, if I have stock with par value of $500 and if the directors declare a 4 per cent dividend, I will get 4 per cent of $500, or $20. If the total stock outstanding has $800,000 par value and I have $500 par value (which would be five shares with the par value at $100), then I own 500/800,000 of the corporation, and get 500/800,000 of the dividends and 500/800,000 of the votes. But the corporation does not promise to pay me $500, and it is usually true that the par value is not equal to the actual value of the stock. It is merely a convenient method of computing dividends, voting power, etc., and is not really a monetary value at all.

Because many people are confused by par values, assuming them to represent actual monetary values when they do not, corporations have in many cases issued stock without any par value, called "no-par stock." In such a case we can still easily compute dividends and voting rights. The directors can declare a dividend of $4 per share instead of a 4 per cent dividend. (When a dividend is declared as a percentage, it always means some percentage of the par value.) If the total number of shares of stock outstanding is 800 and if I own 40 shares, then I get 40 votes out of a total of 800. In recent years no-par stock has come to be even more common than stock with par value.

The actual amount that one can get for his stock if he sells it is called the "market value" of the stock. This is the figure carried in the listings on the financial pages of the newspapers. Market value depends largely on expectation of earnings. If investors think that a corporation is likely to be a profitable one and to pay large dividends, they will be willing to pay large amounts for the stock. If, on the other hand, it is expected that the corporation will declare no dividends, investors will not want the stock. Men may also value stock for its voting rights if they wish to control the corporation, or for other things; but the major thing that influences the market value of stock is the expectation of earnings.

There is one other kind of stock value that is often mentioned. If we look back to our last balance sheet, we see that the owners' equity is $897,000. The owners of the business have $897,000 of assets over and above all the claims of outsiders. If the par value of this corporation's stock is $100, the corporation has 8000 shares outstanding.[1] This means that the corporation has

[1] This we know because, when a corporation's stock carries a par value, the capital stock is always listed in the balance sheet at that par value. In our balance sheet we find capital stock listed at $800,000, and if the par value is $100 there must be 8000 shares. The listing in the balance sheet is always the product of the par value and the number of shares. If the corporation issues no-par stock, it has to assign some arbitrary value to it for purposes of listing in the balance sheet.

assets of $112.12 behind each share of stock. This we call the "book value" of the stock, and we always find it by dividing the owners' equity by the number of shares.

To summarize, there are three kinds of values often attributed to stock: par value, book value, and market value. Par value is arbitrary and unnecessary, and some stock does not have it. It is not a monetary value in the usual sense at all. Book value shows the value of assets that we would have behind each share of stock if we paid off all claims of the nonowners. Market value shows what we can actually get for the stock if we sell it, depending largely on the expected earning power of the corporation. These three kinds of value need not be, and usually are not, the same.

9 · 16. Overcapitalization. On page 146 we assumed that the Veteran Manufacturing Company started in business by selling stock, and that it sold stock with a par value of $800,000 and received $800,000 in cash for it. In practice, corporations may issue stock in return for other things than cash. If they give me $500 par value of stock for my $500 in cash, there is no reason why they should not give you $25,000 par value of stock if you turn over to them $25,000 worth of land as a factory site. Stock may be given in exchange for any property or services, and not merely for cash. Under such circumstances, it sometimes occurs that the promoters of the corporation are rather liberal in their appraisal of property that is offered to the corporation. If I offer them $500 in cash, there is no trouble in knowing the value of my contribution. If I offer them a factory or a building lot or a patent, there may well be honest difference of opinion as to the worth of my contribution. Should I be given $20,000 or $25,000? Someone has to decide and, as has just been said, the tendency often is to be rather liberal in appraising property. Suppose that our new corporation starts out by issuing $60,000 of stock for cash, which is fully paid at $100 a share. In addition it gives $200,000 of stock for a building and building lot. The balance sheet will then appear as follows:

Veteran Manufacturing Company
Balance Sheet
July 8, 1946

Assets		Liabilities	
Cash	$600,000	Capital stock	$800,000
Land and buildings	200,000		

If the building and building lot actually were worth only $150,000, there is no way that we can tell the fact from the balance sheet. Yet, in fact, we would have $800,000 of stock outstanding with but $750,000 of assets. The company would be *overcapitalized* by $50,000. Businessmen would say that there was $50,000 of "water" in the stock. We have *watered stock* when the actual value of the assets behind the stock is less than its par value. Watered stock can

never be discovered by looking at the balance sheet, since that sheet would carry the property at an inflated appraised value to offset the amount of the stock. It would take an independent assessment of the assets to discover whether or not there was really overcapitalization.

Overcapitalization has been widely misunderstood and often vigorously condemned. The major argument against it is that it may make it possible for a company to hide exorbitant earnings. Let us imagine two companies, each with $500,000 of stock outstanding. Corporation A actually has $500,000 of investment behind its stock, while an accurate appraisal would show but $400,000 of investment behind the stock of corporation B. Thus, corporation B is overcapitalized by $100,000. Now suppose that each company is declaring dividends at the rate of 6 percent. They appear to the novice to be making equal earnings. But dividends are always declared on the basis of par value when a corporation's stock has value. Therefore each corporation is paying as dividends 6 per cent of $500,000, or $30,000. The $30,000 is 6 per cent of the actual investment in corporation A, but it is 7½ per cent of the actual investment in corporation B. The latter has made its earnings of 7½ per cent appear to be earnings of 6 per cent. Overcapitalization always makes the earnings appear smaller, because the dividends are declared as a percentage of the par value. When we have overcapitalization, the par value is greater than the actual investment.

Those who understand little of the facts of economics have supposed that there was another difficulty with watered stock. They have argued that, if a corporation has watered stock, it will tend to charge higher prices to consumers in an effort to get enough income to pay reasonable dividends on the inflated amount of stock. This overlooks the fact that corporations usually tend to charge those prices which will yield the largest possible net income, regardless of capitalization. When a businessman tries to decide what to charge for his product, he pays no attention at all to his capitalization. As we shall see later,[1] he is influenced solely by cost and incomes, not by the balance sheet. If watered stock is disadvantageous socially, it is because it helps corporations hide their true earning rates from those who understand little of accounting procedure.

9·17. The Income Account. The balance sheet gives us a description of the property and debts of a corporation or other business unit at a given instant of time. Often, however, we wish to know about the operations of the business through a period of time, such as a year. The most common statement giving information on this subject is the *income account* or *profit-and-loss statement.* This is usually drawn up to show the income and expenditures during a year and the amount of profit made. The terms can be most

[1] See Chaps. 15 to 17.

easily understood if we illustrate with a profit-and-loss statement for a fictitious company.

Profit and Loss Statement
Veteran Manufacturing Company
For the Year Ending December 31, 1948

Total sales	$1,496,000	
Less goods returned	13,000	
Net sales		$1,483,000
Less cost of goods sold:		
Inventory, Jan. 1, 1948	$ 29,000	
Purchases during year	346,000	
Inward freight charges	14,000	
	$ 389,000	
Less inventory, Dec. 31, 1948	31,000	
Cost of goods sold		358,000
Gross profit		$1,125,000
Less operating expense	$ 600,000	
Less selling expense	114,000	
Less administrative expense	94,000	
Less taxes paid	25,000	
		833,000
Net profit on operations		292,000
Add nonoperating income		2,000
		294,000
Deduct nonoperating expenses		1,000
Net profit for the year		$ 293,000

Some of the items in this list may require further explanation. The total value of all goods sold by this corporation during the year was $1,496,000, but $13,000 worth of the goods were returned, so the net sales amounted to the difference, or $1,483,000. This is sometimes called the "gross income" or "total income." Of course this is far from being the amount of the profit. The goods sold cost something. We note that the company had on hand at the beginning of the year goods that had cost $29,000, and during the year they paid $346,000 for additional goods and $14,000 to get them delivered. This was a total of $389,000 of goods all told. But $31,000 worth of these goods were still on hand at the end of the year, so the cost of the goods actually sold was the difference, or $358,000. If we deduct this cost of goods sold from the net sales, we have what is called the "gross profit"—the difference between what was received for the goods and what was paid for them.

There were many expenses of operating the business in addition to the outlays for purchase of goods. We paid $600,000 for such actual operating costs as wages, power, etc. We spent $114,000 for advertising, salesmen's salaries, and the like. The salaries of our administrative officers, our accounting department, etc., amounted to another $94,000. We paid taxes to the extent of $25,000. After we deducted these expenditures, a total of 833,000, we had $292,000 remaining as the actual profit for our business operations during the year. This is perhaps the most important figure in the statement.

Often business firms get some income or make some expenditure in connection with activities that are not really a part of their business at all. For example, a company may sell part of its plant or its land at a profit, yet it is not in the real-estate business. It may get some interest on government bonds in which it has invested, although it is not in the investment business. It may take a loss on sale of securities or real estate, although this, too, would not be properly chargeable to its ordinary business. Therefore it is necessary, where such items occur, to add any income from nonbusiness sources (we list $2000 of this nonoperating income in the statement) or to subtract any expenses for nonbusiness purposes (we have subtracted $1000 of such nonbusiness expense in the statement.) This leaves us with the final net profit for the year, which in this case was $293,000. Of course, if there had been no nonoperating items, the net profit on operations and the net profit for the year would have been identical.

Our profit-and-loss statement often shows the disposition that was made of the net profit for the year. In this case we had $293,000 of such profit, and we might well add to the statement a notation of this general sort:

Disposition of net profit:	
Dividends declared	$143,000
Carried to surplus	150,000
Total	$293,000

Under such circumstances the balance sheet for Jan. 1, 1949, should show a surplus larger by $150,000 than that in the balance sheet for Jan. 1, 1948. In this way the balance sheet and the profit-and-loss account are related. The balance sheet, however, always shows the momentary situation at an instant, while the income account shows the income and expenses throughout a period of time. As has often been remarked, the former is like a snapshot, the latter like a moving picture.

9 · 18. Dividends. Let us now turn back to the illustrative balance sheet that appears just before the end of Sec. 9 · 13. We see that the corporation has capital stock of $800,000, surplus of $75,000, and undivided profits of $22,000. Suppose the directors of the corporation are now meeting and going over the financial statements. The question arises as to the disposition of the

$22,000 of undivided profits. Since they amount to 2¾ per cent of the capital stock, it would be possible to declare up to a 2¾ per cent dividend. On the other hand, the directors are under no obligation to pay any dividend at all. Between the limits of no dividend and a 2¾ per cent dividend, they may do whatever they desire. Suppose that they vote to pay a dividend of 1½ per cent on the capital stock and transfer the remainder of the undivided profit to the surplus account. One and one-half per cent of the capital stock is $12,000, so this amount will be paid out in cash. Each stockholder will receive dividends amounting to 1½ per cent of the par value of the stock that he owns, so that, if I have stock with par value of $500, I shall receive $7.50 in dividends. This payment of a cash dividend will reduce the cash on the asset side of the account by $12,000 and will reduce the undivided profits by $12,000 also. Therefore the balance sheet will still balance. Having paid out $12,000, we shall reduce the undivided profit to $10,000; but we have voted to transfer this to the surplus. We therefore decrease the undivided profit by $10,000 (which wipes it out altogether) and increase the surplus by $10,000, making it $85,000. Our balance sheet will now appear as follows:

Veteran Manufacturing Company
Balance Sheet
January 1, 1949

Assets		Liabilities	
Cash	$ 38,000	Accounts payable	$ 75,000
Accounts receivable	95,000	Notes payable	35,000
Notes receivable	45,000	Taxes due	3,000
Finished goods	175,000	Bonds outstanding	500,000
Raw materials	65,000	Capital stock	800,000
Machinery and equipment	580,000	Surplus	85,000
Real estate	500,000		
	$1,498,000		$1,498,000

We note that the surplus has now grown fairly large.[1] The stockholders have an original investment of $800,000 in the business, but another $85,000 of their money has been invested for them by the corporation's directors, and the decision has been made to keep it permanently in the business. There is no good reason to distinguish this investment permanently from the original investment. Moreover, as the surplus grows larger and larger, we may give a misleading impression, since we declare the dividend as a percentage of the par value of the capital stock rather than as a percentage of the actual investment—the capital stock plus the surplus. If the surplus were as large as the capital stock, a 10 per cent dividend on the capital stock would be but

[1] We should not imply that it is unusually large. Financial corporations, such as banks and insurance companies, often carry a surplus larger than the amount of the capital stock.

a 5 per cent dividend on the actual investment. In such cases corporations often declare a "stock dividend" which is a dividend in the form of new shares of stock rather than in cash. In our present example, the directors might decide to declare a 10 per cent stock dividend. This means that they would issue one-tenth as much stock as they already had outstanding—another $80,000 of stock—and give it to the present stockholders. Each stockholder would get one share of the new stock for each 10 shares that he had before.[1] Now $80,000 of the former surplus would be represented by capital stock. The capital stock item of the balance sheet would be increased by $80,000 to $880,000, since the corporation now has $80,000 more of stock outstanding. The surplus would be reduced by $80,000, since stock has been issued against $80,000 worth of it. This would leave a surplus of but $5000. The student will note that the directors are able to transfer from the undivided profits item to the surplus item by vote and they are also able to transfer from the surplus item to the capital stock item by the issuance of a stock dividend.

The foregoing description merely touches the surface of the problems of accounting and corporation finance into which the student of business or general economics goes much more deeply. Behind the figures that we have used lie many difficult and technical problems of appraisal of property, of record keeping on costs and income, and of determination of the depreciation of property. We have said enough, however, to give the general student an idea of the ordinary operation of the corporation, which is as far as we can go in elementary work.

Questions

1. What characteristics of the business itself would influence you in deciding whether to use an individual proprietorship, a partnership, or a corporation in starting a new business?

2. What advantages does the individual proprietorship have over the partnership? What disadvantages?

3. Does the right of *delectus personae* exist in any nonbusiness relationships, or is it confined to the partnership? Illustrate.

4. In what principal ways does the joint stock company differ from the corporation? From the partnership?

5. Explain carefully the meaning of limited liability.

[1] If a stockholder had three shares of stock, he would receive a certificate giving him the "right" to buy three-tenths of a share of the new stock. He could then either buy other such rights to make up enough to get a full share, or he could sell his right to someone else who was trying to get a share of the stock. Thus rights to stock are bought and sold at the time of a stock dividend.

6. What differences between modern business and the business of the medieval manor explain the growth of the corporation?

7. Is the corporation the cause or the result of large-scale business? Discuss.

8. What is a cumulative preferred stock? A participating preferred stock?

9. How would the figures of Table 9 • 4 have been altered if the stock in question had been cumulative but not participating?

10. How would the figures of Table 9 • 4 have been altered if the stock in question had been participating but not cumulative?

11. If cumulative participating preferred stock has so many advantages, why does anyone buy any of the other kinds of stock?

12. Explain how it might be true in a given case that an investment of $10,000 would control a corporation capitalized at a million dollars.

13. Why are a corporation's assets and liabilities always exactly equal?

14. Under what circumstances would the book value of a stock be lower than its par value?

15. If the stock of the Veteran Manufacturing Company has par value of $100, what is the book value on Jan. 1, 1949, based on the balance sheet in Sec. 9 • 18?

16. How could you state the book value in the preceding case if the stock had no par value?

17. Does either the book value or the par value of a stock influence its market value in any way? If so, how?

18. Section 9 • 16 discusses overcapitalization. Is it possible for a company to be undercapitalized? What would the term mean?

19. Explain the relationship of inventories to the cost of goods sold.

20. If you were passing judgment on the efficiency with which a business firm had been managed, would you prefer to know the net profit on operations or the net profit for the year? Why?

10. THE NATURE OF PRODUCTION

10 · 1. Production. In Sec. 1 · 9, we noted the fact that most things that men want do not appear in nature at all. One does not find shoes or suits of clothes, loaves of bread, dining-room tables or radios growing on bushes. If men want these things at all, they must work to get them. If wants are to be satisfied, there must be production. Yet it is easy for us to misunderstand what is implied in the word "production." The physicist tells us that matter can be neither made nor destroyed, that the same materials exist today that existed centuries ago, no more and no less. How, then, can we produce anything?

There is really no question of fact at dispute in this case, but merely a misunderstanding over the meaning of words. When we say that a company "produced" automobiles or cameras, we do not mean to imply that they brought into being any new material. The things from which the camera was made existed before and will exist long after the camera is worn out. But the camera did not exist before and in this sense we can properly say that a camera was produced. Someone took preexisting materials and made them more useful than they had been before. Wood, leather, glass, and metal were shaped and fitted and combined in such a way that they could satisfy wants that could not have been satisfied by the raw materials separately. Similarly, when the housewife makes a cake, she combines 1½ cups of water, 1 cup of raisins, 1 cup of sugar, ½ cup of shortening, ½ teaspoon of cloves, ½ teaspoon of cinnamon, and two squares of chocolate; boils 5 minutes; cools and adds 2 cups flour, 1 teaspoon baking soda, and ¼ teaspoon salt. She then bakes for an hour at 375° F. When she is through, she has "produced" a cake, and everyone who tries it will agree that it satisfies wants better than would the 10 ingredients taken separately. This is one of the strange cases where mathematics does not seem to be applicable to economic phenomena—the utility of the whole is not equal to the sum of the utilities of the separate parts. By combining the ingredients in the proper way there has been an actual creation of utility even if there has been no creation of

chemical elements. And it is the creation of utility to which the economist refers when he talks about production. Men whose activities make it possible for more wants to be satisfied are producers. *Production is the creation of utility.*

10 • 2. Kinds of Utility. Perhaps the most obvious way in which men add utility to material things is by changing their form; felling trees, sawing them into planks, and manufacturing them into tables, chairs, or wooden legs. In their final forms these goods are much better able to satisfy wants than they were originally. Mining, manufacturing, and agriculture consist largely of altering the forms of things that existed in nature, putting them into a form better able to satisfy wants—a more useful form. Such activities, we say, add "form utility" and are for that reason productive.

After milk has been produced, pasteurized, homogenized, and bottled—after it has been given the usual kinds of form utility—it is still not usually ready for satisfying wants. It must still be taken to the place where the consumer can use it. The quantity of milk in the producing regions is so great that the marginal utility is very low.[1] But if the same milk can be shipped to New York City or to some other metropolitan area, the supply on hand will be small in proportion to the population, and the marginal utility will be higher. If we ship the milk from a surplus area to a deficit area, we are shipping it from a place of low to a place of high marginal utility. We are shipping it to a place where it is more useful, where it can and will satisfy more wants. Thus we have created utility in the milk, entirely aside from any change in form, merely by taking it from one place to another. We would say that we had created "place utility." All the world's transportation agencies —railroads, air lines, motor trucks and highways, steamboat lines, ox carts, and camel caravans—are engaged primarily in adding place utility to goods, carrying them from where they are less useful to where they are more useful.

Just as things may be more useful in one place than in another, so they may be more useful at one time than at another, Eggs saved from the season of heavy production until the off season, apples saved from harvest time into the winter, ice saved from January until the Fourth of July, cotton stored from a year of heavy crop to a year of crop failure—these and many other similar illustrations are cases in which things have gained in utility without change of form or place. Carrying things from places of plenty to places of scarcity adds place utility; storing things from times of plenty to times of scarcity adds "time utility."

Some authors recognize also other kinds of utility that are added to goods. For example, they point out that goods often have more utility for one man than for another. If a man who owns a flute that he cannot play sells it to someone who can play it, there will be greater satisfaction of wants

[1] See Secs. 5 • 7 and 5 • 8.

than there would have been had the change in ownership not taken place. If a man spends his time finding, on the one hand, people who own things that give them little satisfaction and, on the other hand, people who would get more satisfaction from these same goods, and if this middleman can then bring the parties together and effect a change of ownership of the goods, he has made it possible for the goods to satisfy greater wants. The middleman has added "ownership utility."

Commonly we find that all these kinds of utility are added to any good before it is finally consumed. Peas are grown, picked, shelled, and canned—all changes in form. They are stored until wanted by the consumer, who may eat them months or even years after they were grown—an addition of time utility. They are carried from the field to the cannery in trucks, from the cannery to the wholesaler's by rail, from the wholesaler's to the retail store by truck, and from the retail store to the consumer's kitchen by the delivery boy—all adding place utility. They are sold over and over again: by the farmer to the cannery, by the cannery to the wholesaler, by the wholesaler to the retailer, and by the retailer to the consumer, each time going from someone who wants them less to someone who wants them more, and each time adding ownership utility. The goods are not entirely produced until they are in the form and place desired by the consumer, at the time he wants them, and in his possession. It was once common for economists, and is still common for others, to say that the goods are first produced and then marketed. But we see immediately that production does not stop with the original manufacture or change of form. The addition of the other utilities is just as much a part of production as is the addition of form utility. These other utilities are just as necessary for ultimate want satisfaction and are just as desirable socially. It would not help consumers to have form utility added to goods if the other kinds of utility were not to be added also. Every good is, when it reaches the consumer, a bundle of a great many utilities that have been added by many different people, all of whom cooperated (perhaps unconsciously) in producing the good in question; if any one of them had failed in his part of the task, it is quite possible that the consumer would have been unable to satisfy his want.

10·3. Services. When physical material goods are involved, the addition of utility is rather easy to see and understand. We can see the change in form or the change in place or the other changes involved. But many of the producers of society, while satisfying wants, are not satisfying them through the medium of physical commodities. The lawyer gives advice to his client, or the doctor to his patient. The dancing teacher explains the latest steps and gyrations. The public speaker exhorts, entertains, or instructs. In all these cases wants are being satisfied, yet in none of them are there physical commodities being sold—physical commodities to which the seller has given form,

time, or place utility. We say that these people are performing *services,* and we would classify them as producers and their efforts as productive if they are satisfying wants. The key to production is want satisfaction. Do the efforts result in greater satisfaction of wants than there would otherwise have been? If so, the efforts are productive; if not, the efforts are neutral or destructive according as the satisfaction has been left the same or reduced.

We can now define production in more detailed and final terms by saying that production is the addition of utility to material things, or the performance of services for which people are willing to pay. Any activity is productive if it increases human satisfaction.

10·4. Types of Industries. When we look at the production system around us, we see men at work at all sorts of occupations. Electricians, carpenters, dairymen, doctors, barkers at circus sideshows, blacksmiths, butchers, chauffeurs, dentists, painters, musicians, butlers, radio announcers, steel workers, surveyors—there seems to be no end. But a large proportion of the gainfully employed people of the United States work in one or the other of two great groups or classes of industries: what we call the "primary" or what we call the "secondary" industries. The distinction between the two rests on the source of their raw materials. *Primary industries* are those which take their materials directly from natural resources: mining, agriculture, trapping, or fishing. *Secondary industries* get their raw materials from primary industries or from other secondary industries. A steel mill that fabricates materials purchased from a mine would be an example, as would a textile mill weaving cotton that was purchased from another mill that made the yarn. Of course, all the raw materials for all the thousands of things that men use have to come originally from natural resources. There is no other source of supply. Consequently every secondary industry is dependent on primary industries, just as the primary industries are dependent in turn on natural resources. We shall come back to the distinction between these two classes of industry again when we discuss the trends of expense.

The primary industries are, in turn, divided into two subgroups. All primary industries take resources directly from nature, but in some cases the raw materials are replaced or reproduced, conserving the supply; in other cases they are taken away and nothing replaced so that the supply slowly diminishes. Primary industries that replace the materials are called "genetic industries"; primary industries that take raw materials from nature without replacement are called "extractive industries." Often a society has a choice between extractive and genetic sources of a commodity. For example, the meat supply of a society can come, as it often has historically, from hunting. The people take what they find in nature without replacing it. When the supply is used up, they move on to richer hunting grounds. Other societies have ob-

tained their meat from genetic industries, raising flocks and herds and reproducing them as they are used up. Similarly one can get blueberries by picking those which happen to grow wild, or he can raise them and propagate them, increasing the supply.

It is not intended to imply here that one of these industries is better in any ultimate sense than the other. Under some circumstances extractive industries are advantageous; under other circumstances it may be preferable to get the same commodities from genetic industries. When the United States was first settled, it would have been foolish for men to plant trees to raise for lumber. To have talked to them then about conservation of timber would have been to court commitment to an institution. But in other days and under other circumstances of supply and demand, the problem of replacement of forest resources becomes a real one.

One peculiarity of extractive industries, however, is that they are temporary. Raw materials are taken away without replacement, and no matter how slowly they are used up, someday they will be gone. If the petroleum supplies of the United States were mined at the rate of only a quart a year, some day they would be exhausted. This is particularly important in those industries, such as mining, in which there is no possible way of shifting from an extractive to a genetic industry when supplies dwindle. These are the industries that give us the problems of exhaustible resources. Some decision has to be made as to the best rate of use, always with the knowledge that any use at all will bring ultimate exhaustion. This inherent peculiarity of the extractive industries makes such industries, and the societies that are based on them, unstable and temporary. The type of social organization that one finds in a mining camp or a lumber camp is very different from that found in a settled agricultural region. In the lumber camp men cannot afford to build permanent public buildings or construct fine highways. Long before these things would be worn out, the lumber would be exhausted and the society would have moved on. Societies that are built on extractive industries do not ordinarily develop permanent social institutions, such as courts and educational systems. The difference between the extractive and the genetic industries influences men's lives and their modes of thinking as well as their buildings and public institutions.

Although it is impossible to give any exact idea of the numbers of people engaged in the various kinds of industries, a rough approximation can be made as in Table 10 · 1. The student should realize that forestry and fishing are to some extent genetic, although we have classed them as entirely extractive, and that other similar errors of classification make the table at best very rough.

10 · 5. Production for Profit. Adam Smith, great English economist, said

almost two centuries ago that "consumption is the sole end and purpose of all production." At first sight it appears that this is not true, at least in our time. We see men producing things that they never consume at all, or producing far more than they could possibly consume themselves. We say that they are producing the goods for sale rather than for use. Their hope is to sell them at a profit, not to consume them. It has become common for men to speak of our system as one in which we have "production for profit rather than production for use." Unfortunately this phrase is apt to be misleading. A man may work in a factory that makes pianos, and he may never buy a

Table 10·1. Distribution of Employment among the Various Types of Industries, United States, 1940*

Type of industry	Percentage of total employed persons		
Primary industries	20.80		
Extractive industries		2.26	
Forestry and fishing			0.23
Coal mining			1.17
Petroleum and natural gas			0.41
Other mines and quarries			0.45
Genetic industries		18.54	
Agriculture			18.54
Secondary industries	56.70		
Manufacturing		23.41	
Construction		4.55	
Transportation and communication		6.89	
Trade		21.85	
Services	20.98		
Personal services		8 88	
Professional services		8.22	
Government		3.88	
Unclassified	1.52		

*The data in this table are adapted from "Industrial Location and National Resources," National Resources Planning Board, December, 1942, p. 66.

piano or play one. He has, the saying goes, produced for profit rather than for use. But consumption is still his aim. He worked in the piano factory to get wages and, if the wages had not been forthcoming, he would soon have stopped. But he wanted the wages to buy clothes, food, and shelter and, if the wages had not been exchangeable for these consumption goods, he would not have accepted them. He works making pianos but converts the product, through the medium of money, into shoes, potatoes, carpets, and twin beds. We can still say with truth that consumption is the sole end and purpose of his production. Our case has merely become complicated by the fact that a producer nowadays ordinarily does not produce commodities for his own

consumption, but for the consumption of other people; in turn, he relies on other people to produce the things that he wants. Our piano maker is producing goods for someone to consume, and other people are producing goods for him.

In modern society most production is carried on in this way. Most things are produced by men who hope to make a profit out of them rather than to use them; but they hope in turn to spend their profits for consumption goods that other people have made. Here we find it convenient to split off the activities rather arbitrarily for study into the fields of production and consumption. The consumption unit is seeking satisfaction of wants—utility. The production unit is seeking profits. To be sure, these profits will then be used by consumption units, but it is simpler to separate the two kinds of activity for study. Some economists, especially the Germans, have divided economic units into two classes — firms and households, calling those units which are attempting to get the greatest possible profit "firms," and those which are trying to get the greatest possible satisfaction "households." It is the former of these problems that we are now to study; but it is of fundamental importance that the student should keep in mind that the two problems are actually part of one major problem, and not by any means antithetical or antipathetical.

10·6. Specialization: Chronological Division of Labor. We can summarize what we have just been saying by noting that in modern society most men are specialists. Instead of trying to make all the hundreds of things that he wants, the typical producer confines his productive activities to making a few commodities, and these he makes in great surplus, far more than he can use. Our worker in the piano factory makes many more pianos that he will ever use. The poultry farmer raises many more eggs than he could possibly eat. The automobile manufacturer makes more cars than he would have any use for. Everywhere we find men specializing in one line or in a few related lines, and producing surpluses which they hope to exchange for surpluses of other producers in other lines of business. This specialization takes two main forms: chronological division of labor and geographical division of labor. We shall for the time being turn our attention to the former, taking up the latter immediately thereafter.

Chronological division of labor, or chronological specialization, as its name implies, refers to a division in time. First, one man carries on one step in production, then another man does his part, and then a third man adds his share, until finally the finished product emerges. Thus the making of an automobile is subdivided into hundreds and hundreds of steps, starting with the mining of metal, the tapping of rubber trees, and the mining of quartz sand for the glass. If we were to attempt to count all the people who had some small part in getting an automobile manufactured and delivered to us,

we would soon get lost in the endless maze of interrelationships. We would have to count the retailer who sold it to us, and the carpenters and masons who built his warehouse; the railroad men who helped carry it, the men before them who built the railroad, and the men before them who made the picks and shovels that they used; the men who mined the ore, the men who made the mining machinery, the men who mined the ore to make the mining machinery, the men who made the machinery to mine the ore to make the mining machinery, and . . . Every one of these men, and thousands more, contributed toward the final product. Even if we count only those who actually handled the commodity itself, adding some form, time, or place utility to it, we would find that for most products the number is very great. And it is difficult to think of any commodity in common use today that is not made by the joint contributions of many men. Chronological division of labor is the rule, and the exceptions are few. In primitive times men often made things entirely for themselves. They were self-sufficient. The farmer raised the flax, which was spun and woven in his own household, and he could truthfully say, "I made this!" Of course, even this would usually have been an overstatement, since there was almost always division of labor within the family, and various members of the family group contributed various parts in the production of homemade goods. Moreover, unless he made his own spinning wheel and loom with tools that he made himself, he was using the contributions of someone else. It is a rare man indeed who is really independent and self-made, who does not rely constantly on the assistance, direct or indirect, of others.

10·7. Advantages of Chronological Specialization. Men would never have adopted such a complicated method of production had there been no gain thereby. And the nature of the gain is easy to see. When men specialize in a particular line of work, they become skillful at it. They get to know "the tricks of the trade." One has but to watch a skilled artisan at work to see how true this is. The carpenter knows better than others which tool to choose and how to use it. The butcher learns to know the different cuts of meat and how to separate them easily. The gain in skill makes it possible to produce goods of higher quality and in less time. But the gain in time and the gain in quality, important as they are, are by no means the whole story. If we divide the work of production into minute stages, assigning a different man to each, we can put each man on the kind of work to which he is best adapted, the strong man on the job that needs brute strength, and the man with nimble fingers on the job that needs dexterity. We can adapt the man to the job, and let each man work all the time at things that he does well. Also we can use our equipment more efficiently. If one man does all the tasks, most of his equipment lies idle all the time. While he uses the hammer, his pliers, wrenches, screw drivers, planes, etc., lie idle. If the job of pro-

duction can be broken up into steps, one man can use the hammer constantly, another can use the mallet and chisel, etc. We need but one set of equipment for perhaps 20 men instead of 20 sets of equipment. Each tool works constantly instead of only occasionally. To be sure, specialization never progresses so far that every bit of the equipment is used every moment of the time, but increasing specialization increases the efficiency with which tools and equipment are used. Finally, chronological division of labor has the advantage of stimulating invention. The relationship may not be self-evident in this case, but it is nonetheless real. When a job is broken down into its component parts, as is done with chronological division of labor, it becomes easier to see just what steps need to be taken in production, and often it is discovered that many of the processes are simple ones requiring constant repetition. But these are just the processes to which machinery is adapted and, as soon as the nature of the work is seen, it may become evident that machines can be devised to do the work. It is not by chance that the countries with the greatest chronological division of labor have been those most prominent in invention.

10·8. Disadvantages of Chronological Specialization. Unfortunately, when we cast up this balance sheet we find liabilities that offset, at least in part, the assets. The chronological division of labor brings added monotony to work that was often dreary enough at best. A man finds himself destined to go through the same semimechanical processes over and over again throughout his productive life. He finds, moreover, that his work has become impersonal. No one thinks of the product as being "his" product. He is not responsible for it. He gets none of the pride of workmanship that the skilled artisan gets from a fine piece of work. There are those who argue that chronological division of labor tends to lower wages. Their argument is that such division does promote invention, and thus brings in machines to compete with men; that the machines do the work more cheaply and throw men out of work or reduce their earnings. This argument, however, is but half thought through. Machines are not substituted for men unless they will do the same work more cheaply. No one invests in machinery if it will make his output more expensive. With the machines, more work is turned out with less expenditure of human effort, and as a result there is more for men to consume. Real wages[1] rise. To be sure, the wages of the particular people who are displaced by the machine are lowered, and they take a personal loss, which is society's gain. Society as a whole is better off at their expense. If anyone seriously imagines that the introduction of machinery lowers real wages for society as a whole, all he has to do is to think of what would happen if we were to give up the machinery we have—if we were suddenly to lose all our automobile trucks, tractors, railroad

[1] See Sec. 6·1.

trains, plows, harrows, and factories. Obviously the real wages would drop below the starvation level for most of us in short order. The argument that chronological division of labor lowers wages is entirely incorrect unless it refers to the temporary reduction for the particular individuals who are displaced by the machines. For everyone else, the machine brings higher real wages ultimately.

There are, too, those who criticize chronological specialization on more spiritual grounds, who feel that the worker loses self-respect and artistic sensibility under the system of mass production and factory organization. They argue that machine work is itself not so artistic as handwork, and that, in addition, the attitude of the worker and his whole outlook on life are changed to his disadvantage and to the ultimate disadvantage of society by the factory system. Such charges are very difficult to appraise; there is doubtless something of merit in them. Yet one must keep in mind, as offsetting factors, the fact that the gain in skill and in time which accompany chronological division of labor have given workers far more leisure time in which they can develop these other aspects of their lives if they wish. Working hours are much shorter, and the things available for consumption much more numerous and plentiful, than they ever were or now could be under any system of self-sufficiency.

In fact, it is at this point that we can best decide whether or not the gains offset the losses; for in the case of chronological division of labor one can take his choice of using it or of getting along without it. The man who thinks that the losses from chronological specialization are too great can run his life on the self-sufficient basis, grow his own food, make his own clothes, and get along unassisted. As soon as one thinks what this would mean, he sees that the gains greatly outweigh the losses. Few indeed are the people who would accept the life of self-sufficiency if they had to accept the accompanying penalties. It is important for us to remember that the factory system brings evils, and for us to do what we can to escape them or lessen them; but it is not the part of wisdom to overlook the gains that have been made by the general adoption of chronological division of labor.

10·9. Geographical Specialization. As the name implies, *geographical division of labor* is division by areas or localities. We are all familiar with the fact that certain regions tend to specialize in the production of certain commodities. We all know of grain regions and fruit regions, of areas famous for copper mining or for shoemaking, of territories producing silk or coffee or rubber, of cities noted for insurance or stockyards or photographic apparatus. To be sure, such areas do not specialize in the production of one product to the exclusion of everything else. A wheat-growing state has many people who are not engaged in growing wheat. But we think of a state as specializing in wheat production if a much larger proportion of its man

power and capital is engaged in wheat production than is common in other areas.

In some cases, men specialize geographically from necessity rather than from choice. Men cannot mine copper in every locality where copper is to be used. The natural sources of supply are localized, and the extractive industries associated with them have to be localized. In some other cases, although localization of production is not absolutely necessary physically, no sensible person would suggest that production should be dispersed equally. For example, cotton could be grown in Canada, but the expenditure for greenhouses, fuel, etc., would be so high that no practical businessman would think of embarking on the venture. Thus differences in climate or soil often influence the location of industry, especially the primary industries. In many cases, a given commodity could be produced in many different areas, and the selection of a location is largely a matter of cost. One can grow wheat in Kansas or Connecticut. At different stages in the development of the country wheat has been grown in both states. But the cost of production in the one case is greater than in the other. If men want wheat grown in Connecticut, they will have to work harder and get along with less of it than if they had been willing to use wheat grown in Kansas, where the climate and the soil are better adapted to its culture.

These differences in cost may take many different forms. For example, some industries settle in particular locations in order to get the advantage of being close to market. Others select their location in such a way as to be near their supplies of raw material. In either case, the effort is to reduce the shipping cost. Assuming that the raw materials and the market are some distance apart, one can see that as he moves the location of his plant farther from the source of supply and closer to market, the cost of hauling raw materials will increase and the cost of hauling the finished product will diminish. Whether a location close to materials or close to market will be preferable will depend on whether raw materials or the finished product is the easier and cheaper to ship. If raw materials are bulky or perishable and finished products are light, compact, and nonperishable, it will pay to put one's business close to the source of supply, so that one can ship the finished product in preference to the raw materials. It would be foolish and costly to ship tons and tons of ore to market, only to extract a fraction of an ounce of radium. The tendency is to ship the light product in preference to the heavy one. Similarly, if one is trying to decide where he should set up a bakery, he finds that his bread is very perishable and the ingredients are comparatively nonperishable. It would be foolish to put his bakery in Kansas to be near the wheat supply and ship his bread to market. The tendency is to ship nonperishable products in preference to perishable ones. Thus we see that, when the raw material is bulky and perishable but the finished product

is light and durable, the tendency is for the industry to gravitate toward the source of supply; when, on the contrary, the raw material is light and durable but the finished product is bulky and perishable, the tendency is for the industry to gravitate toward the market. Of course, the businessman may also be interested in nearness to labor supply (especially, skilled labor) or the nearness to sources of power. All these things will influence his cost. By the way, the economist habitually measures distance in dollars and hours rather than in miles. From his viewpoint, two places are close together if people and goods can be transported between them quickly and cheaply; they are far apart if the transportation is slow or expensive. In this sense, the old cliché that "the world is smaller than it used to be" is entirely correct. Every improvement in transportation reduces distance when measured in economic terms.

To be sure, businessmen are often influenced by the factors that we have mentioned when determining the location of a business—the location of natural resources, climate, soil, nearness to market, nearness to material, the availability of power and labor all play a part. But one must also acknowledge that businesses are often started in particular places pretty much by chance. Even so, once having started in a particular place, the tendency is for them to stay there and grow bigger, becoming more and more centralized and localized. This has been called "the momentum of an early start," although perhaps it would be a better parallel to have called it "inertia" rather than "momentum." The reason for it is easy to see. Let us suppose that a business has started by chance in Gurleyville and has developed there. You are now thinking of going into the same business. You will find that if you, too, go to Gurleyville you will get the advantage of having repair industries already at hand, the local bankers will know and understand the business and therefore credit should be more easily available, skilled labor will already be on hand in the locality, you can share in the reputation that has been built up for the local product, etc. Consequently you, and others who are thinking of going into the business, will find it advantageous to go where the business is already established. Thus geographical specialization, once established, tends to increase and perpetuate itself. It would take very definite counterforces to decentralize or move the industry.

10·10. Law of Comparative Advantage. When we were describing the chronological division of labor, we listed a number of advantages and disadvantages. When we consider geographical specialization we find, however, but one main advantage and one main disadvantage. The advantage is that, if we specialize wisely, we can get more product with less effort than we can without specialization. If the resources and climates of the world were uniform and if costs of production were the same everywhere, there would be no advantage in geographical specialization and trade. But man finds

differences in cost whether he likes it or not and, given the differences in cost, specialization and trade bring a real gain.

This we can perhaps demonstrate best by taking a simple but hypothetical numerical example. Let us assume that a farmer in Maine could grow 50 boxes of apples by spending $100 for labor, fertilizer, spray materials, and other expenses. If, on the contrary, he tries to grow lemons, his expenditure of $100 will yield but 2 boxes. In Florida, however, the differences in climate are such that a farmer could grow 30 boxes of lemons with an expenditure of $100, or if he preferred he could grow 20 boxes of apples. We can summarize the situation for ease of reference in tabular form thus:

	Apples	Lemons
Maine	50	2
Florida	20	30

Now let us suppose that the farmer in Maine and the farmer in Florida each has $200 to invest. Will it pay them to specialize, or will it pay each one to produce both fruits? We can test this easily by noting the results. If there is no specialization, but each puts $100 into each fruit, the farmer in Maine will get 50 boxes of apples and 2 of lemons. The farmer in Florida will get 20 boxes of apples and 30 of lemons. The total output is 70 of apples and 32 of lemons. But if they specialize, and the Maine farmer puts his entire $200 into apples while the Florida farmer puts his whole $200 into lemons, the Maine farmer will produce 100 boxes of apples and the Florida farmer will produce 60 boxes of lemons. If we compare the results we see:

Product without specialization:
70 boxes apples, 32 boxes lemons
Product with specialization:
100 boxes apples, 60 boxes lemons

It is evident at once that proper specialization increased the total product. There are more apples and more lemons. The Florida and the Maine farmers can now exchange products, and both be better off. The striking thing is that both of them gain by the transaction. It is not a case of one's gaining from the other's loss.

While we still have the figures of this example fresh in mind, it is important to point out that specialization is not advantageous per se. Suppose the Maine farmer had put all his resources into the production of lemons and the Florida farmer into apple production. Our product then would have been 40 boxes of apples and 4 boxes of lemons, which would be far worse than no specialization at all. Specialization is advantageous only if each area

specializes in the thing that it can produce to advantage, trading for the commodities that would be produced at a disadvantage.

So far the case has been perfectly simple and straight-forward, and no one could well quarrel with it except some prejudiced party who had funds invested already in one of the wrong enterprises and wanted society to support his inefficient enterprise at social expense. In practice, the case can and often does become much more intricate, and it is important for the student to discover that it does on some occasions become advantageous to buy from outside areas goods that could have been produced more cheaply at home. The general tendency is for the student to feel intuitively that this could not be true, and consequently it will bear rather close and careful examination. Again we can best follow the reasoning in terms of a simple hypothetical numerical example. Although in practice there would be many production areas and many products involved, we shall again keep the case simple in our illustration by considering but two areas and two commodities. In fact, let us again compare Florida and Maine in the production of apples and lemons, but this time change cost assumptions slightly. We shall assume that an expenditure of $100 in Florida will yield either 20 boxes of lemons or 15 boxes of apples, while the same expenditure in Maine will yield either 10 boxes of apples or 2 boxes of lemons. Again, we summarize the assumptions in tabular form:

	Apples	Lemons
Maine	10	2
Florida	15	20

This time we see that Florida is the more efficient producer of both commodities. She can produce more lemons or more apples than can Maine with the same expenditure. We would say that Florida has a *cost advantage* in both commodities. In this way our present problem differs from our first one, for in that former example Maine had a cost advantage in apple production and Florida in the production of lemons. Our present example shows Florida with a cost advantage in both commodities. The tendency is for the uninformed student to assume that, under such circumstances, both goods will be produced in Florida and neither in Maine. Such an assumption would be incorrect. Let us look at it first from the standpoint of the man in Florida who can, to be sure, produce either fruit or both. If he invests $100 in apple production, he gets 15 boxes of apples, but he has to give up 20 boxes of lemons, which he could have had if he had invested his funds in the lemon business. He got 15 boxes of apples by giving up 20 boxes of lemons. On the other hand, if he puts his $100 in lemons, he gets 20 boxes of lemons, but

only by giving up 15 boxes of apples which he could have had by putting his money into the apple business. He cannot get either commodity without giving up the other. He cannot plant lemon trees and then decide to raise apples on them, nor vice versa. But suppose that there is some way in which he can get 15 boxes of apples by giving up fewer than 20 boxes of lemons. This would be an advantage to him, and he would be glad to do it. Let us make a note of this conclusion for the minute, and come back to it later. *If the man in Florida can get* 15 *boxes of apples by giving up fewer than* 20 *boxes of lemons, he will find it advantageous to do so.*

Let us now turn to the producer in Maine. His cost situation is very different because of differences in climate or soil, or otherwise. If he invests $100 in apples, he will get 10 boxes, but only by giving up 2 boxes of lemons which he could have produced had he put the same funds into the lemon business. If he invests $100 in lemons, he will get 2 boxes of lemons, but only by giving up 10 boxes of apples which could have been produced had he gone into the apple business instead. He gets 2 boxes of lemons only by giving up 10 boxes of apples. *If the Maine farmer can find any way to get* 2 *boxes of lemons by giving up fewer than* 10 *boxes of apples, it will pay him to do so.*

Let us see, now, what would happen if Florida were to confine her activities to the production of lemons, putting all her resources into that crop, and Maine were to specialize in apple production, and if they were to exchange products at the ratio of, say, 3 boxes of apples for every 2 boxes of lemons. Let us now compare this new ratio of 3 apples to 2 lemons with our two italicized statements. At the ratio of 3 to 2, Florida gets 30 boxes of apples for 20 boxes of lemons, which is far better than the 15 boxes she could get at home; and Maine gets 2 boxes of lemons by giving up 3 boxes of apples, which is also better than the 10 boxes of apples she had to give up at home. *Both parties have gained!* We can figure their gains in another way. Suppose a man in Florida invests $100 in apple production. He gets 15 boxes of apples. But if he invests in lemons, he will get 20 of lemons. Exchanging them at the ratio of 2 lemons for 3 apples, he now has 30 apples by trade instead of producing 15. Either way he spent $100. By producing apples himself, he got 15 boxes; by trading, he got 30 boxes. The gain is apparent. Similarly in Maine, if a man invests $100 in lemon production, he gets 2 boxes of lemons; if he invests the same $100 in apples, he gets 10 boxes of apples, which, at the ratio of 3:2, he can swap for 6-2/3 boxes of lemons. If he produces lemons himself, he gets 2 boxes for his $100; if he raises apples and trades them for lemons he gets 6-2/3 boxes for his $100. Again the gain is apparent. It illustrates the very important fact that often men can get more of one commodity by producing an entirely different commodity and trading. We all realize this when we think about it. For example, the Minnesota

farmer who likes bananas knows enough not to try to grow them. He knows that he can get more bananas by growing potatoes and trading for his bananas. And that is what happened in our numerical illustrative example.

Let us see if we can find the principle that underlies this example. We note that Florida had a cost advantage in both commodities. But Florida was 10 times as efficient a producer of lemons, and only 1½ times as efficient a producer of apples. She had a cost advantage in both cases, but when we compare them we find that she had a *comparative advantage* in lemons—lemons gave the *greater cost advantage.* And we can now *enunciate* the principle that is implicit in the examples; namely, the Law of Comparative Advantage: *Any area finds it advantageous to specialize in the production of those commodities in which it has the greatest cost advantage or the least cost disadvantage as compared with its competitors.* The principle is also sometimes called the "Law of Comparative Cost," and some authors change the wording so that the word "comparative" is substituted for the word "cost" in both places.

In the example just taken, Florida was in the pleasant position of having a cost advantage in both apples and lemons. Yet she found it to her advantage to specialize in lemon production where her cost advantage was greater, and to import her apples from Maine even though they could have been produced more cheaply at home. Maine, on the other hand, had the misfortune to be at a disadvantage in both products. But her disadvantage was smaller in the case of apples, where she was two-thirds as efficient as Florida, than it was in the case of lemons, where she was only a tenth as efficient as Florida. So Maine found it advantageous to select the product with the smaller cost disadvantage. Florida could have raised either fruit profitably, but she could make more profit at lemon culture, so she specialized in that field. She found that she could get more apples and more lemons by specializing and trading than she could get by scattering her energies over both products.

In practice it often turns out that areas import products that could have been produced more cheaply at home. In other words, the Law of Comparative Advantage is not some abstract theory with no practical application. Thus in the United States, corn yields are higher in the cotton belt than in the corn belt. The cotton belt has advantages in both corn and cotton production. But the advantage in cotton production is far greater; the comparative advantage is in cotton. Therefore the cotton belt emphasizes cotton, and the heavy corn-producing area is somewhat farther north where corn can be grown almost as well but where cotton growing would be at a very great comparative disadvantage.

In our discussion of Maine and Florida we found that specialization would increase the total output and give both areas more product. But we

said nothing about the fact that, if the areas are to specialize and trade, they will have a transportation cost that would not exist if each area produced both products. Specialization must not only make a gain in product, but it must make a gain great enough to offset the cost of transportation. The fact that the gain is often large enough to do this is evidenced by the fact that geographical specialization is common; men do find it to their advantage to specialize and trade. In fact, they find it so advantageous that, when governments try by various means to impede or stop such trade for purposes that we shall discuss later,[1] they find that men often run great risks rather than give up the advantage of trade.

The superficial thinker is apt to agree with David Grayson[2] that, "Looking at bargains from a purely commercial point of view, someone is always cheated." We now know two reasons for disagreeing with this statement. First, we realize that utility is purely subjective,[3] so that two men may get entirely different satisfactions from a given good. If Jones dislikes apples and likes lemons, while Smith dislikes lemons and likes apples, and if Jones has a box of apples and Smith a box of lemons, they can both gain by trade. Each will give up something that has little utility and get something that has greater utility. But now we find another way in which both parties to a trade may gain. We discover that, if parties specialize and trade, they may be able to increase the total amount of the output, so that there is more for everyone, and that Smith and Jones have more apples and more lemons by trading than they could possibly have if each tried to be self-sufficient. This point is worth continued emphasis, because it is perhaps less widely understood than any other of the principles of economics.

10·11. At What Ratio Will Trade Take Place? In our numerical example of comparative advantage, we assumed that the Florida and Maine farmers would specialize and that they would then trade their products at the rate of 3 boxes of apples to 2 boxes of lemons. Let us now see where this ratio came from.

If we look back at the original data of this problem,[4] we see that in Maine a farmer had his choice of producing 10 boxes of apples or 2 boxes of lemons, the ratio being 10:2. In Florida the ratio was 15:20—15 boxes of apples to 20 boxes of lemons. These ratios are determined, of course, by the climate, the soil, and other factors which fix costs of production and which are beyond the control of farmers or anyone else in either state.

[1] See Chap. 23.

[2] GRAYSON, DAVID, "Adventures in Contentment," p. 17, Doubleday, Doran & Company, Inc., New York, 1923.

[3] See Sec. 5·2.

[4] See p. 171.

The Maine ratio of 10:2 is, of course, exactly the same as a ratio of 100:20, which can easily be compared with the Florida ratio of 15:20. The expenditure that will produce 20 boxes of lemons in Florida will produce 15 boxes of apples there. The expenditure that will produce 20 boxes of lemons in Maine will produce 100 boxes of apples there. So we have two comparable ratios:

Maine: 100 of apples to 20 of lemons
Florida: 15 of apples to 20 of lemons

It will pay the areas to trade at any ratio between these two; at a ratio of 98:20; a ratio of 75:20; or a ratio of 18:20. In our earlier example, we selected the ratio of 3:2, which is the ratio of 30:20. This ratio lies between our extremes, and therefore will be of advantage to both areas. But we might as well have chosen any other ratio between these two extremes. For example, we might have chosen a ratio of 40:20, which would be 2:1; that is, we might have given our original production figures and then have said that both areas would gain if they exchanged at the rate of 2 boxes of apples to 1 box of lemons. If we go through the same reasoning as before, we shall see that both parties would have gained at this rate. Maine could now get 2 boxes of lemons by giving up 4 of apples, which is better than giving up 10 of apples as she would do without trade. Florida could get 40 of apples by giving up 20 of lemons, while at home she could get but 15 of apples. Both parties would gain by trade, and at any ratio between 100:20 and 15:20. Our general rule is that the ratio of exchange must lie between the ratios of production in the two different areas. In our problem the ratio of production in Maine is 100 of apples to 20 of lemons while the Florida ratio is 15 of apples to 20 of lemons. There is no way to tell exactly what ratio of exchange would exist in practice. As the exchange ratio came closer and closer to 100:20, Maine would gain less and less by trade and Florida would gain more and more. On the other hand, as the ratio approached more and more closely the Florida ratio of 15:20, Florida would gain less and less and Maine more and more. The particular ratio that was chosen would depend on which area drove the better bargain. The problem is indeterminate except to say that the ratio at which trade is actually carried on will lie between 100:20 and 15:20.

There is no advantage in trade between areas if their costs are the same, or even if their costs are in the same ratio. If Florida is twice as efficient as Maine in lemon production and also twice as efficient in apple production, trade has no advantage. But when the ratios of advantage differ, then the ratio at which the commodities actually exchange must be somewhere between the production ratios in the two different areas. From the standpoint of either area considered alone, the advantage of trade becomes greater as the

trading ratio differs more and more from the ratio at which the goods can be produced at home.

10 · 12. Disadvantage of Geographical Specialization. We have seen that geographical or territorial division of labor has a very real advantage in that it increases total productivity and gives consumers more goods. But it has an accompanying disadvantage in that it makes men dependent on other people, and other people are not always dependable. If Florida decides to specialize in lemon production and plants no apple orchards, she will run into difficulty if her trade with other areas is broken off. If a washout on the railroad, a strike on the steamship lines, a war, or something else interferes with trade, Florida will find herself without any apples and with no orchards from which to get them; and she will find herself with a great surplus of lemons that cannot be used. If we could be sure that trade would always continue, we would not have to worry about this disadvantage, but experience shows us that trade interruptions have not been uncommon historically. In domestic trade they have come from natural catastrophes usually and have been of relatively short duration. In international trade, interruptions have commonly been phenomena of war and sometimes have lasted for many years. They have also arisen from the commercial policies of the nations involved. A specialized area has sometimes discovered that its former customers are cut off by tariff barriers or trade quotas, so that it can no longer sell and consequently is not in a position to buy. These matters we shall discuss later when we consider the problems of international trade.[1]

Any area discovers, then, that there are certain gains to be made from trade while trade is continued but that, if the area specializes and trade is interrupted, there will be certain losses. The problem is whether the gains during the time of trade more than offset the losses when trade is interrupted. If interruptions are uncommon and if the commodity that we import is not absolutely necessary, so that we can dispense with it temporarily without too serious effects, then we are likely to decide that the gains of trade in ordinary times are too great to be foregone. If, on the contrary, interruptions are common or if the commodity is so necessary that we cannot afford to run the risk of being without it even temporarily during an interruption of trade, we may decide to produce it at home even though we lose in ordinary times by doing so. The matter is one of judgment, and one on which even sensible people can disagree. But that there may be a gain from specialization and trade when trade is carried on is not a matter of judgment but a fact.

10 · 13. Factors of Production. In this chapter we have been describing the nature of production and some of its problems. We should now point

[1] See Chap. 23.

out that things do not produce themselves, but that they are commonly produced only by the cooperative efforts of many people and agencies. These have already been mentioned in this chapter, but the time has come to draw them together in more formal and orderly fashion. For example, when we were talking about the primary industries, we pointed out that all the raw materials for every commodity have to come originally from nature. A secondary industry may buy its raw materials from another secondary industry, but some time back in the chain of preceding events these materials came from natural resources. There could be no production of goods of any kind without natural resources. They are basic to production. The economist, for some reason difficult to explain, calls these natural resources "land," and he would say that land is one of the "factors of production" or "agents of production." By this he would mean that production cannot be carried on without the use of natural resources—of land. But natural resources are not ordinarily sufficient. We also pointed out earlier in this chapter that most things that men want do not exist at all in nature or at least not in sufficient quantities to satisfy men. Therefore, we said, if men are to get what they want, they have to work. We must have the expenditure of human effort, or we get no radios, no speed boats, and no alarm clocks. So the economist would recognize "labor" as a second factor of production and would include, as labor, not merely unskilled labor but all human effort used in production.

In some cases perhaps, commodities are produced by land and labor alone, by human effort expended on natural resources without the help of any other factor. But usually men find it advantageous to use tools, machinery, and equipment of various sorts to help in production. Such tools and equipment multiply the output tremendously, as we can see if we consider the output of a farmer working his field with his tractor and mowing machine and other tools, and contrast this with what he would produce if he worked unassisted (merely land and labor) without his tractor or mowing machine and even without a rake, hoe, or spade—just himself with his bare hands and his elbow grease. The tools and equipment with which men work are called by the economist "capital," and we would list it as a third "factor of production."

No attempt is made here to define these factors carefully and precisely, but merely to list them. At present we shall content ourselves with the statement that most productive effort requires the joint contributions of natural resources, human effort, and tools—of land, labor, and capital. The next three or four chapters will be given over to study of these factors of production, or agents of production, individually. There we shall define them more precisely, and there we shall become better acquainted with the major problems connected with each of them.

10 · 14. Proportionality of the Factors. Although it is true that many factors of production are required to produce almost every commodity, nevertheless the process is by no means one of mixing factors at random. Most recipes in the cookbook require a number of ingredients, yet if the cook merely mixes materials at random the result is not apt to be palatable. Similarly, a random mixture of productive factors is not likely to yield a profit. The businessman in any field of production needs to know the techniques applicable in his line of business. He needs to know what raw materials are needed and where they may be had at reasonable prices. He needs to know how many workers to hire, and what skills and training they need. He needs to know what kinds of tools and equipment are necessary. In other words, he needs to know what factors to use and in what proportions to combine them. The amount of profit (or loss) that he makes will depend largely on his skill in making decisions in this field.

10 · 15. Factors Affecting Profit. A businessman is usually in a position to decide some things about his business, while other things are beyond his control. For example, the farmer or manufacturer is ordinarily able to decide for himself how much output to produce and put on the market. He can grow 5 acres of melons or 1 acre or none at all. He may keep 12 cows or 30 cows. He may produce 1000 pairs of shoes a month or 5000. The amount of output is under his control, and he may adjust it in whatever way seems to him desirable, which, in practice, means that he will ordinarily try to adjust it in such a way as to get the largest possible profit. Usually the businessman is not able to determine his own costs. The farmer may be able to decide for himself how many acres of melons he will raise, but he cannot arbitrarily decide that he will buy fertilizer at $1 a ton or 10 cents a ton, that he will hire labor at 10 cents an hour or 10 cents a month. The prices of the things that he buys are usually fixed by forces beyond his control. Sometimes it is true also that the prices of the things that he sells are beyond his control. Thus the wheat farmer cannot set the price of wheat at $10 per bushel in any effective way. To be sure, he can refuse to sell his own wheat at any price below $10 a bushel, but in such a case he will sell no wheat and the price of wheat in general—the market price—will not be appreciably affected by his decision. He is too small a part of the production system to play any effective part in price setting even for his own products. He takes the market price as he finds it. On the other hand, the manufacturer of a patented product may be able to set the price for it wherever he likes, at least within wide limits. Even in such a case as this, however, he will not be entirely a free agent. If he sets the price very high, consumers will not use his product; if he lowers the price, he will be able to increase the amount sold. This matter we discussed when we were considering problems of demand in Chap. 6.

The producer, then, can usually set his volume of output wherever he wishes, but he is influenced by the conditions of cost and by the demand for his product, both of which are beyond his control. What it will be profitable for him to do will depend on the costs and demand that he finds ready made for him. His problem is to select that particular volume of output which will maximize his profit and yield for him the largest possible net income. As we shall see, this does not mean that he will necessarily choose the largest possible output, nor that he will curtail output to the smallest possible amount. Sometimes the greatest profit can be made by producing a large output, sometimes by producing a small one. In any particular case the producer will have to find the point of highest profit by experimentation, by varying his output to see what happens to his profit.

10·16. Fixed and Variable Factors. In practice and within limits, the businessman usually finds that, when he wishes to alter his volume of business, there are some factors of production that he can change in quantity easily and some that can be changed only with difficulty or not at all. Thus a farmer may own a farm, and his investment in it may be fixed. He may be able to hire 2 men or 10, and he may be able to vary the amount of fertilizer applied or the number of times the fields are cultivated, but perhaps he is unable to vary the amount of land. Similarly a manufacturer may have a fixed investment in a factory and machines. He can employ more workers if he likes and run two or three shifts, or he can cut down on the number of workers and leave half the factory and machines idle; but perhaps he is in such a situation that his changes in output will be made, within reasonable limits, without any change in his investment in the factory itself or in the machines. Factors of production that are changed in quantity in order to get changes in output are called "variable factors," while those which stay unchanged as the output changes are called "fixed factors." One man may, of course, vary his output by varying the number of men, while another varies the amount of capital. No factor is fixed for everyone. In any particular case, those factors which are altered in amount to get changes in output are variable and those which are left unaltered are fixed factors.

Sometimes it may seem at first sight as though a businessman had no fixed factors and could alter them all. Thus, for example, a farmer who wants to raise more potatoes may be able to hire more men if he wishes, perhaps he can rent more land if he desires, he can buy more fertilizer if it seems advantageous, and in fact he seems to be able to alter all the factors. In this case we are overlooking the fact that he is himself a fixed factor. There are limits to his own managerial abilities and, as he tries to supervise more and more men and more and more machinery and land, these other factors will be varied while he is fixed. Often in practice the most effective limitation on the size of a business is the managerial ability of the businessman. Thus

every business has some one or more fixed factors and some one or more variable factors, and the problems of management center pretty much around the decision as to how much of the variable factors to apply to the given fixed factors.

10 · 17. Diminishing Productivity. One might suppose that the differences between businesses would be so great that each would be a law unto itself, yet in fact we discover some common tendencies that seem to underly all sorts of business enterprises; especially, we find the same sorts of things occurring constantly as we alter the volume of the output. The changes can be most easily understood by means of a simple numerical illustration. Without trying to make our figures jibe particularly with those which one would get in any actual case, let us take a purely fictitious example such as that of Table 10 · 2. Here we assume that a farmer is trying to vary the amount of potatoes that he raises on a given field. How can he get more potatoes from the same field? Perhaps by putting on more fertilizer, by cultivating oftener, by spraying more thoroughly, or by taking more care that every potato is recovered at harvesting, etc. In any of these cases he will have to incur a greater expense to get more potatoes. He will have to use more of the factors of production—more labor or capital, since we are assuming that the amount of land is fixed. As he hires more men, let us say, in an effort to get more potatoes from his field, what happens to his output? We shall assume that his output varies in accordance with the figures in the first two columns of Table 10 · 2. If he puts no labor into potato production at all, he will get no crop at all. If he puts the time of one man into potato production, he will get 50 bushels, etc. We note that, in general, he gets more product each time he adds another man, although presumably this would stop somewhere. If he put too many men on this field, they would get in each other's way and reduce the output. Although no farmer in his right mind would do such a thing, we have shown this possibility in our table by assuming that the thirteenth man actually reduces the total output.

Our first principle, then, which we have illustrated in the first two columns, is that as the entrepreneur (a name commonly given to the man who makes the business decisions—the manager) increases his expenditure for factors of production, he ordinarily increases his output, although it would be possible for him to go so far as to reduce it.

The last two columns of Table 10 · 2 are derived from the first two. The third column is found by dividing each figure in the second by the corresponding figure in the first column. If, for example, the farmer hires 6 men, he gets 570 bushels of potatoes. Dividing, we find that this amounts to 95 bushels per man. These figures we call the "average outputs," or "average products," or "outputs per unit of input." In our particular case the outputs were potatoes and the inputs were men, so our results are in terms of bushels

of potatoes per man; if we use 6 men, we get 95 bushels per man. We note that in our example the output per man rises for a time, reaches a peak when we hire 5 men, and thereafter falls again. The point represented by 5 men in this example is called the point of "highest average output." Of course, if we were given the first and third columns, we could find the second column by multiplication. Thus if the average output is 83 bushels when we hire 10 men, the total output is 83 times 10, or 830 bushels.

The last column of Table 10·2 is found by subtraction. To get any particular figure in this last column, we subtract from the corresponding figure in the second column the figure above it. For example, if we want to

Table 10·2. Illustration of Changes in Output Accompanying Changes of Input

Number of men employed	Number of bushels of potatoes	Average output, bushels	Marginal output, bushels
0	0		
1	50	50	50
2	150	75	100
3	270	90	120
4	380	95	110
5	480	96	100
6	570	95	90
7	651	93	81
8	720	90	69
9	783	87	63
10	830	83	47
11	858	78	28
12	864	72	6
13	858	66	−6

know the marginal output when we are employing 8 men, we note that their total output is 720 units while the total output of seven men is 651 units. Subtracting 651 from 720 we get 69 bushels, the marginal output of eight men. In fact, *the marginal output obtained from any particular quantity of a variable factor is the difference between the total output obtained when we use that quantity of the variable factor and the total output that we would obtain if we used one less unit of the variable factor.* Since in our example the variable factor is the number of men, we can say that the marginal output of any number of men is the difference between their total output and the total output that would be obtained if we used one less man. Now we come to the other principles illustrated by Table 10·2; namely, if we keep one or more factors of production fixed and vary the others, we sooner or later reach points beyond which the marginal and average outputs start to di-

minish. They may in some cases diminish right from the start. On the other hand, it is possible that early applications of variable factors may increase efficiency and bring larger marginal and average outputs. But the important thing is that, sooner or later, the marginal and the average outputs start to fall and continue to fall, until ultimately even the total output starts to fall, as when we use 13 men in the table.

This tendency, as far as we know, is common to all businesses and industries. There are no known exceptions to it. In fact, the reader will realize that it is true by considering the alternative. If this tendency did not operate, if, on the contrary, we could add more and more fertilizer to a field and continue to get larger and larger marginal and average outputs, then it would pay men to select the one most fertile acre of land in the world and grow all their crops on it. If we could get enough potatoes from the 1 acre to feed one family, then by doubling the fertilizer and labor we could get enough for two families, and by increasing our inputs sufficiently we could feed a thousand or a billion families from the single acre. Immediately we realize that this would not occur. Sooner or later as we tried to increase our output from this fixed piece of land, we would become less and less efficient. Our marginal and average outputs would drop.

This tendency is described in what we know as the "Law of Diminishing Productivity": *At any given state of the arts, if we add more and more of any variable factors of production to fixed amounts of other factors, we reach points beyond which the marginal, the average, and the total outputs diminish.* To be sure, if we find new strains of crops or new methods of production; that is, if we change the state of the arts, it may be possible to produce more than we used to without diminished efficiency. But then, under the new state of the arts, it will still be true that some new points will be reached beyond which the marginal, average, and total outputs will diminish. A change in the state of the arts might change every figure in Table 10·2, so that the maximum points no longer occurred at the same places. If we can judge by past experience, it is safe to assume that new maximum points would appear, since we have never known any exception to this general tendency in the past.

The Law of Diminishing Productivity means, then, that an entrepreneur cannot change his use of the factors of production at will without affecting their efficiency. If he uses more and more of any factor, the marginal output or the average output associated with that factor drops. But, just as surely, the average output of the fixed factors rises. Thus, if the field that we kept fixed for the purposes of Table 10·2 contained 1 acre, the yield per acre continued to rise long after the yield per man started to fall. The highest yield per man came when we hired 5 men; but the largest yield per acre came when we hired 12 men. One factor becomes more efficient as another becomes less effi-

cient. The entrepreneur does not try to get the highest yield per acre, or the highest yield per man, or the highest yield per dollar of capital invested. Almost any farmer in the United States could get higher yields per acre if he wanted to. But greater efficiency can be obtained in any one factor only at the expense of reduced efficiency in others. The entrepreneur cannot pick out some one factor and increase its efficiency without regard to the others. He must consider them all together, and try to select the particular combination of factors that will yield the largest possible profit. This will not usually be the combination that gives the greatest possible output from any one of the factors considered by itself, but, as we shall see in Chap. 14, will ordinarily be some compromise.

10·18. Summary. Production consists in the creation, not of matter, but of utility. Most productive activities center around changing the form of things, holding them from a time of plenty to one of scarcity, or transporting them from a place of surplus to one of deficit. Any of these activities add utility to goods, and they are therefore called productive. The economist also recognizes the fact that many activities are productive without being connected with physical objects at all. These productive activities he calls services.

Industries that derive their raw materials directly from nature are called primary industries and are subdivided into genetic and extractive industries according as they do or do not replace the materials that are used up. Those industries which derive their raw materials from other industries are called secondary industries. Genetic industries and the societies based on them tend to be more stable than extractive industries and the societies based on them.

Modern society is characteristically one of specialization. The nature of modern production cannot be understood unless one understands its cooperative aspects—its maze of complicated interrelations. In the production of almost every product one sees the combined efforts of many men, each carrying out his own small part of the task. Such chronological specialization brings with it major advantages in the way of increased efficiency, and concomitant social and spiritual disadvantages. Overlying this pattern of chronological specialization is also a pattern of geographical division of labor. Differences in environmental factors make it profitable to specialize in the production of different goods in different places, and it commonly turns out to be true that the people of a given region can have more of a product to consume if they do not try to produce it, but rather obtain it through trade.

Production in modern society is, for the most part, carried on by firms that are seeking a profit. Goods are not usually produced for the use of the maker, but for sale to others. In producing these goods it is necessary to have the cooperation of many agents, which we classify under the broad headings of land, labor, and capital. These agents can be combined in various propor-

tions but, as a businessman increases the amount of the variable factors that he combines with given quantities of the fixed factors, he runs into the phenomenon that we call diminishing productivity. Increased use of variable factors ordinarily increases output, but before long the marginal output and then the average output begin to drop. Our later study will indicate how the businessman determines the best stopping place—the most profitable combination of the factors of production.

Questions

1. Which of the following would you class as productive in the technical sense of that term: a farmer, the keeper of a retail store, a prize fighter, a big-league ball player, a newsboy, a boy who mows his neighbor's lawn for hire, a boy who mows his father's lawn, a man who mows his own lawn, an opera singer, a man singing to himself for his own enjoyment, a highwayman hired by a syndicate of criminals to hold up a mail train, a physician, a worker in a steel mill, a grammar-school boy making a birdhouse? If there is any question about any case, what further information would you need in order to decide how to classify the person?

2. How many kinds of utility can you identify in your fountain pen?

3. Name three or four primary industries; three or four secondary industries; three or four extractive industries; three or four genetic industries.

4. In the case of exhaustible resources, even a very low rate of use will bring ultimate exhaustion. Under such circumstances, what considerations would influence you in deciding what is the best rate of use? Discuss.

5. What are some of the problems that arise in a society characterized by production for profit which do not arise in a society characterized by production for use? To put the question in other terms, what new problems appear in a society that changes from self-sufficiency to specialization and trade?

6. Can you name any products in which there is no chronological specialization in production—ones completely produced by one man without the aid of others?

7. Name some products with bulky raw materials and compact finished products. Some with compact raw materials and bulky finished products. Some with perishable raw materials and more durable finished products. Some with relatively durable raw materials and more perishable finished products. What is the effect in each case on the location of the industry producing the product?

8. Suppose that an expenditure of $1000 will produce either 2000 watermelons or 10,000 loaves of bread in Paris. In Berlin the same expenditure will produce, let us say, either 1500 watermelons or 5000 loaves of bread.

   a. Which country has the cost advantage in the production of melons? Which in bread?

b. If the areas specialize and trade, which area will produce melons? Which bread?

c. Which of the following ratios of exchange would make it profitable for the two areas to specialize and trade?

1 melon to 4 loaves of bread.
1 melon to 8 loaves of bread.
1 melon to 1 loaf of bread.
6 melons to 21 loaves of bread.
6 melons to 28 loaves of bread.
6 melons to 38 loaves of bread.

9. Describe two ways in which both parties to a trade may gain.

10. Name a commodity that is produced by land alone, without the use of labor or capital.

11. Suppose that a farmer is raising cabbages. What might some of his fixed factors be? Some of his variable factors?

12. Suppose that a man operates an automobile repair shop. What are some of his fixed and some of his variable factors?

13. State in your own words the tendencies illustrated in Table 10•2.

14. Suppose that in some table similar to Table 10•2 a man is getting an average output of 550 bushels and a marginal output of 400 bushels. Do you think that the average output will rise or fall when the man increases the volume of his business? Why?

11. NATURAL RESOURCES

11 · 1. Land Defined. As was pointed out at the end of Chap. 10, economists have fallen into the habit of using the word "land" as synonymous with all natural resources. The entire natural environment can be divided into three categories: those things which are beneficial to men and assist them in satisfying their wants; those things which are neutral, giving neither advantage nor disadvantage; and those things which are harmful. Only the first of these groups is included among natural resources. In other words, natural resources are useful parts of the natural environment. This means that, although the sum total of the things in the natural environment may be constant, the natural resources are not constant. As men discover new wants and new uses for materials, things that were once useless become useful; they become natural resources. Thus, there was a time when men knew no use for rubber. At that time, and for men, rubber trees were not natural resources. Or, to take a better illustration, the coming of the white men to what is now the United States actually increased the resources of the area. The physical materials had been here before, but the residents of that time knew no use for them. Today we list among our resources petroleum, bauxite, helium, and tungsten. These things existed in the United States before the coming of the white men, but they were not then resources. Men had not been ingenious enough to make them resources. Resources depend on human ingenuity and understanding as well as on physical characteristics. Strictly speaking, materials that exist in the natural environment do not become a part of the natural resources of the community until they are discovered and a use has been found for them. If undiscovered or useless, they have no influence on economic activity and cannot be included among the factors of production.

Even after we limit "land" in this way, we have a good deal of trouble often in being sure just what is land and what is not. For example, it might be natural to suppose that a farm is "land," that it is a natural resource. But a moment's reflection makes it clear that any typical farm includes many things that nature did not provide, improvements made by men. The land

may have been cleared of trees and stones, drained or irrigated, fertilized artificially, etc. In trying to make this distinction, David Ricardo, English economist of the first quarter of the nineteenth century, spoke of the "original and indestructible powers of the soil." Today we are less certain that the powers of the soil are indestructible, but we still try for purposes of analysis to distinguish the contribution of men from the contribution of nature, and we would say that an ordinary farm embodies contributions of both. Strictly speaking, land is a bounty of nature—a gift of the gods rather than the result of human planning and toil.

When we consider that *land includes all natural resources that are used in production,* it becomes immediately evident that the concept is a broad one. Not only do we include farm and city "land" as the term is used in common speech, but include such additional resources as the oceans, seas, and rivers; the climate and the rainfall; the topography and minerals; water power; and all those other features of the natural environment which influence men's economic activities. In the strange jargon of the economist, Niagara Falls is land, a stream of water in irrigated territory is land, and a flock of wild geese is land (at least in a society of hunters).

11 · 2. Quantity of Land. Land is peculiar among the factors of production in that its quantity is but little subject to conscious and planned variation. To be sure, resources change as we discover new uses for materials, or new techniques of production. From the economic standpoint, if we discover new strains of crops or new cultural practices that double the yields of crops, we have in effect doubled the amount of cropland. If we find ways to utilize ores that could not be used before, we have in effect discovered new ore deposits. But such discoveries are more or less haphazard. A nation can follow policies that will tend to increase or to decrease its population or its capital equipment; but any given time a nation has to take its natural resources about as they are. It may try, of course, to extend its own territory by conquest or colonization, and thus to gain control of new resources. The political control of resources can be and often has been altered radically. But the resources available within any given region change slowly and sporadically and unpredictably; they are much less subject to manipulation than are the human resources or the capital resources.

The general tendency of most societies has been for slow increases in population and capital, with the land approximately fixed. Under these circumstances the natural resources tend historically to become slowly more scarce, and consequently to raise more and more pressing problems. The tendency is most evident, of course, in the case of exhaustible resources, since they decrease in quantity not only relative to population, but absolutely. In newly settled territories it is often true that most kinds of land are, for practical purposes, free goods to be had for the asking. Under such circum-

stances the major problems are those of getting sufficient labor and capital to exploit the resources. As the population grows and capital accumulates, however, the situation slowly changes, until in some of the older countries labor has become almost (not quite, to be sure) a free good, while land is extremely scarce and very valuable. On the one hand, we have areas that badly need man power, encouraging immigration in an effort to make use of their wealth of natural resources; on the other hand, we have areas that are heavily overpopulated, looking for colonies and *lebensraum.*

The history of the United States offers, at different periods, illustrations of both situations in some degree. From the time of the first settlement by white men until some time late in the nineteenth century, men thought of our natural resources as plentiful—almost limitless—to be exploited. Then, as the frontier disappeared and as population continued to grow at an increasing rate, the attitude changed to one of concern over natural resources, and we saw the beginning of what was called the "conservation" movement under President Theodore Roosevelt.

11·3. Economic Properties of Land. Land has many properties and characteristics that are important to the chemist, the agronomist, the engineer, etc. From the standpoint of the economist, however, those qualities of land are important which fit it to serve in production. Lands vary so that it is quite impossible to list here all the things for which they are used and which give them value; but we can list some of the commoner classes of things that give value to land in an economic sense. These we call the "economic properties of land."

Unquestionably most important of all is the mere *extension* of land, the fact that it gives standing room and support. This by itself, however, would not get us very far. Physical support is but a small part of economic support. Among the economic properties of land, surely one of the most significant is *soil fertility,* the ability to raise crops. Lands differ markedly in fertility, of course, and soil that is well fitted for the production of one crop may not be at all useful for another. Yet taken all together, a much larger part of the natural resources is used for crop production than for any other economic purpose. Men are willing to pay for fertile soil if they have to, because of its productivity; and, other things equal, they will pay more for land the more fertile it is. It is not by chance that farm land in Iowa and Nebraska is worth more than in upland Connecticut or the Berkshire hills of Massachusetts, even though these latter areas are closer to the great market areas. The differences in soil fertility and the resultant differences in crop yields are responsible for the difference in value.

Yet some very valuable land is not particularly fertile. Other factors have their share of influence on land values. A sterile desert area may be very valuable if one discovers in it deposits of gold or silver or if one strikes

oil on it. Thus the *mineral content* of land influences its value, whether the mineral be gold or gravel, silver or sand. Some land, also, is valued for its *power resources.* Many of the early settlements took advantage of water power, and even today the great sources of hydroelectric power are exceedingly valuable. Power is commonly a natural resource and hence listed as land by our definition.

It would be difficult, however, to explain some of the most outstanding examples of high land value on the basis of soil fertility, minerals, or power. Land in the center of a large city may be worth hundreds or thousands of dollars a square foot, yet no one supposes that this price measures the soil

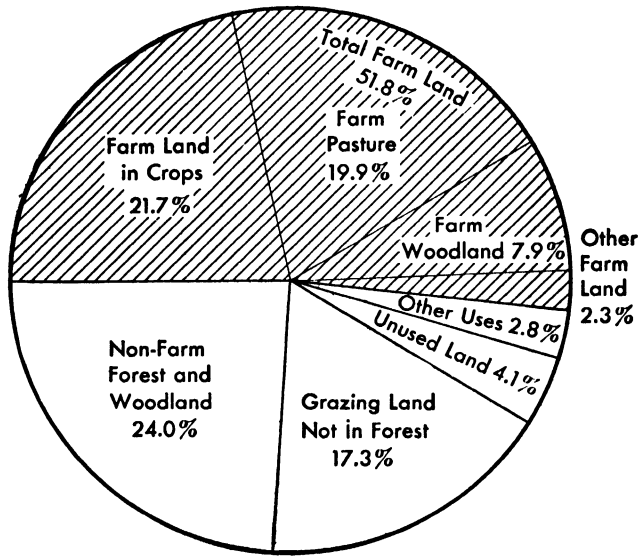


FIG. 11 • 1.—Land Utilization in the United States, 1930. (*Based on Report of the National Resources Board,* Dec. 1, 1934, p. 109.)

fertility, the mineral content, or the power possibilities. The economist says that such land is valued for its *location* or its accessibility. Railroads, good harbors and steamship lines, airports, highways, canals, trolley lines, subways, any sort of transportation tends to make an area more accessible and to improve its location. When we make it easier to bring goods in and to take them out, we make land more valuable. Most large cities are situated at seaports or on navigable streams. On any navigable stream of importance one usually finds a settlement at the head of navigation. Modern cities have tended to spring up at the junctions of railroad lines. The opening of the Erie Canal and the extension of the railroad network in the United States did a great deal to influence the centers of population and the values of land.

11·4. Land Use in the United States. Ever since the original settlement of the United States, the principal use of its land has been agricultural. Today over half the total area of the country is in farms and almost half the total farm area is in crops. A study of the major uses of land in the United States in 1930,[1] with estimates of the probable uses in 1960, is summarized in Table 11·1 and Fig. 11·1. Farm land and such semifarm uses of land as

Table 11·1. Major Uses of Land in the Continental United States, 1930, and Estimated Uses, 1960*

Class of land	Percentage of total area	
	Actual, 1930	Expected, 1960
All land in farms	51.8	50.0
Cropland	21.7	23.3
Crops actually harvested	18.9	20.3
Crop failure	0.7	0.7
Idle cropland	2.1	2.3
Pasture land	19.9	18.2
Plowable	5.7	4.1
Not plowable	14.2	14.1
Woodland	7.9	6.5
Pastured woodland	4.5	3.6
Nonpastured woodland	3.4	2.9
Other farm land	2.3	2.0
Land not in farms	41.2	50.0
Forest and woodland	24.0	24.8
Grazed	13.1	13.7
Not grazed	10.9	11.1
Other grazing land	17.3	16.0
Public	10.7	9.7
Private	6.6	6.3
Nonagricultural, nonforest use	2.8	6.4
Lands of little or no use	4.1	2.8

* National Resources Board Report, Dec., 1, 1934, p. 109.

forestry and grazing accounted for 93 per cent of the total area of the country in 1930 and are expected to use 91 per cent of the total area in 1960. In contrast, the land in urban use took up but 0.7 per cent of the area in 1930, and is expected to take up the same proportion in 1960. Yet these comparisons are by no means the whole story. Although the agricultural and allied uses of land in 1930 occupied some fourteen times as much territory as did the nonagricultural uses, nevertheless some of the land used for the latter purposes was very valuable and on it lived a very considerable

[1] National Resources Board Report, Dec. 1, 1934, p. 109.

part of the population. The United States is not quite the largest nation of the world in area, but it has a wider range of natural resources, soils, and climates than any other country.

11·5. Mineral Resources of the United States. Roughly a million men and their families are dependent on mining for a living in the United States, turning out in ordinary years between 5 and 10 billion dollars' worth of product. The United States is by a wide margin the largest producer and the largest consumer of minerals in the world. Approximately two-fifths of the value of the world's mineral production comes from the continental United States and, if we add the products of her possessions, the value is raised to about 50 per cent—as much as all the rest of the world put together. No other area is endowed with such a great quantity or such a wide variety of minerals. If we add the natural water power to the mineral sources of power, we find the United States producing nearly half the mechanical energy of the world.[1]

The importance of minerals is a modern phenomenon, with the major use confined almost to the memory of living man. The recent increase in the use of minerals has been vividly depicted by Leith[2] in the following passage:

> The quest of iron and copper and flint for use as weapons, and of gold and silver and precious stones for adornment and art, runs far back into history and is associated with many stirring events of exploration and war. But minerals were used on a relatively small scale, even as late as the French Revolution, and constituted only a minor factor in the environmental conditions influencing human activities. With the advent of the industrial revolution in Great Britain, a little more than a century ago, began the real exploitation of earth materials in a way to influence essentially our material civilization. In this short time, at an ever accelerating rate, minerals have become a fundamental basis of industrialism, to be ranked with soil, climate, and other major influences on our activities. Few realize the sweeping range of this change—the fact, for instance, that in a hundred years the output of pig iron, copper, and the mineral fuels has increased a hundred fold; that more mineral resources have been mined and consumed since the opening of the century than in all preceding history; that in the United States more minerals have been mined and consumed in the last twenty years than in its preceding history; that the per capita consumption of minerals in the United States has multiplied fifteen times in forty years; that the world production of several essential minerals has been doubling about every ten years. The last twenty years have seen as much world gold production as the four hundred years following the discovery of America. A single Lake Superior iron mine now produces every

[1] *Ibid.*, p. 391.

[2] LEITH, C. K., "World Minerals and World Politics," pp. 3-4, Whittlesey House, McGraw-Hill Book Company, Inc., New York, 1931.

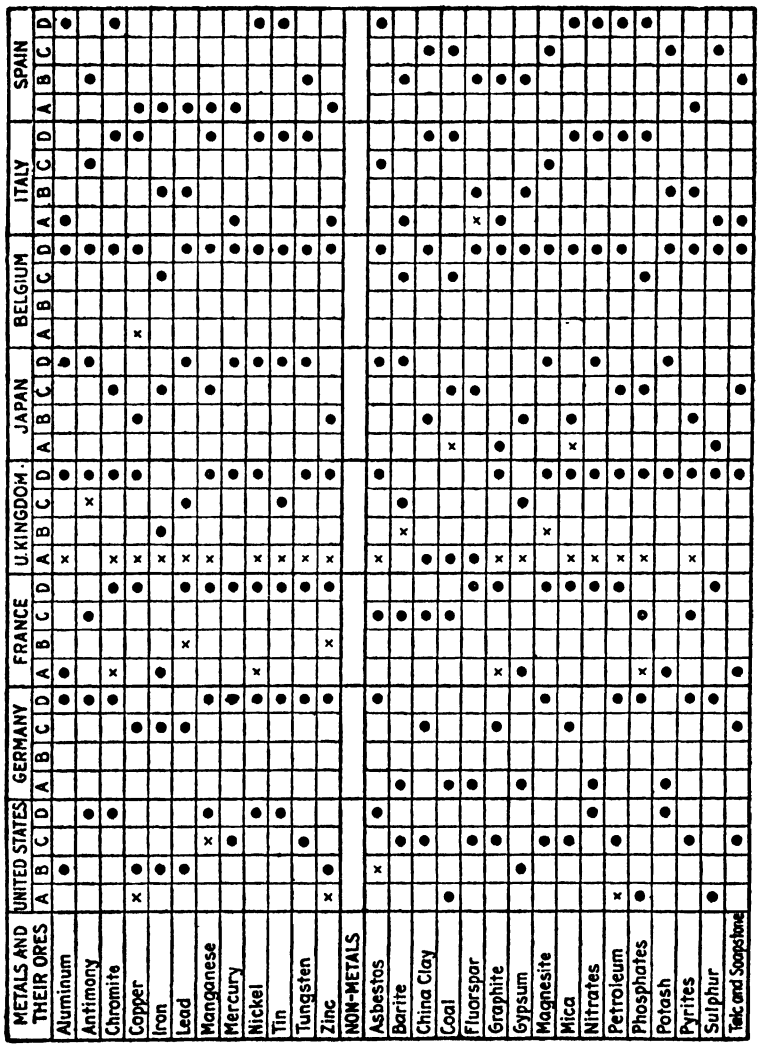
two weeks a volume of ore equivalent to the Great Pyramid of Egypt, which required the toil of vast hordes for several decades and has been long regarded as one of the most stupendous works of man. In 1929 the United States produced more zinc than all of the world did in the first fifty years of the last century. The copper production of the world in 1929 was more than twice as great as the estimated production for all history up to the nineteenth century, and the United States figure for the same year was greater than all the production of the country up to 1888.

Machine production is absolutely dependent on the metals; the metals, in turn, are absolutely dependent on the fuels. An area that combines rich metallic ores close to abundant coal supplies is certain to be a leading industrial section. The Lake Superior iron ore deposits of the United States are of enormous size, largely so near the surface that the ore can be taken out with steam shovels operating on the surface. They are also very rich, with an average metallic content of between 50 and 60 per cent. Over 40 per cent of the iron production of the world is in the United States, which also mines about a third of the world's coal. Great reserves of both minerals are known to exist. So important are they in industry that coke and coal have long led all other commodities in tonnage hauled on the railroads, about 37 per cent of the total; iron, iron ore, and steel come next, with about 10 per cent. All agricultural products together were next in importance with about 8 per cent, followed by forest products with about 6 per cent. Next came two more groups of minerals: petroleum and its products with about 4 per cent, and gravel, sand, and stone with another 4 per cent. Thus we see that the minerals, among them, provide well over half of the frieght tonnage.[1] In value of output, coal would rank among the first half-dozen commodities in the country.

Of much more recent, but rapidly growing, importance among the minerals is petroleum. The development of the internal-combustion engine and the subsequent development of the automobile and the highway system have brought petroleum and its products to a place of outstanding industrial and commercial importance. Petroleum is now recognized as a key resource in wartime. When, on May 4, 1945, German Field Marshall Karl von Rundstedt explained to the Associated Press the reasons for Germany's defeat as he evaluated them, he put the Allied air power first, but second place he assigned to the lack of fuel, both oil and gasoline. (It is of interest to note, by the way, that a large proportion of the reasons assigned by von Rundstedt were economic. After the air superiority of the Allies and the lack of fuel, the other things that he mentioned, in order, were destruction of the German railroad system, the loss of raw material areas such as those of

[1] See "Freight Commodity Statistics," Bureau of Statistics, Interstate Commerce Commission, Washington, D. C.

Rumania, and the smashing of the home industrial sections such as those of Silesia and Saxony.) The United States is outstanding in petroleum production, ordinarily accounting for about 70 per cent of the world's output.



A, Minerals available in large quantities for export. B, Minerals adequate to meet domestic demands without appreciable excess or deficiency. C, Minerals inadequate to meet domestic demands, partially dependent on foreign sources. D, Minerals for which the country depends almost entirely on foreign sources.

FIG. 11 • 2.—Position of certain countries with regard to minerals, 1934. (*Courtesy of C. K. Leith, "World Minerals and World Politics," p. 73.*) *Dots show country's domestic situation. Crosses show situation including resources outside its own boundaries but under its commercial or political control.*

Taken all in all, the United States is the only country in the world that has most of the major industrial minerals in adequate quantities. It is the world's leading producer of iron, coal, petroleum, aluminum, lead, zinc, natural gas, and phosphates. Among the minerals of lesser importance, it

also leads in the output of such things as arsenic, borax, and molybdenum. Leith[1] points out that the output of energy in the United States from coal, oil, natural gas, and water power amounts to nearly half the world's total and that, since the total power available from men and beasts is insignificant in comparison, it is not far wrong to say that the United States is today actually doing nearly half the world's work.

Rich as the United States is in minerals, it is almost entirely dependent on foreign sources of nickel, tin, platinum, antimony, chromite, and manganese. To a lesser degree it is dependent on foreign sources of mercury, tungsten, magnesite, pyrites, and talc. The situation of several leading countries with respect to various minerals is shown in Fig. 11·2.

In the case of some few minerals, the discovery of new deposits or the development of new methods for the recovery of the mineral is, in effect, increasing the quantities available. For example, despite the fact that 1.7 billion barrels of crude oil were produced in the United States in 1944, the proved oil reserves at the end of the year were 20.5 billion barrels as compared with 20.1 billion at the beginning of the year.[2] Similarly helium gas, which had been available in but small quantities in 1917 at a cost of $2500 per cubic foot, was being produced at about ½ cent per cubic foot by 1933 as the result of new processes developed by government scientists in the interim.[3] Aluminum, still at the time a laboratory curiosity, sold at $100 per pound in 1856. The year afterward the price had dropped to $27 a pound; two years later it was down to $17. By 1888, the price was $4 a pound, and late in the 1930's, just prior to the second World War, it had fallen to 20 cents a pound.[4] Yet these cases of increased output and lower prices as the result of new discoveries and improved technology are movements counter to the general trend. Minerals are exhaustible resources, and the richest and most conveniently located deposits are worked first. As time goes on, as stocks dwindle and producers have to shift to poorer ores or sources farther from market, and as population increases, the general tendency is for minerals to grow more expensive. New discoveries and new techniques of production are the only offsets to this tendency. The discovery of new sources is largely a matter of chance, and in the cases of most metals the discovery rate has been slowing down materially within the past few decades. Moreover the mining industries seem to be for the most part subject to what we shall later call the tendency for "increasing costs." All these factors emphasize the

[1] LEITH, op. cit., pp. 48-49.

[2] Facts on File, vol. V, no. 227, p. 63 G.

[3] National Resources Board Report, Dec. 1, 1934, p. 416.

[4] All figures on aluminum prices save the last are from Lancelot Hogben, "Science for the Citizen," p. 458, Alfred A. Knopf, New York, 1938.

importance of the conservation of mineral resources. Yet the mineral industries are, in fact, among the most wasteful of industries. A government agency[1] reported that, in 1934 in a single producing field in the United States, a billion cubic feet of natural gas was being blown into the air daily, more than twice enough to supply the entire United Kingdom, and almost enough to supply every householder in the United States now consuming either natural or manufactured gas. Careful field studies in 1923 showed that there was avoidable loss of bituminous coal in the United States to the amount of some 150 million tons a year, enough to supply the entire requirements of the German Reich.[2] "For every ton of coal produced, our methods of mining have placed a second ton beyond recovery; for every 1000 feet of natural gas turned out, a similar quantity has escaped (into the air); for every barrel of petroleum that has seen useful service, nine barrels have been wasted."[3]

To summarize: Although no nation is self-sufficient for all minerals, the United States has the largest, richest, and most varied supplies of any nation. The rate of use has been increasing rapidly despite the fact that reserves are limited. Despite the fact that careful planning is desirable in the use of all exhaustible resources, there are tremendous wastes in the use of the nation's mineral resources. The hopeful part of the picture is that new sources of supply are constantly being found (although for most minerals at a decreasing rate), that advances in production methods are increasing the efficiency both of mining and of use of minerals, and that at least for the metals one can always rely on past production by using scrap metal to supplement new output. Just prior to the second World War approximately one-third of American steel was being made from scrap, and this source was becoming increasingly important for other metals.

11·6. **Water Resources.** Strangest of all economic nomenclature to the layman is that which includes water as part of the land. Yet the water resources of an area are among its most important features from the standpoint of human occupancy and use. With the single exception of the air, no other natural resource is needed so urgently or used in such large quantities. Water is also used for a large variety of purposes, among the most important of which are: for (1) domestic use for drinking, washing, cooking, etc.; (2) irrigation of crops, both artificial and natural; (3) the development of mechanical power; (4) industrial uses as a raw material and solvent, etc.; (5) transportation; and (6) scenic and recreational uses. Water may also

[1] National Resources Board Report, Dec. 1, 1934, p. 392.

[2] *Ibid.*, p. 401.

[3] GILBERT, C. G., and J. E POGUE, "America's Power Resources," quoted by Stuart Chase, *The Tragedy of Waste*, p. 237, The Macmillan Company, New York, 1926.

appear as illth, rather than as wealth, giving problems of drainage, flood control, and the like.

Water resources are by no means evenly distributed, and this fact has done much to influence the location of population centers and the development of industry. Great areas of the earth are too dry for any present human use. Perhaps 12 to 15 per cent of the earth's land area is dry desert. Regions are properly classed as "arid" if the rate of evaporation exceeds the rate of rainfall. Since evaporation rates vary greatly with latitude, altitude, wind velocities, etc., there is no one amount of rainfall that can be thought of as everywhere necessary for the support of human activities.

Water for domestic use is furnished largely by wells in the rural areas, but mainly by surface water for those people who live in towns and cities. About half of the population of the United States depends on ground water for home use, but three quarters of the public water systems use surface water. The supplies of both ground water and surface water (especially the latter) have been greatly influenced both by drainage and by changes in the forest cover. The waste of mineral resources can well be matched by human depletion of water resources through lumbering, draining, and contamination of supplies. The latter takes place both by the disposal of sewage and by pollution of streams with waste products of manufacturing and mining.

Water is also an important source of power. Until the middle of the nineteenth century it was the preeminent source of power. Then, with the development of the steam engine, coal and other fuels forged ahead. But water has one great advantage over the fuels in this respect in that water power is not an exhaustible or wasting resource. If properly handled, it can be a permanent source of power. In the middle years of the twentieth century water power has again begun to grow in importance. Its early uses were of a nature that confined the use of the power to the spot where it was generated. Coal could be transported; water power could not. But with the development of hydroelectric power it has become possible to transmit the energy of waterfalls to considerable distances. Although circumstances alter cases and no hard and fast rule can be laid down, nevertheless one can say roughly that at present the average distance to which electric power can be transmitted with profit is about 250 miles.

The amount of water power that can be developed at a given site depends both on the quantity of water and on the height of the fall that can be attained. Since the latter is always important, land relief or topography plays a major part in determining the availability of water power in any area. In the days of early New England every little stream that flowed through that hilly country was a continuing source of power for small mills and home manufacturing plants. An area equally well watered, but more nearly level, would have been without water power. For commercial

use, one has to consider not only the amount of flow and the height of fall, but also the regularity of flow. A stream that yields a tremendous volume of flow for 2 or 3 months, and is dry for the rest of the year, cannot well be used as a source of power. The regularity of flow is increased by forest cover and by natural or artificial lakes, so that again we find human activities influencing the water resources of an area.

Water power and steam power differ in a number of important economic aspects. The original investment in a hydroelectric plant is far greater than that in a coal-steam plant. The expenditure for the dam, for transmission lines, for the purchase of land be flooded, for the relocation of highways and railroads—all these make the first cost of water power very great, perhaps three to four times that of an equivalent steam plant. But once built, the hydroelectric plant has great advantages. The depreciation is very low as

Table 11·2. Energy from Mineral Fuels and Water Power, United States, 1940*

Source	Comparative energy
Bituminous coal	12,072
Anthracite coal	1,400
Domestic oil	8,119
Imported oil	256
Natural gas	2,806
Water power	2,620
Total	27,273

* The figures in the table are the equivalent heating power of the various fuels in trillions of British thermal units. Water-power equivalents have been computed from kilowatt-hours of power The data are from "Statistical Abstract of the United States, 1942," p. 453.

compared with that of the steam plant, there is no fuel to buy, only a quarter to a fifth as many men are needed for operation, and the mechanical efficiency of the water plant is typically three or four times as great as that of the steam plant. If we can get ahead of our story enough to foresee some of the problems of management that we shall take up soon,[1] we might say that the fixed costs are much higher in the hydroelectric than in the steam plant but that the variable costs are considerably less. To be sure, it is not uncommon for the two types of power to be combined in use as supplementary rather than as competing sources of energy.

Table 11·2 indicates the relative importance of various sources of power in the United States in 1940. One notes at once that these sources of power all come from the "land" as the economist defines that term. They are all natural sources of power and, in fact, they are all indirect transformations of sun power. To them one would have to add wind power, tidal power, the direct power of the sun's heat, and many other relatively undeveloped but

[1] See Chap. 14.

potentially great sources of natural power, as well as human and other animal power. However, although the figures of Table 11·2 are thus not complete, the other sources of power are minor as compared with those included there.

The water power of the United States is still relatively undeveloped. The lower cost of developing power from fuels has made it natural for industrialists to turn to them first despite the fact that they are exhaustible resources while water power is not. In January, 1941, the actual developed water power capacity of the United States was 18.9 million horsepower, while the potential water power available 90 per cent of the time was 42.8 million horsepower. If one were to include water power available as little as 50 per cent of the time because of seasonal variation in stream flow, the figure would rise to 57.2 million horsepower.[1]

While speaking of the uses of our water resources, one must not overlook their uses in transportation. Water transportation is very cheap as compared with most other types, but under most conditions it is slow. In the United States in 1940, the total water-borne commerce was 608 million short tons, of which 497 million were domestic and 111 million foreign. These figures can be compared with the 1069 million tons of freight originating on the steam railways of the United States and 7000 tons of freight and express carried on American air lines in the same year.[2]

The problems of development and protection of water resources are many and difficult, raising not only economic but also puzzling legal and political questions. The power of the Federal government over interstate commerce and navigation has long been recognized. Within the past two decades the government has been taking more and more active interest in problems of flood control, pollution of surface and ground water, generation of power, etc. It is already clear that the development of our water resources is not to be left to the forces of competition and private enterprise entirely, although it is by no means clear as yet just what the relative positions of the Federal government, the individual states, and private enterprise are to be.

11·7. Forest Resources. The forest resources of a nation are of importance both as sources of wood and because of their influence on stream flow. Forested areas are of great use in slowing the runoff of water and making stream flow more uniform; but in this section we shall consider the forest resources as raw materials rather than as conservers of water.

Lumber may be obtained either from an extractive or from a genetic industry. When forested regions are first opened up, the settlers naturally and properly operate on the extractive basis, clearing the woods and even burning them off in order to get the use of the land. As needs for lumber

[1] Data from "Statistical Abstract of the United States, 1942," p. 464.

[2] *Ibid.*, pp. 487, 510, 511.

grow with increasing population and as supplies grow smaller, there comes a time when it is necessary to reverse this policy and to begin planting and growing trees to replace those which are cut. It is estimated that the territory that is now the United States was once about 40 per cent forested. Reference to Table 11 • 1 will show that this figure has now dropped to about 32 per cent even if we count farm wood lots. Virgin timber stands in the eastern wooded region are now but about 9 per cent of the original area, while even in the western timber region of the United States the virgin timber areas have shrunk to about half their original size. These figures are a good indication of the extractive nature of the industry in this country. At the beginning of the second World War, the annual growth of lumber in commercial forests in the United States was about 11 billion cubic feet while the total volume removed was about 13 billion; in the saw timber sizes, the annual growth was about 32 billion board feet while the removals were roughly 48 million. Thus the extractive nature of the industry was continuing, even though at a diminished rate. Despite increased population and demand, the actual lumber consumption has decreased by about a quarter or a third since its peak in 1907.

Until about 1850 the lumber center of the country was in New England, with its white pines. Thereafter it shifted south and west, first into New York and Pennsylvania, then along the Great Lakes, and at about the turn of the century into the south with its yellow pine. By the time of the outbreak of the first World War the center of production had shifted to the Pacific northwest, with its pines, firs, and spruces, and there the domestic center has stayed ever since, although the south is still of considerable importance. The United States is by far the largest user of forest products in the world, consuming roughly 40 per cent of the total wood and 50 per cent of the total wood pulp of the world. Increasingly it is becoming necessary to rely on imported supplies, and wood and its products now typically travel hundreds and even thousands of miles from forest to market. Most of the domestic supply is privately owned, although the Federal and state governments have become increasingly large holders of forest land until they now own, between them, approximately a third of the remaining timber. Private ownership and exploitation have often resulted in great waste of timber, although in fairness one must say that much of the waste was unavoidable. Even so, the U. S. Forest Service has estimated that careless practices have themselves been responsible for the wasting of 13 per cent of the entire amount of timber cut.

Forest products include not only timber, but also drugs, tanning materials, naval stores, cellulose, and many other commodities of commerce. But no matter what the product, forestry has the peculiarity that it is a very long-time enterprise. Unlike ordinary farm crops, a forest usually yields no

return for decades. The owner must make his original investment in land and trees, he must run the risk of loss from fire or hurricane, and he must pay the taxes on his land and crop year after year with little or no income. If the owner is himself mature when the forest is first planted, it is unlikely that he will live for the harvest. This time factor raises many problems which have bothered those interested in developing a rational forest policy. They have felt that some revision is needed in the tax system under which the owner may pay the entire tax when the trees are cut and the income received, rather than being forced to pay piecemeal as the forest grows. The problem involved is by no means peculiar to forestry. The same problem arises in connection, say, with a war bond. The government allows the purchaser, in computing his income for tax purposes, to wait until the bond matures and then pay the tax on the entire income on the bond or, if he prefers, to compute each year the amount by which the bond has appreciated in value and to pay his tax on this annual appreciation. Year by year, as the trees grow, the actual present value of the forester's holdings increases. It is not necessary for him to cut the timber to get an income; he can sell out to get an income. Yet the time involved in forestry does make it particularly difficult for the industry to be carried on as a genetic industry under private ownership unless the land is owned by a large corporation, which, of course, is perpetual. Either the government or a private corporation can afford to take a longer view than can the typical private owner.

The problem of establishing a sensible forest policy for the nation is complicated, involving consideration of such factors as labor costs, interest rates, tax policies, freight rates, and the use of substitute materials. Certainly we are still far from a satisfactory permanent policy in the United States.

Questions

1. Name some parts of the present natural environment that are not natural resources. Under what conditions would they become natural resources?

2. What proportion of the value of a typical American farm do you suppose resides in the "original and indestructible powers of the soil," and what part in man-made improvements?

3. Give actual examples of location as a factor influencing the value of land. Choose situations in your own neighborhood.

4. What are the principal iron-ore producing regions of the United States? The principal coal regions? The principal petroleum regions?

5. Figure 11 • 2 shows a number of minerals for which the United States is dependent on foreign countries. What are our principal sources of supply of each? For what is each of them used primarily?

6. In the first paragraph of Sec. 11 • 6 are listed five major uses of water. Do you think that they are listed in order of importance? If not, what would be their correct order?

7. What are the main advantages and disadvantages of water as a source of power?

8. Barring new discoveries and inventions, would you expect water power to become more or less important in the next thousand years as compared with other sources of power? Why?

9. In 1920, some geologists and others were predicting that the supply of petroleum in the United States would be exhausted within 25 years. By 1945, the total known reserves were larger than the total known reserves in 1920. Does this indicate to you that there is no problem of conservation in the use of petroleum? Discuss.

10. In what ways has the peculiarity of American law on land ownership, combined with the liquid condition of petroleum, affected the mining of petroleum?

12. HUMAN RESOURCES

12·1. Labor. The word "labor" is used in common speech to mean many things: unskilled work, hard work, organized unionism, etc. To the economist the term refers to human effort, but not to all kinds. Some human effort satisfies wants directly; it is pleasant. Thus a man playing tennis, skiing, or mountain climbing exerts a large amount of energy, but he is not productive. Just as we include among the natural resources only those parts of the physical environment which are put to productive use, so we define *labor as human effort used in production.* It is immaterial whether this productive human effort is that of a day laborer or of his employer, that of a dentist in his office or a judge on his bench. All sorts of human effort are classed as labor if they are used productively.

The exertion of energy may be, as we have said, pleasant. But labor usually is not so, because in order to be productive it is necessary that it be directed into particular channels and concentrated and continued there. There is an element of compulsion in connection with it. If one does not work, he does not get the product, be it vegetables from his garden or money in the pay envelope. Unless one anticipates that the product will be forthcoming, he does not do the work. He must see more satisfaction by working than by abstaining, or he will not work.

12·2. Importance of Human Resources. In Chap. 11 we gave some attention to the problem of natural resources. Different parts of the world differ greatly in their endowment in this respect, and the differences are reflected in the wealth and income of the people of the areas involved. But the human resources of an area are also important in determining the amount that is produced, and thus the amount that can be consumed. For example, when the white men came to what is now the United States it is likely that there were not more than a million Indians in the area.[1] They

[1] REUTER, E. R., "Population Problems," p. 49, J. B. Lippincott Company, Philadelphia, 1937.

had the same natural environment that we have today save for the fact that some of the resources have been exhausted in the meantime. Yet these Indians lived in what we would consider abject poverty. The difference between their life and ours is largely a difference in the human resources—the energy and ability of the people. The inhabitants of the tropics, living amongst almost unparalleled natural riches, often have barely enough to keep body and soul together. The inhabitants of the Scandinavian peninsula, by contrast, live in a forbidding climate and cultivate relatively unproductive soil, yet they manage to make for themselves a far better living than do the tropic savages. The industry, ingenuity, perseverance, sobriety, cooperativeness, honesty, strength, and intelligence of the people—the human resources—must be considered along with the character of the land or natural resources in explaining the wealth or indigence of a people.

In considering the human resources of any society, we find it helpful to take up separately the qualitative and the quantitative aspects of the problem. The quantity of the labor force and its quality may to some extent be but different aspects of the same problem, but society attacks them as two different problems, and we shall so consider them.

12 · 3. **Quality of the Population.** The quality of any population is the result in part of heredity and in part of environment and training. It is not our intent here to take part in the long-standing quarrel as to the relative importance of heredity and enviroment, of nature and nurture. Suffice it to say that both factors are important in molding the human resources of a country, and we shall have to look briefly at each.

First, as to heredity, we may state a few general conclusions. An elementary textbook in economics is no place for an extended discussion of human heredity. The student who is interested in such matters will do well to turn to some of the popular treatments of the subject or, better yet, to take a course in genetics and acquaint himself with the facts at first hand.

Sociologists have long illustrated the problem with the history of two or three extreme cases—families whose pedigrees and accomplishments have been painstakingly studied and recorded. Thus the Jukes, the Kallikak, and the Nam families show generation after generation of degenerate, shiftless, criminal people, while the Edwards family by contrast shows in each succeeding generation large numbers of scientists, lawyers, teachers, physicians, and public citizens of unusual merit and productivity. The student will do well to read the history of one or more of these cases. But he must be warned at the start that the differences between them are by no means certainly hereditary. The families have differed in heredity, but also in environment. More recent studies of identical twins have attempted to learn what happens to persons of identical heredity under different environments.

If we can summarize the results of these studies briefly in the words of one of the country's leading population experts of today,[1]

> The upshot of these studies on the influence of heredity and environment on twins and foster children is to show beyond question that heredity does place definite limits to the development of the individual in certain respects but that in many important respects these limits are not narrow, and hence environment can produce significant differences between individuals even when they have identical heredity. Moreover, the more significant the characteristics are from the standpoint of the adjustment of the individual to the community, the broader seems to be the field within which environment can act on the traits provided by heredity. . . . What has been said above about the relation of heredity and environment does not mean that we should not concern ourselves about the improvement of society by the improvement of human heredity, but it does mean, as will be pointed out later, that we shall probably make far greater progress in improving the quality of our population by providing an environment favorable to the development of the heredity we already have than by concentrating any considerable amount of effort on the improvement of heredity itself.

Yet in the most extreme cases, such as the inheritance of certain types of insanity, feeble-mindedness, or epilepsy, the problem of inheritance is of such an urgent nature that society cannot afford to overlook it. The inheritance that plays the greatest part in affecting human life is probably not that of physical and mental characteristics, but the inheritance of environment, if we may use that term—the chance as to whether one is born into a family of high income in a good home with intelligent parents or whether, with the same physical inheritance himself, he is born into a poor family with untrained parents in a poorly equipped home.

When we turn to the environmental side, we have to include all those things which surround the individual and influence his development as he grows up. The environment, then, includes the training and the social surroundings as well as the physical surroundings. To be sure, the environment cannot furnish traits or abilities that were not inherited, but it can do a great deal to shape those things which were inherited. Perhaps genius, like murder, "will out," but it is much more likely to realize its full potentialities if it is given the proper medium for development. Any college student today, given the right environment, can learn the calculus, which, in another environment, had to await the coming of one of the greatest minds of all time.

A very large part of the effort of society goes toward influencing the quality of our human resources. Obviously this would include our entire educational program, on which we spend several billion dollars a year; but

[1] THOMPSON, WARREN S., "Population Problems," pp. 349-350, McGraw-Hill Book Company, Inc., New York, 1942.

it would also include the efforts of the church and the home and the influence of one's social and professional associates. Heredity determines the character of the raw material with which we have to work; environment

Table 12·1. Population of the United States, 1610-1945*

Year	Population, thousands	Year	Population, thousands	Year	Population, thousands	Year	Population, thousands
1610	†	1860	31,513	1890	63,056	1920	106,466
1620	2	1861	32,351	1891	64,361	1921	108,541
1630	6	1862	33,188	1892	65,666	1922	110,055
1640	28	1863	34,026	1893	66,970	1923	111,950
1650	52	1864	34,863	1894	68,275	1924	114,113
1660	85	1865	35,701	1895	69,580	1925	115,832
1670	114	1866	36,538	1896	70,885	1926	117,399
1680	156	1867	37,376	1897	72,189	1927	119,038
1690	214	1868	38,213	1898	73,494	1928	120,501
1700	275	1869	39,051	1899	74,799	1929	121,770
1710	358	1870	39,905	1900	76,094	1930	123,077
1720	474	1871	40,938	1901	77,585	1931	124,040
1730	655	1872	41,972	1902	79,160	1932	124,840
1740	889	1873	43,006	1903	80,632	1933	125,579
1750	1,207	1874	44,040	1904	82,165	1934	126,374
1760	1,610	1875	45,073	1905	83,820	1935	127,250
1770	2,205	1876	46,107	1906	85,437	1936	128,053
1780	2,781	1877	47,141	1907	87,000	1937	128,825
1790	3,929	1878	48,174	1908	88,709	1938	129,825
1800	5,308	1879	49,208	1909	90,492	1939	130,880
1810	7,240	1880	50,262	1910	92,407	1940	131,970
1820	9,638	1881	51,542	1911	93,868	1941	133,203
1830	12,866	1882	52,821	1912	95,331	1942	134,665
1840	17,069	1883	54,100	1913	97,227	1943	136,497
1850	23,261	1884	55,379	1914	99,118	1944	138,083
1855	27,386	1885	56,658	1915	100,549	1945	139,621,000
1856	28,212	1886	57,938	1916	101,966	1946	
1857	29,037	1887	59,217	1917	103,266	1947	
1858	29,862	1888	60,496	1918	103,203	1948	
1859	30,687	1889	61,775	1919	104,512	1949	

* Data for 1610–1780 are estimates from Table I, p. 9, "A Century of Population Growth, 1790–1900," U. S. Bureau of Census; for 1790–1850 are data of the centennial census; and from 1855 on are census estimates as of July 1 each year. Censuses have been taken at various dates, but never on July 1. Hence these July 1 estimates do not coincide exactly with the census counts even in census years.
† Less than 500 population (210 in Virginia).

and training are the processes by which we shape that raw material into the finished product. It is just as vain to attempt to separate heredity and environment as it is to separate the quality of the lumber and the quality of the workmanship that go into a piece of furniture.

Yet one more word on this subject. Different groups or races of people have succeeded in very different degrees in meeting their economic problems. For example, a few paragraphs back we pointed out the difference in the achievement of the Indians and of the white men who followed them in the United States. It is only fair to say that, although there are great differences between the individuals who go to make up any social group, as far as can be learned there are but minor differences between the abilities and potentialities of social groups or races themselves, given the same environment and training. Indians can be trained today to make the same sort of living from our environment that the white men make. The differences that we see are greater within groups than between groups, and the differences

Table 12·2. Percentage Increase of Population between Censuses, United States

Census	Percentage increase over preceding census	Census	Percentage increase over preceding census
1790	....	1870	26.6
1800	35.1	1880	26 0
1810	36.4	1890	25.5
1820	33.1	1900	20.7
1830	33.5	1910	21.0
1840	32.7	1920	14.9
1850	35.9	1930	16.1
1860	35 6	1940	7.2

that exist between groups are largely the result of differences in environment, training, and opportunity.

12·4. Quantity of Labor. Relatively little is known about the history of the growth of the world's population; even to-day the figures are known but roughly. The League of Nations estimated the world population at the start of the second World War as about 2.2 billion persons, and gave, as the corresponding figure for the year 1900, the population of 1.5 billion. Since many countries take no census and since no country takes a census that is entirely accurate, these figures are open to a rather large percentage of possible error; yet they will have to serve as a rough index in the absence of anything more reliable. For the United States, and especially in recent years, the facts are known somewhat more accurately. The figures (see Table 12·1) for the years before 1790 are rough estimates; those after that are based on census counts and are presumably somewhat more accurate.

If the student will bother to carry out the necessary calculations, he will find that the rates of decennial increase have been getting smaller and smaller in recent times (see Table 12·2). Although the population is still growing,

the rate is becoming smaller and smaller, and one can look ahead to the time when the population will cease to grow and become static if present trends continue. Let us now look beneath the surface to see if we can find something of the causes of this trend.

Population growth is made up of two parts: (1) migration into or out of the country and (2) change in population from births and deaths among those already there. We shall take up each of these in turn for study, starting with the latter.

12·5. Birth Rate. If for the moment we neglect the effect of immigration or, for purposes of our present argument, assume that we are studying a country in which there is no migration either inward or outward, we realize that changes in the size of the population must come from births and deaths. In order to compare situations at different times or in different places, we usually put birth and death figures in relative terms, as *birth rates and death rates per thousand.* Thus, if we say that there were 1404 deaths in Nevada and 149,946 in New York in 1940, it looks at first glance as though deaths were much more common in New York. But when we recall that the population of Nevada was 110,247 and that of New York was 13,479,142 in 1940, we realize that there were many more people in New York to die. In fact, the death rate in Nevada was, we see, 12.7 per thousand, while that in New York was but 11.1 per thousand. The rate of deaths was lower in New York than in Nevada.

For many years there has been a tendency for a slowly falling birth rate, both in the United States and in many other countries. In the United States the records of births have been incomplete until very recently. Not until 1933 did the birth registration area include the entire country. As a result, it is impossible to do more than estimate the birth rate prior to that time. The estimates of rates for the period since 1915 appear in Table 12·3. Other data seem to indicate that the birth rate was 50 per thousand or even higher in 1800, that by 1850 it had fallen to between 40 and 45 per thousand, and that by 1900 it was about 30 per thousand. Although the data of the preceding sentence are very rough and approximate, they may well be compared with the data of Table 12·3 in understanding the general trend. The data of the table show an upturn toward the end, with the onset of the war. It is too early to tell, as this is written, whether this is a temporary war phenomenon or a change in the trend. The data are shown graphically in Fig. 12·1.

Although we do not have reliable data on births before 1915, we can judge something of the facts from census data showing the numbers of children and the numbers of prospective mothers. If, for each census, we compute the number of children under five years of age for every 1000 women of childbearing age, we note the very marked tendency for a reduction in

childbearing. The figures appear in Table 12·4. These data, and many others, lead to the inevitable conclusion that the reductions in the birth rate, which we see in the first part of the period during which we have accurate data,

Table 12·3. Birth Rates in the Birth Registration Area of the United States, 1915-1942

Year	Birth rate per thousand	Year	Birth rate per thousand
1915	25.0	1931	17.8
1920	23.7	1932	17.3
1921	24.2	1933	16.6
1922	22.3	1934	17.2
1923	22.2	1935	17.0
1924	22.4	1936	16.8
1925	21.5	1937	17.1
1926	20.7	1938	17.6
1927	20.6	1939	17.3
1928	19.8	1940	17.9
1929	18.9	1941	18.9
1930	18.6	1942	21.0

are but the continuation of a reduction that has been long-continued. It has almost certainly been falling for more than a century in the United States, and is presumably now not more than half what it was a century and a half ago.

Table 12·4. Number of Children under the Age of Five, per Thousand Women Aged 16-44, United States, 1810-1940

Year	Number of children	Year	Number of children
1810	976	1880	635
1820	928	1890	554
1830	877	1900	541
1840	835	1910	508
1850	699	1920	489
1860	714	1930	407
1870	649	1940	342

The general long-run tendency for a falling birth rate is the result of many interrelated causes. To some extent it is the result of changes in the economic function of the family in American life. A century or so ago, children were in most cases economic assets to the family. At a tender age they

began to carry their share of the farm work at a time when hired help was difficult to obtain and expensive. The farmer with 10 children was economically better off than the farmer with no children. Today the situation is changed. Children are ordinarily economic liabilities, regardless of their desirability in other ways. Although conditions vary greatly with different families, it has been estimated that a child of fifteen represents a family investment of some $1500 to $2500 in net cash outlay and, although in some cases a small income is returned to the family, generally the cash income returned is negligible or entirely absent. Each child in the family adds to the burden of expense both of money and of energy, and there is usually little or no added income to offset the new outgo. Many people do not have a

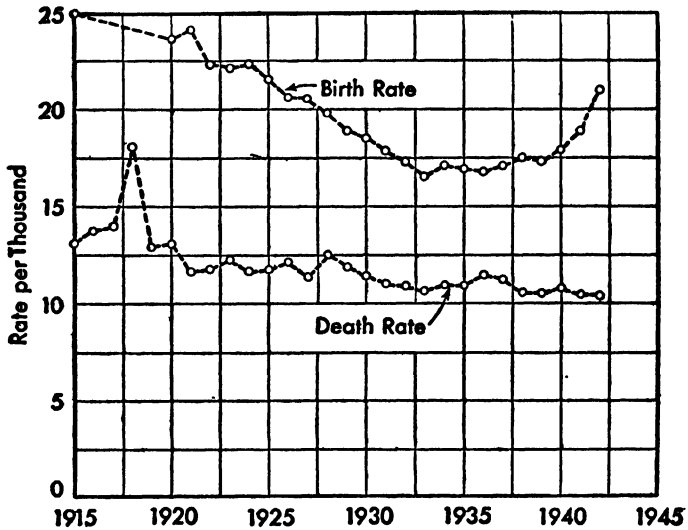


FIG. 12 · 1.—Birth rates and death rates, United States, 1915-1942.

sense of responsibility well enough developed to understand the obligation of each generation to repay to the following generation the debt that it owes to the preceding generation—to repay in some small measure to one's children the debt that one owes to one's parents. The matter is further complicated by the difficulty of securing adequate housing in the city life that is led by such a large proportion of our modern population. The lengthening of the period of education and the consequent retardation of the age of marriage also play a part in the reduction of the birth rate. The new status of women and their increasing independence and opportunity to work have reduced somewhat the likelihood of marriage and childbearing. The spread of information concerning methods of contraception has certainly played some part also.

Up to this point we have talked of the birth rate for the country as a whole, without breaking it down into its constituent parts. Actually, however, the birth rates in different groups of the population are very different. Those among people in rural communities are higher than those among city dwellers. Birth rates among recent immigrant stocks are higher on the average than those among the old Yankee stocks; rates among people of low income are higher, on the average, than those of the middle class or wealthy groups. Many sociologists have pointed out what seem to them to be indications that birth rates are higher among those of less intelligence than among those of higher intelligence, that the stocks that would pass on the best heredity are those which multiply most slowly. The reader will realize at once that there is a great deal of room for disagreement as to which stocks are really the most desirable, and which qualities of human beings we should try to perpetuate if we had the opportunity. Some students have been disturbed over differences in birth rate in different parts of the world, pointing to the falling rates among the western nations and seeing the threat of a "yellow peril."

These problems arising from differences in the birth rate among different groups within the population are called the problems of the "differential birth rate." There is no question but that the differences in the birth rates exist, but there are great differences of opinion as to how much they matter; even among those who think that they matter there are grave differences of opinion as to what practically can be done about it. Most countries practice sterilization of the unfit in at least the most extreme cases. Segregation of the sexes is common at least among the mentally unbalanced. Some people have advocated a conscious effort on the part of society to improve the human stock as breeds of livestock have been improved through generations of planned mating, although presumably the methods employed would have to be somewhat different. The *eugenics* movement is directed toward the improvement of human heredity. It, too, would establish a differential birth rate, but one in which the heavy rates were among the better stocks and the light ones among the weaker stocks. Some students of population problems feel that the differential birth rate is slowly disappearing by itself and that the problem will ultimately take care of itself.[1]

12·6. Death Rates. The death registration area of the United States was not organized until 1900 and did not include the entire country until 1933. Consequently we are again handicapped by paucity of reliable information with regard to trends. Death rates in the United States since 1900 are given in Table 12·5, and those since 1915 are depicted graphically in Fig. 12·1. It will be noted that death rates, like birth rates, show a tendency

[1] See, for example, Thompson, op. cit., pp. 184-186.

for slow diminution. Although no figures are available for the United States as a whole before 1900, it may be of some interest to notice that in Massachusetts, where records were kept much earlier, the death rate was 27.8 per thousand in 1789, had fallen to 21.4 in 1855, and to 20 in 1875. The student will notice, by the way, as he inspects the data of Table 12·5 or Fig. 12·1, the effect of the influenza epidemic of 1918 on the death rate for that year.

The slow fall in the death rate has been the result of mixed causes, like that in the birth rate. Immediately one thinks of the discoveries in the field of medicine and public health—the sulfa drugs, penicillin, or, if we go farther back, the discovery of the relationship of germs to disease and the

Table 12·5. Death Rates per Thousand, United States, 1900-1942*

Year	Rate	Year	Rate	Year	Rate
1900	17.2	1915	13.2	1930	11.3
1901	16.4	1916	13.8	1931	11.1
1902	15.5	1917	14.0	1932	10.9
1903	15.6	1918	18.1	1933	10.7
1904	16.4	1919	12.9	1934	11.1
1905	15.9	1920	13.0	1935	10.9
1906	15.7	1921	11.5	1936	11.6
1907	15.9	1922	11.7	1937	11.3
1908	14.7	1923	12.1	1938	10.6
1909	14.2	1924	11.6	1939	10.6
1910	14.7	1925	11.7	1940	10.8
1911	13.9	1926	12.1	1941	10.5
1912	13.6	1927	11.3	1942	10.4
1913	13.8	1928	12.0		
1914	13.3	1929	11.9		

* Data from "Statistical Abstract of the United States."

development of vaccination. These have surely been potent factors in the reduction of the death rate. But it is no disparagement of the medical profession to point out that a very considerable part of the reduction is presumably due to other forces. The increase in the general level of education has played an important role, as has the raising of the general income level—the economic level at which families live. The reduction in the average size of family has itself played some part, since death rates are always very high among small children and are higher for the later born than for the earlier born children in the same family. Shifts of population from country to city, changes in the racial makeup of the population, and many other factors have played some part in the general trend.

Although many people realize that death rates have been falling, it is

commonly overlooked that the change has been almost entirely at the younger ages. Death rates have always been very high among children in the first few days and weeks of life. Graunt estimated that in London about 1662 over a third of the children born failed to reach the age of six. In France in 1750 we are told that only 54 per cent of the children reached the age of five. In the United States during the first 10 years of the twentieth century 12.7 per cent of all babies born died before their first birthday. In the decade of the twenties, this figure had dropped to 7.2 per cent, and in the thirties to 5.7. Better understanding of the diseases of infancy and childhood has made it possible to make great gains in death rates at these ages. But, unless someone discovers a key to immortality, those who are saved from death during infancy and childhood, having escaped the threat of diarrhea and diphtheria and tuberculosis, remain to succumb to the difficulties of old age—heart disease, hardening of the arteries, cancer, etc. The fact that these latter diseases

Table 12·6. Expectations of Life for White Males and Females in the United States Since 1900, at Various Ages

Year	At birth		At age 20		At age 65	
	Male	Female	Male	Female	Male	Female
1900—1902	48.23	51.08	42.19	43.77	11.51	12.23
1909—1911	50.23	53.62	42.71	44.88	11.25	11.97
1919—1921	56.34	58.53	45.60	46.46	12.21	12.75
1929—1931	59.12	62.67	46.02	48.52	11.77	12.81
1939—1941	62.81	67.29	47.76	51.38	12.07	13.56

are becoming more common as causes of death is because more people live to be old enough to suffer from them. The average length of life is growing; the average age of the people of the United States is greater than it was a few decades ago. And, although great strides have been made in preventing and curing the ailments of childhood, there has been little or no improvement in this regard after middle age.

The *expectation of life* at any age is the average length of time which people of that age still have to live. The improvements in death rate among children has increased this expectation at the lower ages very significantly. But at ages beyond middle life it seems to be no greater today than it was in colonial times; in fact, not significantly greater in the United States with all the medical and sanitary knowledge than it is in backward countries like China or India. We can see the nature of the changes that have taken place in Table 12·6, which shows the expectations of life at various ages for males and for females since the beginning of the present century. It will be noticed

first of all that women live longer, on the average, than do men, either because their occupations are less hazardous or because of inherent differences. But, as one runs his eye down any column in the table, he notices at once what has happened to life expectancy with the passage of time. Toward the left-hand part of the table the expectancies are much greater than they used to be. A baby boy born in 1940 could expect to live (if his life were of average length) some 14½ years longer than his father born in 1901. But people at the age of twenty in 1940 had an advantage of but 5½ years over the twenty-year-old of 1901, and at sixty-five there appears to be little or no change in average length of remaining life.

As a matter of interest, we present in Table 12 • 7 a more detailed picture of the last census given in Table 12 • 6. It may help in interpreting this table if we follow across one line, explaining as we go. Taking for illustration the age of twenty, we see that out of every 1000 white males who reach their twentieth birthday, 2.12 will die before their twenty-first birthday. Of 100,000 white baby boys born 20 years ago, there will be but 92,293 left to celebrate their twentieth birthday, the others having died younger. Of those who reach their twentieth birthday, the average length of life still remaining is 47.76 years. Some will live much longer than that and some will die tomorrow, but the average length of after lifetime—the expectancy of life—is 47.76 years. The following three columns are interpreted in the same way, save that they refer to lives of white females. It is on tables such as this that insurance companies base their premiums.

The expectation of life at birth is thought to have been about 18 years in 1790, and has risen to almost 63 years today, a gain in average length of life for new-born children of 45 years. But the *span of life*—the extreme length of time that men live—seems to be no longer today than it was in 1790. Men lived occasionally to be one hundred years old then and they do so still, but the extreme upper limit of life has not been increased to 150 or 200 years; the span of life does not appear to have increased at all. At this point we note that one cannot expect continued reductions in the death rate. In a stationary population—one that was not growing but was maintaining a stable age distribution—a continued death rate of 20 per thousand would mean that the span of life was 50 years. A death rate of 10 per thousand would mean that the span of life was 100 years. A death rate of 25 per thousand would indicate a span of life of 40 years. For a stationary population we find the span of life by dividing 1000 by the death rate. If we try that for the figures of Table 12 • 5, using the figure for 1942 (which is a death rate of 10.4), we find a span of life of about 96 years. We know that the average length of life in the United States is far below that figure, but we do not, of course, have a stationary population. We have one that is growing in size and

changing rapidly in age composition. Yet we could not reduce our present death rate continuously, in the absence of immigration, unless the span of life

Table 12·7. Life Tables for White Males and White Females in the United States, 1939-1941*

Age	White males			White females		
	Deaths in ensuing year per 1000 alive at given age	Number of survivors to this age out of 100,000 births	Expectancy of life at this age, years	Deaths in ensuing year per 1000 alive at this age	Number of survivors to this age out of 100,000 births	Expectancy of life at this age, years
0	48.12	100,000	62.81	37.89	100,000	67.29
1	4.87	95,188	64.98	4.32	96,211	68.93
2	2.65	94,724	64 30	2.20	95,796	68 23
3	1.90	94,474	63.46	1.61	95,585	67.38
4	1.53	94,295	62.58	1.28	95,431	66.49
5	1.38	94,150	61.68	1.10	95,309	65.57
10	1.00	93,601	57 03	0.70	94,890	60.85
15	1.43	93,089	52.33	0.96	94,534	56.07
20	2.12	92,293	47.76	1 45	93,984	51.38
25	2.43	91,241	43.28	1.82	93,228	46.78
30	2.79	90,092	38.80	2.20	92,320	42.21
35	3.63	88,713	34 36	2 78	91,211	37.70
40	5 13	86,880	30 03	3.68	89,805	33.25
45	7.66	84,285	25.87	5.23	87,920	28.90
50	11.55	80,521	21.96	7.62	85,267	24.72
55	17.37	75,156	18.34	11.28	81,520	20.73
60	25 48	67,787	15.05	17.14	76,200	17.00
65	36 85	58,305	12.07	26.43	68,701	13.56
70	54 54	46,739	9 42	42 33	58,363	10.50
75	83.13	33,404	7.17	68.89	44,685	7.92
80	124.71	19,860	5.38	108.19	28,882	5.88
85	181.04	9,013	4.02	162.94	14,487	4.34
90	248 94	2,812	3.06	231.41	5,061	3.24
95	320.76	552	2.40	307.96	1,109	2.47
100	389.35	65	1.96	387.39	139	1.95

* Data from "Statistical Abstract of the United States, 1943," Table 82, p 83. The table is based on the census population of 1940, and on the deaths of 1939-1941.

were to increase remarkably. Continued immigration of young people and large numbers of births used to make our average age low. The reduction in immigration and the reduction in births combine to bring about a slow in-

crease in the average age of the population and, as we become more fully adjusted to the change, we may well see some increase in the death rate unless some as yet undreamed-of discovery in medicine and public health actually increases the life span.

12·7. Natural Reproduction Rate. Whether a population increases or decreases in size (still barring for the time being the effects of migration) depends on the birth rate and the death rate in that population. If the number of births is large and the number of deaths small, the population increases in size; if the number of deaths is large and the number of births small, the population decreases. It is a simple problem in arithmetic. If, starting with a given number, we add and subtract something, the resulting number will be greater than that with which we started if the addition is greater than the subtraction; smaller, if the subtraction exceeds the addition. A static population would result if the birth and the death rates were exactly equal.

If, now, we look back at Fig. 12·1, we notice that the birth rate has been higher than the death rate in the United States. This indicates that the population has been growing, even without immigration. It has been growing from natural reproduction. If we substract the death rate from the birth rate, we get the *natural reproduction rate,* which shows us the rate at which the society is growing or shrinking from internal causes—barring migration. Thus, if a city of 1000 people has 25 deaths and 19 births in a year, there will be a net loss of 6 persons. The birth rate is 19 per thousand; the death rate is 25 per thousand; and the natural reproduction rate is the difference or −6 per thousand. Negative natural reproduction rates indicate that the population is falling; positive ones that the population is increasing. Inspection of Tables 12·3 and 12·5 will show, for example, that the natural reproduction rate in the United States in 1942 was 10.6 per thousand, indicating a rate of annual increase of about 1.06 per cent in the population from births and deaths alone.

Careful statistical work by experts in the field of population study seems to verify the impression that one gets at first glance from Fig. 12·1, namely, that, although the birth rate exceeds the death rate in the United States and both have been showing a tendency to fall, the birth rate has been falling more rapidly than the death rate, so that the two are slowly approaching each other. The natural reproduction rate, of course, has been diminishing under such circumstances. If the prewar trends continue, we can look ahead to the time in the future when the birth rate and the death rate are equal, when the natural reproduction rate falls to zero, when the population stops growing, barring migration. Of course, no one can be sure that present trends will continue, but if they do it is estimated that the United States will stop growing, barring migration, about the year 1980 or 2000. A glance back at

Table 12·2 will give the student another indication of the tendency up through the 1940 census.

Before leaving these problems of birth rates, death rates, and natural reproduction rates, it would be well to point out to the student that the field of vital statistics is a difficult one and requires much experience. The vital statistician would point out convincing reasons for refusing to use crude and uncorrected birth rates and death rates as we have been doing, and would assert that comparisons of such crude rates may well lead to entirely unwarranted conclusions. The reader should not assume that the simple statistical data given here, taken alone, prove anything. At most, they suggest to the student who approaches the population problem for the first time some of

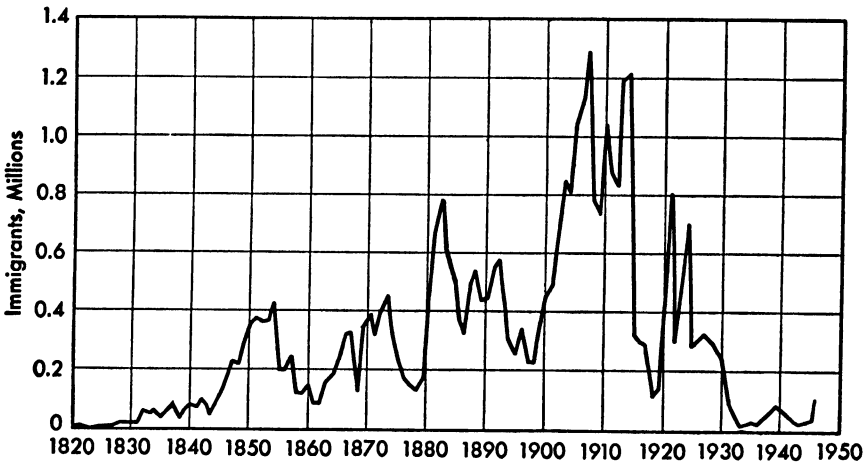


FIG. 12 · 2.—Immigration into the United States, 1820-1946.

the forces that are at work to determine the size of the population. For this reason an effort has been made to reinforce the figures of this chapter with general statements as to the nature of the problem and the trends that seem evident to students who have had the opportunity to go into the matter in far more detail than we can do in an elementary approach to the problem.

12·8. **Human Migration.** When societies were typically organized on an extractive basis, migrations occurred regularly. Pastoral tribes kept their herds in a given territory until the pasturage was exhausted and then moved on. Migration was on a tribal rather than an individual basis. As individual freedom became more common and men were no longer tied down to the land, migration became less an organized group affair and more an individual or family decision. This family migration, like the migration which had preceded it, was largely the result of economic causes. People moved in the hope of making a better living. To be sure, other causes also played a

part: efforts to escape religious or political persecution, and the wish to escape personal difficulties and maladjustments. But population movements have always been primarily economic and have continued so down to the present day. Within the past few decades, to be sure, we have seen some countries organizing migration for military and national considerations, but even in these cases the causes have been basically economic ones.

Since migration has been largely an effort of men to better their economic lot, it has usually taken place from densely to sparsely populated areas. In the heavily populated regions, men are abundant and wages tend to be low; in the sparsely populated regions, more men are needed and higher wages are offered to attract them.

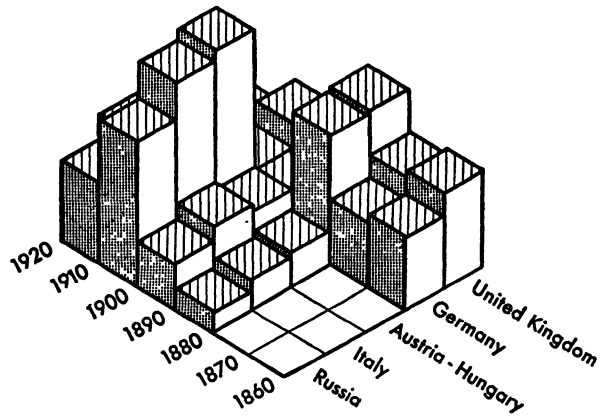


FIG. 12 • 3.—Immigration to the United States from selected countries, 1860-1920.

12 • 9. Immigration to the United States. The greatest of the modern population movements was the great migration to the United States, which started as a tiny trickle with the first settlements, grew into a sizeable stream with the disastrous failure of the Irish potato crop in 1846 and the revolutions of Germany and central Europe in 1848, and reached flood stage from the turn of the century until the beginning of the first World War. The succeeding waves of immigration can be seen in Fig. 12 • 2. The student will note that, although there were temporary interruptions in times of war and depression, nevertheless there was a marked general movement toward heavier and heavier immigration throughout the period up until the time of the outbreak of the war in 1914. During most of this period such immigration was generally looked upon as a desirable thing, and the United States government did everything that it could to encourage and increase it.

Most of the early migration to the United States came from England. Beginning in the 1840's there was very heavy immigration from Ireland and

Germany. More than 2 million Germans came to this country between 1871 and 1890, and the German migration averaged almost 100,000 a year from 1850 to 1890. The peak of Irish immigration had come a little earlier, in the 1840's and 1850's. During the period just before and just after the Civil War, there was heavy migration from the Scandinavian countries. But about 1880 or a little later, the sources of our immigration slowly began to shift. The religious and political persecution in northern Europe was about over. The best of the free land in the United States was taken up, and we began to need unskilled factory workers rather than independent artisans and farmers. As a result of these changes, the number of immigrants from northern and western Europe started to diminish, while the number from

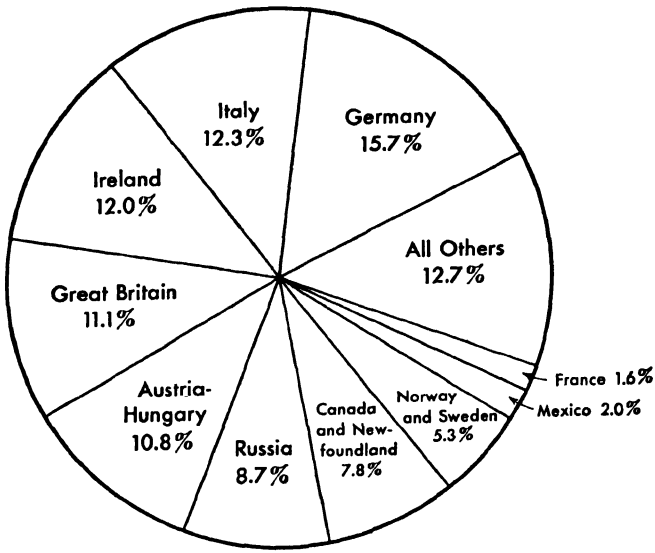


FIG. 12 • 4.—Total immigration to the United States from various countries, 1820-1940.

the Mediterranean countries and from southeastern Europe generally increased suddenly and tremendously. The general character of the shift can be seen in Fig. 12 • 3, which shows the immigration from selected countries from 1860 to 1920. We note that the northwestern flow reached its crest in the 1880's, when the southeastern flow was just beginning to swell. As the migration from England, Ireland, Germany, and the Scandinavian countries diminished, the influx from Italy, Austria-Hungary, Russia, and the Balkans more than made up for the loss. The immigration from northwestern Europe is commonly spoken of as the "old immigration," as contrasted with the "new immigration" from southeastern Europe and, although the change from the one to the other was gradual, the date of the change

is usually given as 1882. This date was an important one in the immigration experience of the United States for several reasons: (1) Immigration in that year reached a peak that was not surpassed for over 20 years (see Fig. 12·2). (2) The first important laws regulating immigration were passed in 1882. (3) It was at about this time that the shift in sources of population was first becoming noticeable. Therefore the custom has grown up of taking the year 1882 as the dividing point between the new and the old immigration.

Immigration into the United States reached its highest level in the decade before the first World War. Reference to Fig. 12·2 will show that several times in that period the immigration reached a figure of over a million immigrants a year at a time when the entire population was only about 90 million (see Table 12·1). In 1907, the immigration for the year was roughly 15 per cent of the total population! Although this flood of immigration was cut sharply by the war, as soon as the conflict was over it began to swell again and might well have reached its olden proportions had not steps been taken to stop it—had not the nation reversed its former policy of encouraging immigration and begun to discourage and restrict it.

During the period since records have been kept of the number of immigrants, Germany has sent us the largest number, closely followed by Italy, Ireland, Great Britain, and Austria-Hungary (see Fig. 12·4). But the countries that have sent the most immigrants are not necessarily those which are most heavily represented in our population. A country that sent its immigrants early, and whose immigrants and their descendants had high reproduction rates, will have more of its stock here now than a country that sent the same total number of immigrants, sent them late, and sent people with low reproduction rates. Figure 12·5 shows the national origins of the white population of the United States at the time of the 1920 census and, because there has been relatively little immigration since that time, it is likely that the diagram will serve at least very roughly to indicate the origins of our present white population. The contrast between Figs. 12·4 and 12·5 is interesting and instructive. Just over 40 per cent of the American white stock comes from Great Britain and Northern Ireland; if we add Germany, the Irish Free State, Norway and Sweden, we account for about 70 per cent. Canada brings the total up to over 75 per cent.

During the years 1800-1930 the white population of the United States increased some 106 million persons. The population experts, Thomson and Whelpton[1] estimate that about 72 per cent of this increase came from natural reproduction, and about 28 per cent from immigration. Immigration was

[1] Adapted from data in W. S. Thompson and P. K. Whelpton, "Population Trends in the United States," p. 303, McGraw-Hill Book Company, Inc., New York, 1933.

relatively most important in the decades 1880-1889 and 1900-1909, in each of which it is estimated that more than 40 per cent of the population increase came from immigration. Some authorities, starting with the famous American economist, General Francis A. Walker, have held that immigration does not bring about an increase in the population over and above what it would otherwise have been; but that on the contrary the immigration reduces the birth rate among the population already here, and the immigrants thus merely replace the domestic stock that would otherwise have been born. There is no way, of course, that one can prove or disprove such a theory, since there is no way to know what the birth rates would have been in the

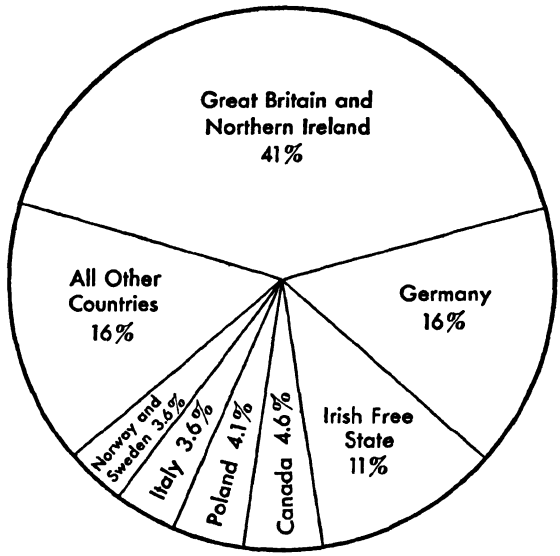


FIG. 12 • 5.—Make-up of the white population of the United States in 1920 by national origin.

absence of immigration; yet careful study of birth rates and immigration statistics leads to the conclusion that, although Walker is partly right in that the birth rate has declined as immigration has increased, nevertheless there has been an actual and significant net addition to the population by immigration.[1]

12 • 10. Regulation of Immigration to the United States. Countries pass various sorts of legislation affecting immigration, most of which can be classified under one or another of three headings: (1) legislation designed to influence the quality of immigrants by forbidding the immigration of persons with undesirable characteristics is called "selective legislation," (2) legislation that prohibits entirely the immigration of persons from specified areas

[1] Thompson and Whelpton, *op. cit.*, pp. 304-311.

is called "exclusive legislation," and (3) legislation that sets quantitative limits to the number of immigrants who may be admitted is called "restrictive legislation." Alternatively, we may say that nations follow the policy of selection, exclusion, or restriction.

The United States, as we have said, encouraged immigration throughout its early history. Far from trying to restrict or reduce the numbers, efforts were made to increase them. But, as early as 1882, the United States began to apply both selective and exclusive policies. The selective policies have been intended to prohibit the admission of immigrants who were morally or physically undesirable, and to this end the law has been amended again and again in efforts to specify the qualities that are not acceptable in immigrants. The law at present lists, as prohibited classes, "idiots, imbeciles, feeble-minded persons, persons who have had one or more attacks of insanity at any time previously; persons of constitutional psychopathic inferiority; persons with chronic alcoholism; paupers; professional beggars; vagrants; persons afflicted with tuberculosis in any form or with a loathsome or dangerous contagious disease . . ." and so on, including anarchists, felons, those who advocate the overthrow of the government by force, political assassins, etc. The list is a long and a growing one. Perhaps the most controversial of the selective provisions is that which specifies that no one may be admitted who cannot read or write some language, the so-called "literacy test," which was long a subject of public controversy and which was finally passed over President Wilson's veto.

Since 1882, the United States has practiced exclusion of persons from certain areas, which have been designated in the law by laying out certain boundaries of latitude and longitude and specifying that people born within those limits may not be admitted. In practice the boundaries have been so drawn as to exclude natives of most of Asia, especially the yellow and brown races. Recently, and purely as a wartime gesture, an exception has been made which grants China an immigration quota of 100 persons per year. Various reasons have been suggested for the exclusion of these people. The layman usually imagines that it was done because of the competition that we would otherwise have to face from low-paid oriental labor; but of course every immigrant, whatever his source, raises this problem. It is but natural that immigrants usually come to the United States from areas where wages are lower and, if we were to exclude immigrants from such areas, we would have to exclude almost all immigration. Actually the major difficulty with oriental immigrants was that they were not easily assimilated into our population. They tended to segregate themselves into communities of their own, retaining the language and customs of their mother country, and not intermarrying with the population here.

12·11. Restrictive Legislation in the United States. At the time of the first World War certain problems arose which led many Americans to doubt the wisdom of continued unrestricted immigration. In part, perhaps, the ordinary nonwar tendencies would have raised the same questions, as national resources became more and more scarce and as population grew larger and larger. But the war raised the problem that men then called "hyphenated Americanism." We discovered that there were thousands of people living in the United States who felt a very definite allegiance to some other country. The advent of war had in itself cut immigration to about a quarter of its ordinary level, and with the entry of the United States the flow almost stopped. But as soon as the war was over, the movement started again, promising soon to reach its old level (see Fig. 12·2). Pending further study and the establishment of a permanent system, Congress passed a temporary immigration act, effective in 1921. Under this act countries were divided into three groups: (1) those which were not allowed to send immigrants at all under the exclusion acts, (2) those which were still allowed to send immigrants without limit, and (3) those which were allowed to send a limited number of immigrants each year. We have already said that the first group came mainly from Asiatic countries. The second group comprised the countries of North and South America, which were allowed, and still are allowed, to send immigrants without limit—the so-called "nonquota countries." Our major attention will be given to the third group, and particularly to the way in which the quotas for the various countries were set.

Under the temporary act of 1921, any country that was given a quota was allowed to send each year a number of immigrants equal to 3 per cent of the number of its nationals who were living in the United States in 1910. If we wanted to find the quota for, say, France, we looked in the 1910 census and found that there were 190,000 people living in the United States in that year who were born in France. The French quota was therefore 3 per cent of 190,000, or 5700. The quota for every other country was set in the same way, with quotas varying from a high of 77,000 for the United Kingdom and 68,000 for Germany down to 92 persons for Luxemburg.

By 1924, Congress was ready to adopt a permanent immigration policy, and the act passed in that year is still in effect. Under this act of 1924 the quotas were changed temporarily, pending study that we shall mention shortly, to "2 per cent of 1890" instead of "3 per cent of 1910." That is, temporarily the quotas were altered by giving each country a quota equal to 2 per cent of the number of its nationals who were living in the United States at the time of the 1890 census. If, for example, we again look up the facts on France, we find that at the time of the 1890 census there were 198,000 people living in the United States who were born in France. The French

quota was therefore set at 2 per cent of this number, or at 3960 persons per year.[1] Under the 1924 act, no quota was to be set at less than 100, so, if the computations for any country yielded a quota that was smaller, the country was automatically given a quota of 100 people. The quotas under this system ran from the highs of 51,000 for Germany and 34,000 for the United Kingdom down to 100 for several countries.

The student will notice that the change in 1924 reduced the number of immigrants from 3 to 2 per cent, thus reducing the number eligible for admission. It also shifted the base back from 1910 to 1890. This, of course, reduced the number admissible even more, but the main purpose of the shift to 1890 was to influence the source of immigration rather than the quantity. The 1890 census was taken only a few years after the shift from the old to the new immigration. It was still true in 1890 that most of the people resident in the United States who had been born abroad had come from northwestern Europe, while by 1910 most of the foreign-born residents were from southeastern Europe. Thus Congress, in shifting the base back to 1890, was indicating a preference for immigrants from northwestern Europe as compared with those from the southeast. While the change reduced the German quota from 67,607 to 51,227, it reduced the Italian quota from 42,057 to 3,845. The total number admissible under quotas—357,803 under the 1921 act—was now reduced to 164,667, a reduction of 44 per cent.

But Congress had decided in 1924 to change entirely the basis for quota making. The quotas under the temporary act of 1921 and the temporary quotas under the act of 1924 had been based on foreign-born population alone. Thus, if a man of English extraction but American birth was living in the United States in 1910, he was not counted at all in determining the quota. Only those living in this country and born abroad were included in the count for quota making purposes. Congress felt that it would be fairer or more desirable to base the quota for France, not on those who were living in the United States at a particular time and who had been born in France, but on the total number of people in the United States of French extraction. It was decided to shift for a permanent system to what is called a "national origins" basis. Consequently, the act of 1924 set up the temporary quotas just described, based on "2 per cent of 1890," but also set up an organization to determine the national origins of the American people as a basis for permanent quotas. The study was a difficult one and the outcome by no means satisfactory, since it turned out to be impossible to make very accurate determinations of the national origins. However, after one or two postpone-

[1] The figures have been rounded off in the illustrations of this and the preceding paragraph. The actual French quotas were 5729 under the 1921 act and 3954 under the temporary quotas of the 1924 act.

ments, rough estimates were finally made, and new national origins quotas went into effect in 1929 and have been in operation ever since. These quotas were provided for under the act of 1924 and are the permanent quotas under that act. Thus we have had three sets of quotas:

1. The quotas under the temporary act of 1921, equal to "3 per cent of 1910."
2. Temporary quotas under the act of 1924, equal to "2 per cent of 1890."
3. Permanent quotas under the act of 1924, based on national origins.

The act of 1924 specified that, as soon as the national origins were ascertained, the quota for any country was to bear the same proportion to 150,000 which the number of inhabitants of the United States in 1920 owing their origin to that country bore to the total population of the United States from all nationalities subject to the quota law. This statement is rather involved and cumbersome. Let us illustrate: The total number of persons in the United States in 1920 who were descended from stocks of quota countries was 89,507,000. Of these, 1,842,000 were estimated to be of French descent. Thus France had 2.06 per cent of the quota population of 1920, and her quota was therefore 2.06 per cent of 150,000, or 3086 persons. Similarly the quota for any country could be found by a corresponding proportion, as follows:

$$\frac{\text{Country's quota}}{150{,}000} = \frac{\text{country's descendants}}{89{,}507{,}000}$$

In this proportion the phrase "country's descendants" means the number of people who were living in the United States in 1920 and who were thought to be descended from stock of the country in question. Obviously since the two denominators are constants we can restate the formula as follows:

$$\text{Country's quota} = 0.0016758\ (\text{country's descendants})$$

If we multiply the number of the country's descendants in thousands by 1.68 we shall find the approximate quota. If we take our example again, the number of French "descendants" was estimated as 1842 thousand. Multiplying by 1.68 we get,

$$\text{French quota} = (1.68)\ (1{,}842) = 3095$$

The actual French quota was 3086. Our answer is only approximate because we have rounded numbers. Perhaps it would be close enough to say that we can find the approximate quota for any country by taking one and two-thirds times the number of thousand descendants of that country. This would give an approximation of 3070 for the French quota.

Under the national origins system, many countries would have received extremely small quotas save for the provision, still retained, that no country that was granted a quota would be assigned one smaller than 100. Under

this provision approximately 40 countries are given quotas of 100, and the total number of quota immigrants therefore exceeds 150,000. If every country filled its quota, the total number of quota immigrants would be just under 154,000 per year.

It must be remembered, of course, that immigrants are still allowed to come without numerical limit from the other independent countries of North and South America. Thus in 1944, out of a total of 28,551 immigrants,

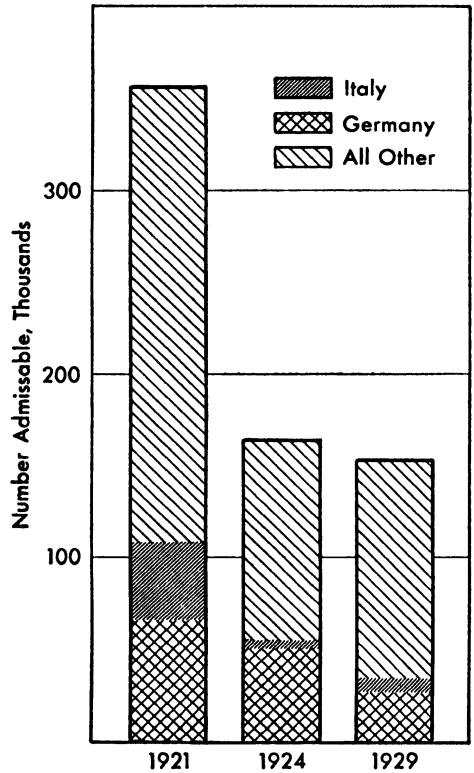


FIG. 12 • 6.—Total numbers admissible under quotas under the three American quota acts.

23,084 or over 80 per cent came from nonquota countries. There is really, then, no numerical limit on total immigration, but the limit on quota immigration, established at 357,803 in 1921 and lowered to 164,667 in 1924, was finally set at about 154,000 in 1929 and has stayed there ever since.[1] The relative sizes of these quotas are shown in Fig. 12 • 6.

12 • 12. Immigration Regulations. So far we have been describing the

[1] The total under the national origins system varies slightly from year to year as national boundaries vary, but it has always been between 153,000 and 154,000. On June 30, 1944, it was 153,879.

purely numerical aspects of the immigration laws. There are many more things that could be said, and much more that one would have to know if he were to have any thorough understanding of our immigration policies and procedure. We shall confine our remarks to a few additional points and to the situation under the present act.

Every effort is made before a prospective immigrant leaves his homeland to determine whether or not he is eligible for admission to the United States. This is done in order to escape the difficulty, common when we first started immigration restriction, of having people come to this country only to find out at the last minute that they were not admissible. At present, prospective immigrants must obtain a visa on their passports from the American consul in their home country. These consuls are directed to determine whether or not applicants are feeble-minded, have prison records that would bar them, or are otherwise inadmissible. Also the consul in each country is required to keep track of the number admitted each year, so that he can tell whether or not the quota for the country to which he is assigned has been filled. If he finds that the applicant is admissible, he puts his visa on the passport, signifying that the applicant may come to the United States.

Aliens entering the United States are divided into three classes or groups: quota immigrants, nonquota immigrants, and nonimmigrants. Only those of the first class are subject to the quota restrictions that we have been describing. Nonimmigrants would include such persons as the ambassadors and other representatives sent here by foreign countries, tourists or others who are here temporarily for travel or business, aliens traveling through the United States on their way to some foreign destination (as a Mexican traveling from Mexico to Canada), aliens once lawfully admitted who later go from one part of the United States to another through foreign territory (as an alien who is living in New York and who goes to the state of Washington over the Canadian National Railways), foreign sailors touching at American ports, or foreign businessmen who enter this country solely to carry on trade between the United States and their own homeland under a treaty of commerce and navigation. Nonquota immigrants include husbands, wives, and minor children of American citizens under certain restrictions; nationals of independent countries of North, Central, or South America; immigrants previously lawfully admitted and now returning from temporary visit abroad; ministers and college professors under certain restrictions; women who were born in the United States but lost their citizenship through marriage to foreigners; students coming temporarily to study, etc.

As has already been said many classes of aliens are denied admission entirely: those who are seriously defective physically, mentally, or morally;

those who advocate overthrow of government by force; contract laborers, who have been induced to come to this country by offers of employment; those who seem likely to become public charges; members of the so-called "yellow races" unless exception is made, as it now is with China, by special treaty; persons unable to read any language, etc. The laws on these matters are long and complicated, and only the most general description has been attempted here.

12·13. Economic Effects of Immigration. There can be little question but that, in most cases, immigrants themselves improve their lot by migration. Even though recent immigrants are, on the average, among the poorer people in the United States, they are in most cases considerably better off than they were in the "old country." To be sure, there are certain psychic and spiritual losses that are bound to occur when the immigrant finds himself in a new environment, unable to express himself adequately or to understand exactly what is going on, seeing his children learn new customs, strange to him, and grow away from old customs and habits that he holds dear. But there can be no question that the gains on the average far overbalance the losses and that the migrating people are themselves better off as a result of the change. As for their fellows in the "old country" whom they leave behind, it is somewhat more difficult to be sure whether the result is desirable or undesirable. On the one hand, population pressure and competition for jobs are reduced, and unemployment becomes somewhat less likely. On the other hand, a country that has borne the full expense of raising and educating these people now loses them before they have returned their product. The emigrant is often above the average for his group in initiative and energy, and the loss may well be an important drain on the human assets of the community. In recent years many countries have tried to encourage large population for political or military reasons, and in such a case emigration of people in the prime of life would also be looked upon as disadvantageous. This problem we shall discuss shortly. From the standpoint of the country to which the immigrant comes, much depends on the stage of economic development. The problem is very different in a young, new country, which is underpopulated in proportion to its natural resources, from that in an old densely populated country. The receiving country, to be sure, is getting workers for whose training and rearing some other country has borne the expense; it may be receiving workers with particular skills that are needed, or potential soldiers for its armies, or workers to man industries that could not otherwise have been developed. On the other hand, as we have already seen,[1] after a certain stage is reached in a country's development, additions to the population reduce the average real income and the level

[1] See the discussion of diminishing productivity, pp. 180*ff*.

of living. Immigration may improve the lot of the immigrant, but only with a concomitant hurt to the society receiving him. This harm will not be present in the early stages of a country's settlement, and this is the reason for the change of attitude toward immigration which took place in the United States, shifting us from a policy of encouragement and sympathy to one of discouragement and restriction. Also we should not overlook the sociological and spiritual problems that arise in a society of heterogenous elements, not understanding each other and not sympathizing with each other's aims and ideals.

12·14. How Large a Population Is Desirable? We have now considered the two major sets of forces that determine the size of a country's population: the natural reproduction rate and migration. Throughout most of its history the United States has been growing as the result of the joint effect of both these forces, although the natural reproduction rate has been growing smaller and immigration has been reduced to a very small fraction of its former amount. In fact, there were a few years during the depression of the 1930's when the United States suffered a net loss through migration, the number of emigrants exceeding the number of immigrants. Yet, on the whole, the population of the United States has been growing, at first rapidly, and of late years more and more slowly (see again Tables 12·1 and 12·2.) The 1940 census showed a population of something over 130 million people. How large should the population grow? Should we look ahead to a population of 200 million, 500 million, or a billion? Is there any limit, any optimum size for the population of the United States?

This problem is common to all areas and can perhaps be better visualized if we consider a smaller area. In the south Pacific Ocean lies Pitcairn Island, with an area of about 2 square miles. Previously uninhabited, it was settled in 1790 by 15 men and 12 women, a population of 13.5 per square mile. By 1856, the population had increased to 193, or almost 100 per square mile. At this time most of them did in fact move away but, if the population had continued to increase at the same rate, the 1900 population of these 2 square miles would have been 716, the 1950 population would have been 3175, and the population in the year 2000 would have been 14,000. How many people could Pitcairn Island support? Sooner or later, obviously, additional population could spell nothing but misery, poverty, and even starvation. Just how large the population could grow before the effects became deleterious is hard to say, but that some point would be reached beyond which larger population would be accompanied by lower levels of living is indisputable. And this is true not only for Pitcairn Island, but for every area on earth.

Unfortunately there are many people who look on population as an end in itself. They are pleased when population increases and alarmed if it does

not, or if the rate of increase slows down. We should by now realize that increased population may in some cases be socially beneficial and in other cases socially harmful. In the earlier parts of this chapter we were acquainting ourselves with the facts about birth rates, death rates, migration, etc. It has not been our intention to imply that the actual changes in birth rates and death rates have been desirable nor to insinuate that they have been detrimental. We are now at the point where it should be said that population growth may be either harmful or helpful to the people of a country, and that the policies followed presumably should be very different in different places and under different conditions. When we ask how large a population "ought" to be, we are, of course, asking a question to which no answer can be given unless we agree on some basis for evaluating the results. If we agree (and all men do not) that we want to get the highest possible level of per capita real income for the people of the country, then we must realize that sooner or later increases in population will be disadvantageous, that at some point it would be desirable for the population of the United States to stop growing larger. We have already pointed out that the rate of population increase is diminishing in the United States. No one can know exactly what lies ahead, and estimates must be made on the basis of fallible assumptions as to future birth rates, death rates, and migrations. On the basis of various sets of such assumptions, experts have predicted that the population of the United States will become static anywhere from about 1970 to 2000, and that its maximum population will be between 150 and 180 million people. Yet a prudent man will realize that the entire situation might be changed overnight by new discoveries and inventions that might make it possible to support 500 million people as easily as we now support 130 million.

12 · 15. **Malthusian Doctrine.** We have already seen that population is the result of birth rates, death rates, and migration. Students of public affairs have long been interested in the forces that influence the size of population and, in 1798, one of their number, an Englishman named Thomas R. Malthus, published his "Essay on Population," a book that has done much to influence the thinking of social scientists and statesmen ever since. Although more recent knowledge has led us to believe that Malthus was mistaken in some of his conclusions, nevertheless the fundamental theme of his argument is generally accepted today. Malthus started by pointing out that birth rates could be much higher than they were and that populations could grow at more rapid rates than they do. "Through the animal and vegetable kingdoms," he said, "Nature has scattered the seeds of life abroad with the most profuse and liberal hand." Population constantly tends to increase faster than the food supply. But, said Malthus, there are two kinds

of checks that may limit the population. If it grows too large, there will not be food enough for all and the death rate will rise from starvation, famine, disease, pestilence, and war which will come as a result of the overcrowding. These forces, which tend to increase the death rate, Malthus called the "positive" checks. Of course, it is possible (Malthus really thought it rather unlikely) that men will voluntarily limit the population before they reach the sad state where the positive checks come into play. By reducing the birth rate, they can escape the difficulties described. Thus anything that will reduce the birth rate—late marriage, celibacy, sterility, birth control—will keep the population in bounds just as effectively as will an increased death rate. The factors enumerated in the preceding sentence, and any others that tend to reduce the birth rate, were called by Malthus the "preventive" checks, the "prudential" checks, or the "negative" checks. The size of the population, then, was limited by the food supply, and Malthus thought that, whenever anything occurred to increase the food supply, the population would increase to the full extent possible unless held in check by some very powerful forces. The only forces that could act were moral restraint (the negative checks) and vice and misery (the positive ones). In practice Malthus rather thought that the positive checks were the more important and that populations were likely in fact to increase to the very limits that could be supported by the food supply.

It is true that Malthus went further than these general statements. He became specific and quantitative, and it was on these specific and numerical points that he came to grief. Malthus thought that populations tended to increase at a constant percentage rate, in geometrical ratio, doubling themselves every 25 years unless checked, while food supply increased only arithmetically. Thus, while population increased in the series 1, 2, 4, 8, 16, 32, and 64, the food supply increased in the series 1, 2, 3, 4, 5, 6, and 7. There seems no reason today to suppose that these figures are correct or that any particular geometric and arithmetic series are correct. But much of the criticism of Malthus has been directed at these series of numbers without realizing that his most important conclusions do not rest particularly on them.

We can, perhaps, illustrate the Malthusian doctrine very roughly and baldly (and certainly not entirely exactly) by taking this example. Let us suppose that we have a family of man, wife, and two children with an income of $1000 a year. What will happen if their income is increased to $2000 a year? Malthus would have said that they could: (1) support a larger family at the same level which they enjoyed before, (2) Support the same family of four at a much higher level, or (3) increase the size of the family somewhat and at the same time increase the level of living somewhat. He thought populations tended, in general, to accept the first of these three possibilities—

to increase numerically whenever anything happened to increase subsistence. Thus he thought that populations always expanded until held in check by the lack of subsistence—by the positive checks.

Malthus wrote just as the Industrial Revolution was getting well under way. He had no way of foreseeing the tremendous increase in productivity that would come as a result of power machinery, electricity, railroad transportation, etc. Within a relatively short time after he wrote his essay, the world witnessed one of the greatest increases in output per man in history. We may well ask, then, what did happen as a result, and whether or not Malthus' predictions turned out to be well founded. We find when we look at the record that the people of Western civilization actually chose, at least for the time being, the third rather than the first of the possibilities listed in the preceding paragraph. European population had been roughly and approximately static for some time before the Industrial Revolution, but from the time of Malthus' first essay in 1798 until the outbreak of the first World War in 1914, the people of European stock had increased from about 200 to about 600 million,[1] increasing in a century and a quarter from about one-quarter to about one-third the population of the earth. Thus when the discovery of new territory and new techniques made it possible to consume more, people did in part adjust by larger families, as Malthus had supposed they would. But, contrary to his expectation, they did in part take it up by voluntarily limiting the sizes of their families in order to enjoy a higher level of living. We have already noted that the average level of living increased markedly during this period. To be sure, no one knows what would happen if the Industrial Revolution were now to stop—if we made no new inventions and perfected no new techniques. It is possible that, under those circumstances, Malthus might yet turn out to be right and that the final adjustment would be one at or near the starvation level—one where the primary limits were disease and misery and vice rather than voluntary limitation as the result of yearning for a higher level of living—one where the positive rather than the negative checks played the major role. Be that as it may, Malthus did the great service of emphasizing for us the fact that the size of the population does make a difference and that, unless population is held in check by other means, it will most certainly be held in check by starvation and disease. Those people who thought that the latter checks were inevitable earned for the economics of their day the appellation of "the dismal science." Modern economists would differ from Malthus primarily in pointing out that the future of any people is largely in their own hands.

This last statement, by the way, furnishes a clue to the seeming paradox

[1] THOMPSON, *op. cit.*, p. 1. Note that the figures given are for people of European stock, not merely those living in Europe.

that many economists, who favor free trade in goods among nations because of their understanding of the principle of comparative advantage, nevertheless favor restriction of immigration. They would argue that, if the people of an area decide that they wish to get the advantages of a higher level of living by means of voluntary limitation of population, they then should not be forced to share their gains with those of other countries who chose the other course of increasing the population and forcing down the level of living. Obviously such a view involves difficult decisions in the field of ethics which cannot be settled by any collection of statistics nor by any amount of scientific reasoning.

12·16. Population and Diminishing Productivity. The economist's interest in the problems of population is very different from that of the census taker. The economist fixes his attention, not on the size of the population itself but on its effects. We recall from Chap. 10 that, if some factors of production are kept fixed while others are varied, the average and marginal product of the variable factors diminishes. In practice the natural resources (land) of any society are at least approximately fixed, while the population can and does vary. Birth rates, death rates, and migration all bring about changes in the country's human resources. If, as is often the case, the population grows larger and larger with fixed amounts of land, there is the tendency for the marginal and the average product of labor to diminish. Unless the tendency can be offset by the discovery of new techniques, the people become poorer and poorer. Every statesman must, therefore, be interested in the forces that determine the size of his country's population.

12·17. Summary. We have seen that the human resources of a nation are important in determining how its people live, just as are the natural resources. The qualities of the human resources are in part the result of heredity and in part the result of environment and training, and every society is forced to take various kinds of direct or indirect steps to influence each. The quantity of human resources is in part the result of the natural reproduction rate and in part the result of migration. As to the former, we have witnessed in recent years reductions in both birth rates and death rates. The birth rates are higher than the death rates, but are falling faster, and if present trends were to continue the natural reproduction rate would become zero within the next generation or two. The birth rates are very different in different groups within our society, although many students think that this tendency for a differential birth rate is tending to disappear by itself. As for migration, the United States has changed its policy from one of encouragement and stimulation to one of discouragement and restriction, and an immigration that was once the largest mass movement of people in the world's history has shrunk almost to the point of disappearance. With the natural reproduction rate falling and with immigration falling at the same time, the rate

of population growth of the United States has decreased markedly, and some experts expect that the population will become static within the lifetime of people now living. Whether this would be a desirable thing or not is a matter of dispute; at least we must admit that there is a limit to the population that can be supported in the United States, and that long before the maximum possible limit is reached we would be forced to a bare subsistence level of living. Sooner or later we reach the point where societies have to choose between larger population and poverty on the one hand, and static population with present levels of living on the other.

Questions

1. A chauffeur drives an automobile for a living, while another man drives his car for pleasure. Both men do exactly the same thing, yet the economist would say that the activities of the one were labor and those of the other were not. Does this seem reasonable to you? Why or why not?

2. In Sec. 12 • 2 there is some comment about the number of Indians that the natural environment of the United States supported, and the number of white men now supported in the same natural environment. In Sec. 11 • 1 we learned that natural environment and natural resources are not identical. Putting the two together, would you say that the natural resources of an area are influenced by the human resources? Explain.

3. What laws, if any, does your state have that are intended to influence human heredity?

4. A city of 142,000 has 2840 births in a year. What is the birth rate?

5. Suppose we have two adjacent towns, one with a modern maternity hospital and the other with none. What will be the effect on the birth rates in the two towns? Why? Will the birth rates reflect accurately the population changes of the two towns? Why?

6. Suppose that large numbers of the natives of surrounding states move to California in their old age when they retire. What will be the effect on the death rate in California? In the other states? Why might it be true that the states with the most healthful climates might nevertheless show the highest death rates? Why might the death rate be expected to rise rather than fall in a city where a fine, new hospital was built?

7. On the basis of your discussion of questions 5 and 6, explain some of the reasons why vital statisticians hesitate to work directly with crude (uncorrected) birth rates and death rates.

8. A city has 296 births and 218 deaths in a year. The birth rate is 14.8 and the death rate is 10.9 per thousand. What is the population of the city? What is the natural reproduction rate in the city?

9. In a stationary population, the death rate is 20 per thousand. What is the birth rate? What is the span of life?

10. Why is it that one cannot confidently expect continued reductions of the death rate in the United States?

11. Why have cancer and heart disease become more and more important as causes of death in the United States?

12. "If we could find complete and absolute cures for every disease and ailment of mankind except old age, we would within a relatively few years find 100 per cent of our people dying from old age, and alarmists would say that old age was a much worse 'disease' than it used to be." Comment.

13. If we say that the expectation of life at the age of twenty is 48 years, and if you are today twenty years old, does this mean that you will live until you are sixty-eight years old? Explain.

14. When we say that the expectation of life at birth was about 18 years in 1790, does this mean that at that time everyone died by the time he was eighteen years old? What does it mean?

15. Explain what caused each of the great declines in immigration to the United States evident in Fig. 12•2. For example, what caused the declines that appear to have centered at about 1862, 1878, 1897, 1918, and 1933?

16. Is it possible for a country to practice restriction, exclusion, and selection all at the same time in its immigration polices? Explain.

17. Can you name all 16 of your great great grandparents, and tell from what country each came? How would you try to ascertain the national origins of the American people? Why could you not do it merely by adding together the total number of all past immigrants from each foreign country?

18. "In setting up the national origins system, the effort was made to see to it that the racial mixture entering the country contained the same nationalities in the same proportions as did the mixture already here in 1920." Do you take it that this is an accurate statement? If so, why? If not, what were the exceptions to it?

19. Suppose that at the time of the 1920 census there were only 215 people living in the United States who were of Danish extraction. (This is not correct, but will serve as an example.) What would be the Danish immigration quota under our present system?

20. Suppose a country's present immigration quota is 3916. How many people of this country's stock were in the United States in 1920?

21. What population do you think you would choose for the United States if you could fix it at any figure you desired, assuming present knowledge and methods of production? What factors would you consider?

22. In Sec. 12 • 14 appears the statement that all men do not agree that we should try to get the highest possible per capita real income for the people of a country. What other goals seem to you defensible?

23. List several positive checks to the growth of population. Do you think that it would be fair to say that either or both of the two great world wars were in any way caused by the pressure of population growth? Would these wars be positive or negative checks to population?

13. CAPITAL

13·1. Nature of Capital. As we continue our study of the factors of production, having looked at problems connected with land and with labor, we turn now to those of capital. The economist defines capital as *those things made by men to assist them in the production of other goods rather than for direct satisfaction of wants.* To be sure, land usually gives no direct satisfaction to wants and is used in production, but land and capital differ in one important respect: land is a gift of nature, while capital is made by men. Land exists whether men make it or not, but men never get capital unless they make it. We sometimes call these tools and machines "capital," "capital goods," "production goods," or "intermediate goods." The student has perhaps been accustomed to thinking of capital as a social group or class of persons, as when he differentiates "capital and labor." He remembers that in the cartoons Capital is represented as a fat man with long-tailed coat and top hat, while Labor is represented as a working man in apron and with dinner pail. To the economist capital is machinery, tools, and equipment—buildings, locomotives, highways, looms, steam shovels, hammers, saws, and vacuum cleaners. The owner of the capital is called a "capitalist," and in this sense the man who hires himself out to work on the town road, supplying his own team of horses and dump cart, is a capitalist as well as a laborer. He is a laborer because he is supplying his own effort; a capitalist, because he is paid for the use of his equipment.

The student is apt to fall into the error of thinking of capital and money as synonymous. This is perhaps because we measure amounts of capital in value terms and say that a corporation has a capital of, perhaps, $500,000. This is just a shortened way of saying that the corporation has capital "worth $500,000." Its actual capital will consist largely of such things as buildings, machinery, and raw materials. Of course, money is capital. Ordinarily it does not give any direct satisfaction to wants but is desired because it is useful in getting things that do yield satisfaction. It may seem strange for the moment to say that money is "productive," especially since it has been common to say

that money is sterile; but we shall see later[1] that money is actually productive, and therefore can properly be classified as part of the capital of society. But it is only part, and a rather small part, of the capital.

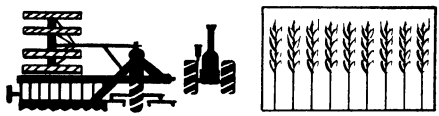
13·2. **How Does Capital Arise?** In order to get capital it is necessary, of course, to have land and labor first. In this respect capital is like consumption goods: it is a joint product of other factors of production. Capital differs from consumption goods in one important respect, which can best be made clear by means of an illustration. Suppose that the staple article of diet in a primitive society is locusts, as it was for many of the Indians of the Southwestern United States. Imagine that the Indians are getting their locusts without the use of any equipment, merely catching them in their hands and eating them. One of the Indians, however, gets the idea that he could catch more locusts if he had a net to trap them with, so while his fellows go out to catch locusts he stays at home for a day or so and fashions a net. The net, of course, is capital, and we see that it could come into existence only if the maker was willing for the time being to give up production, which he could have had, by foregoing the hunt for the day while he made the net. For the time being he had to go without locusts which the others were enjoying. To be sure, he hoped to get even more locusts later with the net than he could without it, but in order that capital may be produced men always have to sacrifice present consumption for the sake of future consumption. They have to abstain —to sacrifice what they could have now for what they hope to have later. In order that the capital can exist in society, we need not only land and labor, as in the production of consumption goods, but also sacrifice or abstinence or waiting.

It seems to be a common characteristic of men that of two equal satisfactions they prefer a present to a future one. If they have to choose between a given satisfaction today and the same satisfaction a year from today, they will choose today's. To be sure, they might prefer next year's satisfaction to today's if next year's were greater, but if the satisfactions are equal men will ordinarily choose the present one. This human tendency is called "time preference." Because of it men do not make capital unless they see a greater satisfaction from its use than they anticipate without it. If our Indian can catch no more locusts with his net than without it, he will not make the net. He would be foolish to sacrifice the consumption of a day or so while he made the net unless he expected that it would increase his productivity. We often note this tendency when capital is measured in monetary terms by saying that, when a man invests his money, he insists on getting back both the principal and interest; he does not give up voluntarily the things that he could buy with his money now unless he gets back enough money to buy even more later on. Just as truly the Indian of our example expects interest. He

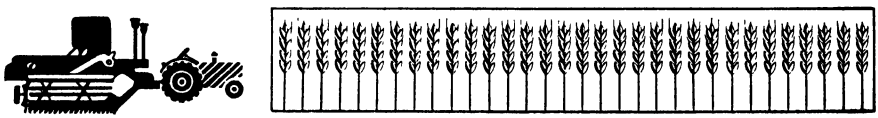
[1] See Sec. 15·3.

expects to get more locusts by investing his time in the net than he could get by investing the same time in locust chasing. While we fall into the habit of thinking of interest as a monetary phenomenon and while some people have even misled themselves into thinking that they could abolish interest if they were to abolish money, as a matter of fact we have interest wherever men have time preference, because such men do not postpone present consumption unless they are paid to do so. The man who puts $100 in the bank and a year later withdraws $105 says that he got back his $100 with interest. The Indian who spends 10 hours making a net and catches with it as many

Hand Methods



Early Machines



Modern Machines

FIG. 13 • 1.—Increased production through use of better equipment. Each head of grain represents one unit harvested by one man in a day. (*Based on National Resources Board, Report on National Planning and Public Works in Relation to Natural Resources, Washington, Dec.* 1, 1934, *p.* 364.)

locusts as he could catch in 12 hours without the net is getting back his 10 hours of locusts and 20 per cent interest.

13 • 3. Productivity of Capital. Men pay for the use of capital for the same reason that they pay for the use of land or for the labor of other men —because it is productive. Contrast the product obtained by two farmers, both of equal ability and both cultivating potato fields of the same size. One of them uses modern machinery, and the other uses no tools or equipment at all—not even a plow, a harrow, a hoe, or a spade. The latter man, without tools, has to tear up the sod with his fingernails and harvest his crop the same way. There is no difference in the land or the labor employed in the two cases. The difference in the product is the result of the use of capital. The capital is productive. Karl Marx, famous socialist or communist writer,

pointed out that labor did not get the entire product—that there was a surplus value, which somehow was filched from the laborer by the capitalist. He did not realize that labor did not produce the entire product—that, if labor alone had been used, the product would have been smaller. Neither did he realize that, if men had not been paid to produce the capital, it would not have been produced, and the product would have been smaller as a result. We can perhaps visualize the productivity of capital somewhat better by inspecting Fig. 13 · 1, which shows the numbers of units of grain harvested by the primitive hand methods using the flail and cradle, by means of the earliest harvesting machines, and by means of modern machines. To be sure, even the smallest product, obtained by the cradle and flail, was obtained by using capital, for a cradle or a flail is capital. But, as the investment in capital increased, the productivity increased also, and the productivity of the capital is self-evident.

Many other examples could be given illustrating the productivity of capital. We are told,[1] for example, that when Elias Howe invented his first crude sewing machine he was troubled to find that no one took it seriously. "It was looked upon as a mere curiosity. But Howe had plenty of confidence in the machine. He challenged any five sewers to do as much work as his one machine. This bold challenge was accepted, and five most expert needlewomen in a large factory were selected for the trial. Ten garments were cut, each garment being exactly alike, and five of these were given to the machine while each of the sewers took one of the other five. The race was begun and anyone watching these experts would have thought the machine was undertaking an impossible task in proposing to sew five times as quickly as one of the sewers. But before these experts had half of their garments done, the machine cried a halt, having completed the whole of its similar task." Again, the Commissioner of Labor in his Thirteenth Annual Report stated that a careful study showed that "the total time required for the production of 12 packages of 1 pound each of pins under the machine method was 1 hour and 33.9 minutes as against 140 hours and 55 minutes required under the hand method—a ratio of about 90 to 1 in favor of the modern method." The same report shows that it took 2 hours and 42.8 minutes of man labor to grow 30 bushels of barley by the methods commonly used in 1895, as compared with 63 hours and 35 minutes by the methods in common use 65 years before. Another authority[2] estimated for various farm crops how much the crop would be reduced if it were necessary to produce it with the same amount of labor actually used, but by old-fashioned hand methods rather than

[1] *Bulletin* 662, U. S. Department of Labor, Bureau of Labor Statistics, November, 1938, pp. 24-25.

[2] QUAINTANCE, H. W., The Influence of Farm Machinery on Production and Labor, *Publications of the American Statistical Association*, Third Series, vol. v, no. 4, pp. 19-27.

by machine. He estimated that the United States barley crop would have been reduced by 95.7 per cent, corn by 60.9 per cent, cotton 64.8 per cent, hay 81.3 by per cent, oats 89.2 by per cent, potatoes by 65.1 per cent, rice by 72.5 per cent, rye by 60.0 per cent, and wheat by 94.5 per cent. To be sure, these and other such illustrations are apt to be faulty in one regard. They count the labor that goes, under present methods, into growing corn, but not the labor that goes into making the machinery used. If we were to get a fair comparison, we should compute on the one hand all the hours of labor that go into growing corn when no machinery is used, and on the other hand we should add together the labor used to operate machines and to make machines and to mine ore from which to make machines, etc. This is a completely hopeless task. But we can, of course, approximate it by comparing the *cost* of producing corn with and without machine, since the cost of this past labor will be covered in the cost of the machinery. When we make such comparisons, the great saving made by the use of machine methods—the great productivity of capital—is still startlingly self-evident. In fact, if we were to try to operate our society with land and labor alone, abolishing all machinery and equipment, a very large proportion of the population would starve to death in short order.

13·4. The Roundabout Process. It might seem the natural thing, if one wanted potatoes, to go out and find them—to use one's efforts in hunting for a supply of the things wanted. But modern man usually goes at the thing indirectly. Instead of starting out to get potatoes, he starts by mining ore and building buildings to make tools: hoes, rakes, potato diggers, spray equipment, bags, motor trucks, tractors, and storage houses. The superiority of man's mental processes as compared with those of other animals makes it possible for him to see that this indirect or roundabout process will, in the long run, yield even greater satisfaction of wants than would the more direct and immediate effort spent on getting potatoes. The use of capital always involves this roundabout process—the expenditure of effort on tools and equipment first, in order that they may be used in the production of consumption goods later. The process of production is prolonged and spread out over many steps, the earlier ones of which have only an indirect and tenuous connection with the ultimate product. For example, the man who mined the ore to make the saw that was used by the carpenter to make the building in which the potato digger was manufactured—this first miner was exerting effort which was ultimately to materialize in the form of French-fried potatoes, but he did not know it. It may well be that this first effort was exerted 5 or 10 years or more before wants were ultimately satisfied by the potatoes. The ore, when first mined, could have been used for many sorts of things, and no one could then tell that it was destined to play any part in the potato business. We would say that the capital was at this stage un-

specialized. As soon as it was made into a saw, however, the uses to which it could be put were somewhat more circumscribed and limited, and the building that was created with its aid was even more specialized indeed. There are very few things for which a potato digger can be used. If, after the ore was mined, men had changed their habits radically and stopped eating potatoes, the ore could have been put to other uses with little or no loss. But after the potato digger was made, the capital had become fixed in a specialized form. Decisions had been made that could no longer be changed. If men no longer wanted potatoes, there would be no easy and convenient way in which the potato digger could be used to weave cloth or load hay. *Specialized capital* is capital that has few uses, possibly only one use. *Unspecialized capital* is capital that can be turned to many different uses. Probably money is the least specialized of all capital in a society such as ours. The man who has money has not committed himself at all as to the use that will be made of it or the wants that will be satisfied.

13 • 5. Capital and Economic Forecasting. This roundabout method of production, which requires that men should take the first steps in production long before the ultimate satisfaction of wants, has the great advantage of increased productivity which we noted in Sec. 13 • 3; but has the disadvantage that it forces men to foresee their wants long in advance. We noted in Sec. 1 • 7 that human wants are not fixed but vary from time to time, unpredictably and unexpectedly. In a simple sort of society where men used no capital but expended their efforts always directly on getting goods to satisfy current wants, there would be no serious results from this inconstancy. But in a society that uses capital and the concomitant roundabout process, the results of changed wants may be serious and far-reaching. If we continue for the moment our recent example of the potato digger, we can see the point easily. Suppose that, after the potato digger is completed, men's wants change in such a way that they no longer want potatoes. This may seem unlikely, yet changes quite as radical and unexpected as this have happened historically. Assuming that this change in wants did take place, we see that the potato digger will be useless or almost so. The work that went into making it will have been wasted, as will all the long line of activities that preceded it back to the mining of the original ore. Men will have been working for months and perhaps for years, exerting their energies in good faith, and at the end it turns out to have been to no purpose. Had the society been one in which no capital was used, had men who wanted potatoes spent their time looking for natural sources of supply rather than in making capital goods, this waste effort would not have occurred. As soon as wants changed and men no longer wanted potatoes, they could have stopped hunting for them. The use of capital in any society always makes it necessary to foresee future wants, to forecast wants sometimes years ahead.

Without getting too far ahead of our story at this point, it might be worth while saying that some people would attribute the great depressions that follow one another as a part of the business cycle[1] at least in part to the use of capital. They would argue that, when any society uses capital, it is necessary to foretell human wants and that these wants can never be foretold with anything like complete accuracy. When mistakes are made, the result is that human effort and investment are drawn into lines that ultimately turn out to be unprofitable. Then there are losses, finished products cannot be sold, losses must be taken on investments in equipment, etc. We shall examine these problems later. Enough now for us to note that the roundabout process does require foresight, and that as long as human wants are changeable our foresight will never be entirely reliable.

13 · 6. **Capital and Saving.** In order that men may produce capital we have seen that it is necessary for them to sacrifice consumption that they otherwise could have had currently. We can state this in another way by saying that if men are to accumulate capital it is necessary for them to save. If they consume all they produce, there can, of course, be no capital. It is necessary for them to produce things over and above their current consumption if they are to have capital. To be sure, saving does not by itself produce capital; it is necessary to have both saving and productive use. The man who saves money and locks it up in his safe or wraps it in a napkin and hides it in the earth has produced no capital; but the man who saves money and invests it in productive equipment, or who puts it in a bank where it is lent to businessmen who need it for productive purposes—this man has made it possible to increase the stock of capital in his society. The man who consumes as much as he produces as he goes along never gives rise to any capital formation. Production must exceed consumption if a society is to increase its stock of capital and, since capital wears out, production must exceed consumption even to conserve the existing supply of capital.

Because capital is productive and because it comes only from saving, thrift and frugality have been looked on as virtues. Most societies have tried in various ways to stimulate saving. But it should be pointed out that saving may be either voluntary or involuntary. We may leave it to individuals to save or not as they wish, and try to make it profitable for them to save by paying them for the use of their capital. This is the way in which an individualistic society operates. But we may also see to it that men save whether they like or not. We can by various means force them to consume less than they produce. For example, a society can get its capital by taxing its members and building publicly owned capital. Or, as we saw in Chap 9, a corporation that has made profits need not declare dividends, but may decide to

[1] See Chap. 22.

retain the profits and reinvest them in the business. In such a case the directors vote to put the profits into the surplus account. The stockholders are not consulted in such a case. Their money is "saved" for them. Thus taxation and corporate surpluses are two common types of involuntary saving. The difference is a rather important one in social philosophy. If one saves voluntarily and accumulates capital that is productive, there is perhaps a valid claim to personal ownership of capital and personal enjoyment of the fruits of its productivity. Those who do not believe in a "capitalist" system[1] cannot well object if the Indian of Sec. 13·2, who has taken a day off to make a net, now uses this net to catch locusts; if he now catches more locusts than his neighbors and is therefore "richer" than they because of his position as an owner of capital, there is certainly some color of justice in allowing him to enjoy his larger income. To be sure, it is still possible to argue that he should be socially minded and should share his surplus with those who are less foresighted or less ingenious than he is. One cannot prove how the locusts "should" be divided. It is quite natural, however, that those who favor public ownership of capital goods should advocate the involuntary methods of saving, while those who believe in private ownership should emphasize more strongly the importance of individual thrift and foresight. They would urge that, if the Indian is not allowed to profit by his net, he will not in fact take the time off to make it, and that society will not get capital (or, at least, not so much capital) if saving is not made attractive to individuals by allowing them a return for their saving.

It has been argued, on the contrary, that men might save even more if they were not paid for it. The argument runs that the man who is saving for his old age, or saving to have something to leave his wife and children, would have to save even more if he were not paid interest. For some people this is doubtless true. But men surely would not make capital at all unless they expected it to be productive. No man in his right mind would spend his time, effort, and money in making a machine if he did not expect that the machine would yield a product larger than could be obtained without it. In our oversimplified case, the Indian would not spend the time making his net unless he thought that his catch of locusts would be increased thereby. There can be no dispute over the fact that capital is productive. The dispute is over the division of the product: Should it go to the man who saves and abstains in order that the capital may be produced, or should it be shared

[1] Although we have not defined this term nor used it heretofore, we should say that the difference between a capitalist society and, say, a communistic society, is not that one society uses capital and the other does not. Every society uses capital, and any intelligent man admits that capital is productive regardless of his social views. In a capitalist society, however, the capital is owned by individuals, while many people think it should be owned by the state. The dispute is one over ownership of capital, and not over the advantage of having capital.

in some other way? This is an ethical problem, and one to which the student may profitably give some thought. At least, it is possible that the answer would depend on whether the saving had been voluntary or involuntary.

13·7. Capital and Concentration of Economic Power. The issues that are lightly touched upon at the close of the preceding section are of considerably more than academic interest because of the fact that it is largely through the ownership of capital that the great fortunes have been made, that the great incomes are now earned, and that economic power is wielded. In Chap. 4 we saw something of the inequality of incomes and of the effects of this inequality. Income-tax figures indicate that the large incomes—$50,000 a year and over, for example—are almost entirely made up of income from property, and hardly at all from wages and salaries or other return for current effort. The phrase "current effort" is used advisedly, since it is evident that returns on capital may well reflect returns to past effort. If a man works hard and long and is frugal, he may build up, by sacrificing current wants, a competence for his old age. After he retires, he lives on the income from his past savings, and it would not be fair to assert that his income was not the result of his effort. But it would not be, as we said above, the result of his current effort. Of course, income from capital may not represent any effort at all on the part of the recipient. He may have been lucky on the stock market, he may have placed his bet on the right horse, he may have married the boss's daughter, or he may have inherited his capital from a rich uncle. The institution of inheritance often makes it possible for men to enjoy the fruits of the ingenuity or productivity of others. But private ownership of capital, coupled often with the right to inherit, is responsible for a large proportion of the income of most recipients of large incomes. Much of the attack on private capitalism which its opponents make is a result of their feeling that inequality has been perpetuated and magnified by private ownership of the means of production. They would also urge that private ownership of capital, especially in conjunction with the corporate form of business enterprise, has made possible the building up of concentrations of economic power, which are socially undesirable. These matters we shall take up in more detail later, in Chaps. 25 and 36.

It might be well before we leave the topic for the present, however, to point out that the problems involved in private ownership of capital are repeated with even greater emphasis when we consider private ownership of land—of natural resources. In the case of privately owned capital we can at least say that the capital was made by men. If they had not made it, it would not have existed at all. But land is a gift of nature. No one made it. Here it is much harder to justify private ownership and private exploitation for private advantage. Yet, even here, the student should be warned that the prob-

lem is not so simple as superficial consideration might make it appear. To be sure, the farmer did not make his farmland; but it is quite likely that he did pay for it, and it is quite possible that his ownership reflects his own productivity and sacrifice. The first owner may have had no moral or ethical claim to the land, but later owners can point to the fact that they paid for what they got. We get what the lawyers call the problem of "vested interests." A man might argue that, if we were to start all over again, he would not want private property in land, yet he might feel that it would be unjust to forfeit the property rights that have been built up under actual conditions. In addition there are many people who feel that private ownership of land and of capital turns out to be of ultimate social advantage by making men more productive than they otherwise would be. This point of view was illustrated by a quotation from the Bradford history of the Plymouth Colony in Sec. 3·4, and the student who wishes to reflect on the topic might profitably read that quotation again at this point. It has not been our intent to settle the issue here at all, but merely to point out again the great complication that is involved in most social problems. Those who are most certain of themselves in the quarrel over public versus private ownership of land and capital are often those who have considered the problems least. This same statement could be made with regard to most social problems. The thoughtful student will realize that on such problems there is always something of real merit to be said on both sides and, if there were not, there would be no problem.

13·8. Challenge of Capital Accumulation. The natural resources that men have to help them in production are approximately fixed, while historically the population to be supported from those resources has tended to increase. As we try to support more and more men on fixed resources, we tend to get reduced output per capita and to reduce people to poverty unless men find new methods of production or have better equipment with which to work. The alternative to increased poverty as population increases is the accumulation of more and better capital—more efficient equipment with which to produce.

Men with high incomes find it relatively easy to save, while men with low incomes find it necessary to use all or almost all of their meager resources for current consumption.[1] The same thing is true of entire societies. A society in which the people are poor finds it difficult to save, while a wealthy society finds it easy. But here we run into one of the basic economic dilemmas: the poor society that wants to improve its lot finds that in order to do so it must save. The rich society tends to improve its lot by saving, and the poor society finds it harder and harder to overcome its handicap. In this way the

[1] See p. 48.

inequality tends to become perpetuated and magnified. The more important it is to save, the harder it is to do it.[1] There is obviously a challenge here to any society to save and to improve its lot while it is still relatively easy—to improve while there is still time.

Questions

1. Which of the following would you classify as capital: a lawn mower, a violin, a toothbrush, a sewing machine, a typewriter, a pair of spectacles, a railroad ticket, a windmill, a wheel barrow, a bushel of seed potatoes?

2. In what way is abstinence or waiting or sacrifice necessary when a building contractor decides to use a steam shovel to dig the cellar hole for a new building? Is not the cellar hole dug sooner rather than later by this method, and would it not be fair to say that the use of the steam shovel actually reduced the amount of waiting necessary? Discuss.

3. If a man saves up money for his old age, is he not displaying exactly the opposite from what we have called time preference?

4. What things that you consume were produced without the aid of capital?

5. Define roundabout process in a simple sentence.

6. Give several examples of highly specialized capital, and several of capital that is relatively unspecialized.

7. When human wants change and men have already expended time, energy, and money toward the production of goods that are no longer wanted, who suffers a loss?

8. Would it be possible for a society or an individual to save too much? What would constitute oversaving? What would be the effects?

9. List the most cogent arguments you can think of in favor of and in opposition to private ownership of capital. How does society gain and how does it lose by private ownership? Do the same arguments hold for private ownership of land?

10. What is meant by vested interests? Do you have any vested interests? Are vested interests confined to capital and to land? Illustrate.

11. What is the effect of war on the accumulation of capital? Why?

[1] See Matt. 13:12.

14. VARIATIONS IN COST; SUPPLY

14·1. Profit: The Aim of Management. The student who thinks superficially is apt to accept the statement that every businessman wishes to get the greatest possible total income. He is also likely to agree unthinkingly that the businessman seeks to reduce his costs to the smallest possible amount. Yet a moment's thought will show him that neither of these supposed aims actually motivates businessmen. Any producer can reduce his costs, for example, by reducing the amount of his output. If he closes up shop entirely and goes out of business, his costs will fall and may even disappear entirely.[1]

Perhaps the easiest way to demonstrate the falsity of the ideas mentioned above is to take an example. Let us suppose that a businessman is trying to decide whether to produce 50, 1000, or 5000 units of output. In the first case his income, let us suppose, will be $100, while in the other two cases it will be $1000 and $2000, respectively. His costs in the three cases are $90, $500, and $1800. We can summarize the situations as follows:

	Situation A	Situation B	Situation C
Units of output	50	1000	5000
Total income	$100	$1000	$2000
Total cost	$90	$500	$1800

If we compare these three situations, assuming that the producer must choose one of them, we find that he does *not* choose the one of lowest cost (situation A) nor the one of greatest income (situation C). On the contrary, if we leave him free to do as he prefers, he will select situation B, since the student can quickly ascertain that the profit is higher under the conditions of that situation than under either of the other sets of circumstances. The aim of the

[1] On the other hand, he may have some fixed costs, associated with fixed factors of the kind described in Chap. 10, and these costs might continue even if the business closed down.

businessman is profit—maximum profit. He is quite willing to incur greater costs *if* he can thereby acquire an even greater addition to his income. He is quite willing to accept a reduction in his income *if* he can thereby achieve an even greater reduction in cost. His aim is *net income* rather than *gross income.*

14 · 2. Sources of Income and Cost. In our society the businessman ordinarily gets his income by selling goods or services. The amount of his income depends first on the quantity of goods that he sells, and second on the price at which he sells them. As we noted in Chap. 6, these two things are not independent. If the businessman tries to increase his income by raising the price of his product, he finds that consumers reduce the amount of their purchases. If the demand for his products is elastic, he will receive a smaller total income at the higher price; if the demand is inelastic, his total income will rise as he raises the price. In either case, however, as he raises the price there will be some reduction in the quantity that he can sell,[1] and this change in the rate of his output will affect his costs. We noted in Chap. 10 that the efficiency of production depends at least in part on the volume of production. We stated there the Law of Diminishing Productivity, which told us that as we vary the volume of output we change the size of the marginal and the average output.

Business costs are incurred in the main by buying or hiring factors of production. The businessman buys raw materials, hires labor, invests in machinery, or buys up natural sources of power. Whether his costs are high or low depends primarily on how much of the factors of production he uses, and on how efficiently he uses them. Thus both income and cost tend to vary according to the volume of output. By changing his volume of output, the businessman can change both his total cost and his total income. He tries to find the volume that will yield, not the largest possible income nor the largest possible cost, but the largest possible difference between them—the largest possible profit.

14 · 3. Physical and Economic Outputs and Inputs. In Chap. 10 we noted the variations in the amounts of output that accompanied variations in the amounts of the productive factors employed. At that time we used a table, now repeated as Table 14 · 1, which illustrated the tendencies. This table showed the physical quantities of the factors of production used and the physical quantities of product obtained. It shows us, for example, that if the producer decides to use 6 men with his fixed factors he will get a total product of 570 bushels of potatoes. Yet we have no way of knowing whether this will be a profitable thing for him to do or not. It is necessary for us to translate the data of Table 14 · 1 from these physical terms to economic terms—to terms of dollars and cents—to terms of cost and incomes. This we shall now proceed to do.

[1] This is in accordance with the Law of Demand. See Chap. 6.

14 • 4. Cost Curves. Let us assume that the farmer is paying $100 for his fixed factors. Since these factors are fixed this amount of $100 stays the same no matter what volume of output the farmer selects. Let us assume that the men who are employed are paid $50 apiece. Let us now go back to the data of Table 14 • 1 and convert them from physical to economic terms, translating them into costs and incomes. First, what happens to the cost of potatoes as we vary the output? If, for example, we hire 3 men, we have to pay $100 for the land and $50 each for the 3 men, or a total outlay of $250. In return we get a product of 270 bushels of potatoes (this figure being taken

Table 14 • 1. Illustration of Changes in Physical Output Associated with Changes in Physical Input

No. of men employed	No. of bushels of potatoes	Average output, bushels	Marginal output, bushels
0	0		
1	50	50	50
2	150	75	100
3	270	90	120
4	380	95	110
5	480	96	100
6	570	95	90
7	651	93	81
8	720	90	69
9	783	87	63
10	830	83	47
11	858	78	28
12	864	72	6
13	858	66	−6

from Table 14 • 1), so our cost per bushel is just under 93 cents. If we make similar computations for the other outputs, we get the data of Table 14 • 2. Each average cost in column 4 of Table 14 • 2 is found by dividing the total cost in column 3 by the total output in column 2. The figures in column 4 are rounded off to the nearest cent save for three cases near the lowest average cost, where they are carried one place further. We note that the average cost diminishes for a time at first, reaches a low point of barely over 69 cents when the output is 651 units, and then rises again. The average costs in every business exhibit this same peculiarity for reasons that are explainable as soon as we separate the fixed from the variable costs. The fixed and variable costs are analagous to the fixed and variable factors of production that were described in Sec. 14 • 2. Any cost that stays constant in total amount as

we vary the output is defined as a *fixed cost,* while any cost that changes in total amount when we change the volume of output is a *variable cost.* In our present problem, the $100 expenditure for land stays fixed, but the wages of the men vary in aggregate as we change the output. Thus, if we hire 3 men, our wage bill is $150; when we hire 4 men, it rises to $200. In our example the wages *per worker* have stayed constant, but the total expenditure

Table 14·2. Derivation of Cost Curves from Data of Table 14·1

No. of men employed	Total output, bushels	Total cost	Average cost	Marginal cost
1	50	$150	$3.00	
				$0.50
2	150	200	1.33	
				0.42
3	270	250	0.93	
				0.45
4	380	300	0 79	
				0.50
5	480	350	0 73	
				0.56
6	570	400	0 70	
				0.62
7	651	450	0.691	
				0.72
8	720	500	0.694	
				0.79
9	783	550	0.702	
				1.06
10	830	600	0.72	
				1.79
11	858	650	0.76	
				8.33
12	864	700	0.81	
13	858	750	0.88	

for wages varies as we vary the output; consequently we classify wages as a variable cost.

Let us notice, now, what happens to our fixed cost per unit of output as we vary our output. The total fixed cost stays constant at $100. If we produce but 1 unit of output, we have $100 of fixed cost per unit of output. If we produce 2 units, we have fixed costs of $50 per unit of output. As our output rises to 3, 4, and 5 units, the fixed cost per unit falls $33.33, $25, and $20, respectively. Since the total fixed cost is constant, the fixed cost per unit grows

smaller and smaller as we expand the output, and this will obviously be true in every business that has fixed costs, no matter what the character of the business. It is a matter of simple arithmetic that, if we keep the numerator constant and increase the denominator, the quotient will diminish. As a result, in every business the tendency is, as far as fixed costs have any influence, for average costs to get smaller and smaller. But sooner or later the Law of Diminishing Productivity starts to operate and, as we expand output, the average output begins to drop and the average variable cost per unit begins to rise. The tendency is, as far as these variable costs have any influence, for average costs to get bigger and bigger. Hence as we continue to expand output, after a while the rising influence of variable costs more than offsets the falling influence of fixed costs, and the average costs start to rise, as they do in the arithmetical example of Table 14 · 2. If we were to plot the data of Table 14 · 2, as in Fig. 14 · 1, we would see that the average cost curve tends to be U shaped, since average costs tend to fall at first, to pass through a minimum point, and later to rise again. We can see on purely a priori grounds that this is to be expected and, in addition, cost accounts of actual business firms always show the tendency in practice. Let us, then, lay down this general principle: *In any business enterprise the average cost curve is U shaped.*

Let us now turn back to Table 14 · 2 and notice the figures in the last column, headed Marginal Cost. We have had enough experience with marginal ideas (marginal utility, marginal output, etc.) to realize that, when the economist uses the term "marginal," he is always talking about a difference. We define the marginal cost of any volume of output as *the amount by which the total cost is increased when we produce one more unit of output.* Thus the marginal cost of 93 units of output is the difference between the total cost of 93 units and the total cost of 94 units. The marginal cost of x units is the difference between the total cost of x units and the total cost of $x + 1$ units.

When, however, we set out to compute the marginal costs in Table 14 · 2, we find that it is impossible to tell exactly what they are because, for example, while we can find in the table the total cost of producing 150 units, we cannot find the total cost of producing 151 units. We see that 150 units cost \$200 and that 270 units cost \$250, but the table does not tell the cost of 151 units. We can, however, make an approximation which will be satisfactory for our purposes. When we increase the total output from 150 units to 270 units, we increase the total cost from \$200 to \$250. Over this range of output as a whole, 120 extra units of output cost us 50 extra dollars. Therefore we can say that, on the average, one extra unit of output would cost us \$50 ÷ 120, or 41⅔ cents. This we round off to the nearest cent, calling it 42 cents, and enter it in Table 14 · 2 as the marginal cost, not for 150 or 270 units of output, but for the range that lies between these two. For this reason the figures for

marginal cost are placed between the other lines of Table 14 • 2, since they show the average marginal cost over the range of output between each pair of figures in the second column. The directions for finding the marginal cost corresponding to a pair of outputs and their total costs is this: Divide the difference in the total costs by the difference in the amounts of output. If we illustrate once more with one other point in Table 14 • 2, we note from the second and third columns of that table that, when the output is 783 units, the total cost is $550, but the total cost rises to $600 when the output increases to 830 units. The difference in cost is $50 and the difference in

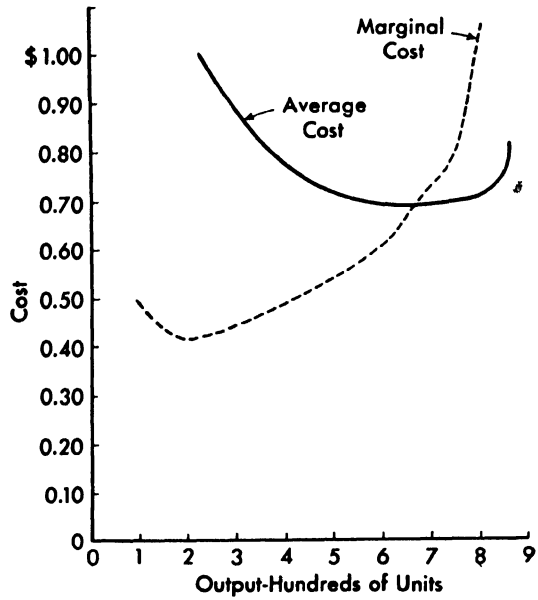


FIG. 14 • 1.—Cost curves based on the data of Table 14 • 2.

output is 47 units. The average marginal cost over the range of output from 783 to 830 units is therefore $50 ÷ 47, or $1.06, which appears in Table 14 • 2 as the marginal cost, but placed between the lines to indicate that it is the average marginal cost over the range from 783 to 830 units of output.

14 • 5. Relations of Average and Marginal Costs. If, now, we look again at Fig. 14 • 1, we find the data for both marginal and average cost plotted from Table 14 • 2. We note the typical U shape of the average cost curve, and we find the marginal cost curve rising through it and cutting the average cost curve at its lowest point. This, too, is a universal characteristic of the average and marginal cost curves of all industries, and we shall soon see the importance of it. The mathematical proof of the relationship between the

curves is not difficult[1], but, since we are here interested in the economic rather than in the mathematical aspects of the situation, let us examine its logic for a moment without the use of higher mathematics.

Suppose we take an arbitrary set of numbers, such as 3, 5, 6, 8, and 13. Their sum is 35 and their average is 7. If we add another number to the group, what will be the effect on the average? That obviously depends on how big a number we add. If we add one smaller than the average of the numbers already in the group, that is, if we add a number smaller than 7, we reduce the average. Thus, if we add the number 1 to the group, our total of the six numbers is 36 and we have reduced the average from 7 to 6. If we add a number larger than 7, such as 13, our total becomes 48 and our average rises from 7 to 8. If we add to the group a number exactly equal

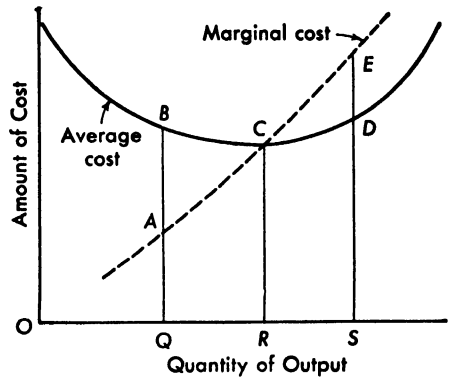


FIG. 14 • 2.—Relationships of marginal and average cost.

to the average, the average will remain unchanged. In our case if we add the number 7 to the group, the total becomes 42 and the average remains at 7. Note, then, the simple arithmetical rule: If we add to a group of numbers a new number larger than their average, the average rises; if the added number is smaller than the average, the average falls; and if the added number is equal to the average, the average remains unchanged. Now let us look at Fig. 14 • 2, which shows typical average and marginal cost curves. If we have the output OQ, the curves tell us that the marginal cost is QA and the average cost is QB. But the marginal cost is the added cost from an additional unit of output, so we see that we have added something smaller than the average, and we are pulling the average down. Sure

[1] See Note 4, Mathematical Appendix, p. 891. Since the method that we have used for computing the marginal cost is only approximate, the marginal cost curve so computed may not cross the average cost curve exactly at its lowest point, but if we were to use an exact method the relationship would be that described—a U-shaped average cost curve cut exactly at its lowest point by a marginal cost curve, which rises through it.

enough, at this point the average cost curve is falling. On the contrary, if we have the output *OS,* the marginal cost is *SE* and the average cost is *SD.* The marginal cost exceeds the average cost. This means that we are adding something larger than the average and pulling the average up. We see that at this output the average cost is rising. Finally, at output *OR* the marginal cost and the average cost both equal *RC.* We are adding an amount exactly equal to the average, and therefore the average is not being changed. Sure enough, at this point the average cost curve is neither rising nor falling but is momentarily horizontal. We can thus make the general rule that whenever the marginal cost curve lies below the average cost curve, the latter will be falling; whenever the marginal cost curve lies above the average cost curve the latter will be rising; and the two curves will cross at the lowest point on the average cost curve. Since this is true of every business enterprise, and must be true on the basis of simple mathematics if the Law of Diminishing Productivity operates, economists often fall into the habit of drawing type cost curves for purposes of illustration, just as they draw type demand curves. When they draw demand curves, they always draw them with negative slope, to illustrate the Law of Demand. Similarly, when they draw cost curves, instead of drawing them at random, they always draw a U-shaped average cost curve cut at its lowest point by a rising marginal cost curve. We shall often use such curves in our future work, and it is well to understand that they are based on the reasoning of the Law of Diminishing Productivity, corroborated empirically by the evidence of actual cost accounts of many business firms.

Finally, we can see the relationship between average and marginal costs by inspecting the data of the last two columns of Table 14·2. We note that toward the top of the table the average costs are falling and that the marginal costs in this part are lower than the average costs. Toward the bottom of the table the average costs are rising and the marginal costs are larger than the average costs. It is evident from further inspection that the average cost and the marginal cost are equal at some point where the output is approximately 650 units, although the exact point of equality is not shown in the table. At any rate, the two seem to be equal at about the place where the average cost reaches its lowest point, which is proper.

14·6. Influence of Selling Price. We have now discovered that, when certain cost facts are given, we can draw cost curves, but before we can tell how many men it will pay our farmer to employ (a problem that we started to discuss in Sec. 14·3 with Table 14·1) we need to know the selling price of the product in addition to the costs. Let us now assume, in addition to all the information which we had before, that potatoes sell for $1 per bushel, and with this added information let us find how many men

it will pay the farmer to employ and what volume of output will be most profitable for him to raise. We can first work out our answer by a long, roundabout way, which has the advantage of being transparently simple and understandable. We can actually work out the amount of profit that the farmer will make when he hires each number of men. For example, if he hires 4 men, his total cost (given in Table 14·2) will be $300. He will get 380 bushels of potatoes (this information is also contained in Table 14·2), and at $1 a bushel, which we have assumed to be the price, his total income will be $380. He will spend $300 and take in $380, so his total profit will be $80. Making similar computations for the other entries of Table 14·2, we get the data for Table 14·3.

Table 14·3. Point of Highest Profit from Table 14·2.

No. of men employed	Total output, bushels	Total cost	Total income	Net profit
1	50	$150	$ 50	−$100
2	150	200	150	−50
3	270	250	270	20
4	380	300	380	80
5	480	350	480	130
6	570	400	570	170
7	651	450	651	201
8	720	500	720	220
9	783	550	783	233
10	830	600	830	230
11	858	650	858	208
12	864	700	864	164
13	858	750	858	108

It is easy to see from Table 14·3 that this farmer will make the largest possible profit if he hires 9 men and produces 783 bushels of potatoes. His profit will then amount to $233, which is more than he can make at any other point. He can, of course, choose any output that he wishes and, if he decides to hire 6 men and produce 570 bushels of potatoes, he will still make a profit; but it will be a profit of $170 instead of the $233 which he could have made. The major problem of business management is that of trying to decide just what volume of business will be most profitable—just how much of the various factors of production to use in order to get the largest possible profit. A skillful entrepreneur judges these factors more accurately than his less skillful competitor and, as a result, makes more profit. It should be noted that two businessmen may pay exactly the same amounts for the things they

buy, and sell their products in competition with each other at exactly the same selling prices, yet one of them may make a large profit and the other may take a large loss. Thus if we had two businessmen, both paying $50 per laborer and both selling their output at $1 per unit as in the numerical illustration which we have just taken, if one of them hires but 2 men and the other hires 9 men, one will be making a profit while the other will be suffering a loss.

When we are given data such as those of Table 14·1, showing the amounts of output that one obtains from various expenditures, and when we are given in addition the facts about fixed and variable costs and about selling prices, we can always make up a table such as Table 14·3 and discover the point of highest profit. But we could have discovered the same answer directly from Table 14·1 by means of a little mental arithmetic, without going through the computations required to get Table 14·3. We could have reasoned this way: Every time we hire a man, we have to pay $50. Since the potatoes sell for $1 a bushel, it will pay us to hire any man who increases our output by more than 50 bushels, but it will not pay to hire anyone who increases output by less than that amount. So we look in the last column of Table 14·1 and ask ourselves, for example, whether we ought to hire an eighth man if we already have seven. We see that the eighth man will increase the output by 69 bushels, worth $69, and we shall have to pay him but $50. We would gladly get $69 by spending $50, so we hire this eighth man.[1] Similarly we hire the ninth man, since he increases the output by 63 units worth of $63, yet he costs us but $50. But we would be foolish to hire a tenth man, who brings in but $47 worth of added product, when we have to pay him $50. In other words, we can enunciate the general principle that *a businessman always finds it advantageous to hire additional units of a variable factor as long as the value of the marginal product exceeds the cost of the factor.* If we remember this principle, we can often compute in our heads the location of the point of highest profit. Let us illustrate the application of this principle once again to the data of Table 14·1. If wages remain at $50 per man but the price of the product rises to $2 per bushel, how many men will it pay to employ? Following our same line of reasoning, we say that, if we have to pay $50 for a man, we will have to increase our output by at least 25 units to be worth his wages if the output sells for $2 per unit. So we glance down the last column of Table 14·1 and see that it will now pay our farmer to hire 11 men and produce 858 bushels of potatoes at the

[1] If we get $69 by spending $50, we will increase our profit by $19. Inspection of Table 14·3 shows that, if we hire the eighth man, we increase our profit from $201 to $220, or by $19, which we could tell merely by inspection of the marginal outputs, without computing Table 14·3 at all.

new price. We could discover the same thing by computing a new table similar to Table 14 · 3, and if our arithmetic is correct we shall arrive at the same conclusion.

The most important principle of all can be illustrated approximately from the figures of Table 14 · 2 or from Fig. 14 · 1. This principle is that, *if the selling price of the product is fixed, a businessman always finds it profitable to expand his output until the marginal cost equals the selling price.*[1] In our case the selling price is fixed at $1 per bushel. The farmer will make the largest possible profit if he expands his output until the marginal cost is $1 per bushel. Inspection of Fig. 14 · 1 indicates that the marginal cost is $1 per bushel when the output is approximately 800 units. Similarly in Table 14 · 2, we see that the marginal cost is $1 per bushel somewhere approximately where 9 men are hired. Since our method of computing marginal cost is but an approximate one, we can get but an approximation of the answer in this case; if we could compute marginal costs accurately, we would find that it paid the farmer to hire the ninth man, but not to hire the tenth man. The reason for this is not far to seek. The marginal cost, we recall, is the amount by which our total cost is increased if we produce one more unit of output. If this marginal cost is less than the selling price, we make a gain by producing an added unit and selling it. It pays us to expand the business. As long as added units can be produced at a cost below the selling price (that is, as long as the marginal cost is below the selling price), it pays us to expand. To be sure, if the marginal cost is almost as great as the selling price, we make only a little additional profit by producing another unit of output—but we do make a little more profit, and any additional profit is worth while. When the marginal cost is above the selling price, added units are costing us more than we can get for them. They are reducing our profit, and we do not produce them. We contract our output.

14 · 7. Graphic Determination of Output. We have already seen that the average cost curve and the marginal cost curve in any business have particular shapes and relationships.[2] Let us assume that Fig. 14 · 3 shows the cost curves for a particular business enterprise. We note that the curves have the shapes and relationships previously described. If the price of the product is 80 cents per unit, how many units will it pay this businessman to produce, and how much profit will he make? The answer can be discovered from the diagram if we use the principles that we have just worked out. First, we remember that the businessman will make his greatest possible profit if he expands his output until the marginal cost equals the selling

[1] For proof of this principle see Note 5 of the Mathematical Appendix on p. 891.

[2] See Secs. 14 · 4 and 14 · 5.

price, if, as in this case, the selling price is fixed.[1] In our present problem, the price is fixed at 80 cents per unit, so it will pay to expand output until the marginal cost is also 80 cents a unit. We see from the diagram that this will be the case if the output is set at about 870 units, so we have our first answer. The businessman in this case will maximize his profit by producing 870 units. But how much profit will he make? Very evidently, if he produces 870 units at 80 cents per unit, his total income will be 870 × 80 cents, or $696. But the average cost curve tells us the average cost corresponding to any particular output. From Fig. 14 • 3 we see that, if the output is set at 870 units, the cost per unit will be approximately 72 cents. The total cost,

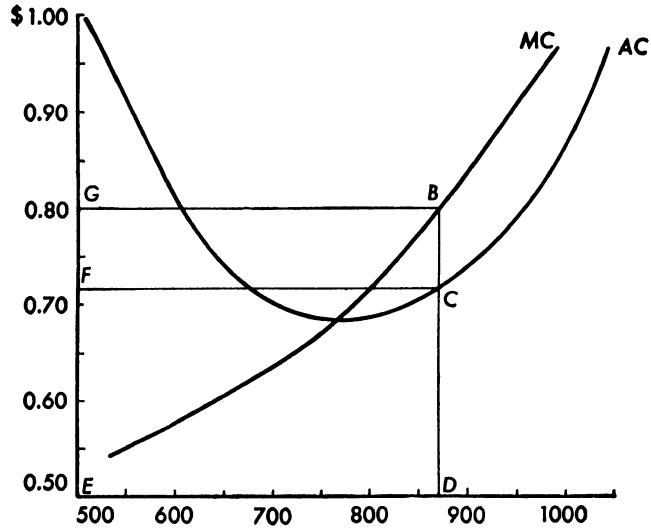


FIG. 14 • 3.—Determination of most profitable output, when price is fixed at 80 cents per unit.

then, will be 870 × 72 cents, or $626.40, which will yield a net income of $69.60. Since we have already discovered that we get the greatest possible profit by selecting the output where the marginal cost equals the selling price, we need look no further. We can see, by the way, these facts on the diagram. If we select the output *ED,* the marginal cost will be *DB,* or 80 cents. The total income will be *ED* units at *DB* cents per unit, or will be represented by the rectangle *EDBG.* But the output *ED* will have an

[1] If the business in question were a monopoly or had some elements of monopoly power, it would be true that the price would fall as output was expanded; but in a competitive case the businessman is too small a part of the market to affect the price, and he can expand output without having the price drop. We shall discuss this matter later in Sec. 16 • 10. In our present example we assume that the price remains at 80 cents per unit of output as the businessman expands and contracts his business. It is only in such cases that the principle holds that highest profit is made by equating marginal cost and selling price.

average cost of *DC*, so the total cost will be *ED* units at *DC* cents per unit, or the rectangle *CDEF*. The larger rectangle will be the total income, and the lower rectangle will be the total cost. The difference between them, or rectangle *BCFG*, will represent the amount of the profit if output is set at *ED* units. This rectangle *BCFG* will be larger in area than any other similar profit rectangle that can be drawn on this diagram.

Table 14·4. Illustration of Changes in Output Accompanying Changes of Input (Same figures as Table 14·1)

No. of men employed	No. of bushels of potatoes	Average output, bushels	Marginal output, bushels
0	0		
1	50	50	50
2	150	75	100
3	270	90	120
4	380	95	110
5	480	96	100
6	570	95	90
7	651	93	81
8	720	90	69
9	783	87	63
10	830	83	47
11	858	78	28
12	864	72	6
13	858	66	−6

14·8. Effect of Changing Costs. We have now discovered that it is possible to find the most profitable output if we are given three things:

1. The technological facts about the industry that determine how much output the entrepreneur can get from various inputs: such information as that of Table 14·1.
2. The fixed and the variable costs of production.
3. The selling price of the product.

All these things are ordinarily beyond the control of the producer. He has to adjust his business to them, selecting the output that will yield the largest possible profit under the given technological conditions, the given costs, and the given price of the product. Let us now see what will happen as these given factors change. For purposes of illustration we shall use again the data of Table 14·1, which we now repeat as Table 14·4 for ease of reference.

When we worked with this table before, we assumed that the farmer had $100 of fixed expense, that the wage rate (which was the variable expense

in this problem) was $50 per man, and that the product sold at $1 per bushel. Under those circumstances we found that the largest possible profit would be made by hiring 9 men and producing 783 bushels of potatoes. The profit made in this way amounted to $233, as we discovered in Table 14·3. Let us now alter the amount of the variable cost, leaving the fixed cost and the selling price unchanged, and see what happens to the point of highest profit. Suppose, first, that the wage rate rises to $75 per man. Following our earlier reasoning, we realize that we shall hire men until the value of the marginal product equals the cost of the factor, which in this case is $75. Since the product sells at $1 per unit, we add men as long as the marginal product exceeds 75 bushels; in this case, we shall find it most profitable to hire 7 men. If the student will follow the same line of reasoning, he will see for himself that we will hire 10 men and produce 830 units if the wage rate is $30 a man, or we will hire 11 men and produce 858 units if the wage rate is $25. It

Table 14·5. Output, Employment, and Profits at Various Wage Rates

Wage rate, dollars	Output, bushels	No. of men employed	Amount of profit
$25	858	11	$483
30	830	10	430
50	783	9	233
75	651	7	26

will be worth the student's while to carry out the necessary mental arithmetic to check the statements in the preceding sentence.

If we gather together the results of the changes just made, we find the first three columns of Table 14·5. The last column is computed for each case separately. For example, if we set the wage rate at $25 per man and hire 11 men, our wage bill will be $275. We shall have in addition $100 of fixed expense, so that our total cost will be $375. We shall sell 858 bushels of product at $1 per bushel, so our income will be $858, and our profit will be the difference between $858 and $375, or $483, as shown in the last column of the table.

Table 14·5 is based, of course, on Table 14·4, which illustrates the Law of Diminishing Productivity. From Table 14·5 we can draw a corollary to the Law of Diminishing Productivity, which runs as follows: *If we increase the variable expense in any business, leaving the other factors constant, it will become profitable for the businessman to reduce his output and to reduce the amount of the variable factor that he uses, and the total amount of his profit will be reduced.* To put this general theorem in terms of our specific example, we can say that if the wage rate rises it will pay the farmer to lay

off some of his hired hands and produce fewer potatoes, and that the total amount of his profit will be reduced. The businessman does not discharge workers and reduce output because he likes to. He would have preferred to go on as before. But under the new cost situation, though he makes less profit than he did before, he still makes the greatest profit that is now possible.

Now let us turn to the effect of changing the fixed costs. Taking the data of Table 14·4 as given and assuming that the variable cost is $50 per man and that the product sells for $1 per unit, we shall try various fixed costs to see what happens. First, we have already worked out our results when the fixed cost was $100. We then found that it was most profitable to hire 9 men and to produce 783 bushels of potatoes. The total profit was then $233. Suppose the other factors stay the same, but the fixed cost rises to $200. If we follow our general principle, we realize that since each man costs

Table 14·6. Most Profitable Output and Employment and Amount of Profit as Fixed Cost Changes

Amount of fixed cost	Output, bushels	No. of men employed	Amount of profit
$200	783	9	$133
100	783	9	233
50	783	9	283
0	783	9	333

us $50 and the product sells for $1 per unit, it will pay us to hire men if their marginal product is 50 bushels or more, so we shall hire 9 men and produce 783 units of output just as before. As we have $100 more expense to pay, our profit is reduced to $133. Similarly, if the fixed cost is cut to $50, we shall find it most profitable to hire the same 9 men and to keep our output at 783 units, but our total profit will be $283. If the fixed cost is cut out altogether, we shall still hire 9 men and produce 783 units, but our profit will rise to $333. The student should check these results by his own computation.

Assembling these results, we find the most profitable outputs and employment and the amounts of profit at each rate of fixed cost to be those given in Table 14·6. Again it is obvious that we can infer a corollary to the Law of Diminishing Productivity, as follows: *Changes in the fixed cost have no influence on the volume of output or on the amount of variable factors employed, but any change in the fixed cost changes the profit an equal amount in the opposite direction.*

It is easy to see why these corollaries work out as they do, if we remember the Law of Diminishing Productivity, which underlies them. This law we can see illustrated in the last column of Table 14·4, where we see, after a certain point is reached, that the size of the marginal product gets smaller and smaller. We have discovered that it pays to expand one's business when the value of this marginal product exceeds the cost of the factor. If we raise the cost of the factor (that is, if we increase the variable expense), we cannot come so far down in the column before the value of the marginal output equals the cost of the factor. We reach the point of highest profit sooner and hence reduce our output and our use of the factor. But if we raise the fixed cost, leaving the cost of the variable factor unchanged, we affect neither the value of the marginal output nor the cost of the variable factor. Since both are unchanged, the point where they are equal is unchanged, and the entrepreneur does not alter his volume of output or employment when fixed costs change.

Table 14·7. Most Profitable Output and Employment and Amount of Profit as the Selling Price Changes

Selling price per unit	Output, bushels	No. of men employed	Amount of profit
$25	864	12	$20,900
2	858	11	1,066
1	783	9	233
0.70	651	7	5.70

The student will find it helpful to work out two or three tables similar to Table 14·3, using various assumptions as to fixed and variable costs and the productivity figures of Table 14·4. He will soon find that such tables as these will also indicate the validity of the corollaries we have just stated.

14·9. Effect of Changing the Selling Price. In the preceding section we altered the variable cost or the fixed cost to see what happens to the point of highest profit. It now remains for us to alter the selling price and learn what effect such changes have. Again we take the figures of Table 14·4 and assume that the variable expense is $50 per man and that the fixed expense is $100. We long since discovered that if the selling price is $1 it will be most profitable to hire 9 men and to produce 783 units of output, making a total profit of $233. Suppose, other things staying the same, the selling price now rises to $2 per bushel. If men cost $50, they can now pay for themselves by producing 25 bushels or more; so we hire 11 men and produce 858 units of output. These 11 men cost us $50 each, or a total of $550, and the added $100

of fixed cost gives us a total expense of $650. We sell our 858 units of output at $2 each and get a total income of $1716. This leaves us a profit of $1066. Similar reasoning shows us that, if the selling price rises to $25 per unit, we shall find it most profitable to hire 12 men and produce 864 units of product, while our total profit will then be $20,900. If the selling price falls to 70 cents per bushel, it will be most profitable to hire 7 men and produce 651 bushels of output, and the total profit will be but $5.70. These results, which should be checked by the student until he is sure that he knows whence they came, are summarized in Table 14 • 7. From this table we can easily derive our third and last corollary of the Law of Diminishing Productivity, which is as follows: *Raising the selling price of a commodity, other things equal, makes it profitable to increase the volume of output and to use more of the variable factors, and the total amount of profit is increased.* In this

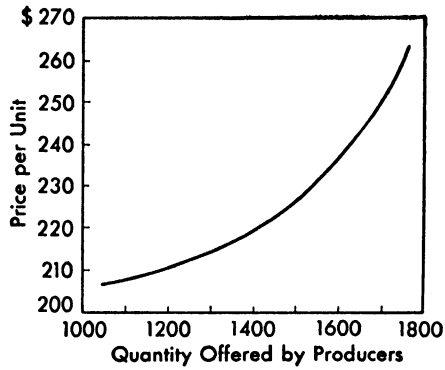


FIG. 14 • 4.—A supply curve.

specific example we can paraphrase the corollary by saying, if the price of potatoes rises, it pays to grow more potatoes and to hire more men, and the total profit will be larger.

14 • 10. Supply. When we were considering consumer problems in Chap. 6 we discovered that, as a result of diminishing utility, consumers found it to their advantage to reduce their purchases of a commodity when the price of that commodity rose. In the preceding section we discovered that producers find it to their advantage to increase their output of a commodity when its price rises. The relationship between the price and the quantity that consumers demanded we called the "demand" for the commodity. This new relationship between the price and the quantity that producers offer for sale we call the "supply" of the commodity. Just as demand was not a single quantity, but a table or schedule of quantities at various prices, so supply is a table or schedule of quantities at various prices. Just as the demand can be shown by a demand curve, so the supply can be shown by a supply

curve. Thus Table 14·8 shows what a supply schedule is like, and Fig. 14·4 shows what a supply curve is like. The supply schedule may show for a single producer what quantities will be offered at various prices. In that case we call it an "individual supply schedule." We can add together the offerings of all producers in a given market, and present a "market supply schedule" showing quantities offered at various prices on the market as a whole. Similarly we can have individual and market supply curves. It is noticeable at once that the supply concept is in almost every respect parallel to the demand concept with which we have already become familiar. In addition, we urged in Chap. 6 that a demand schedule or a demand curve showed what was true at any given instant. Changes in taste or income might change the demand. Similarly, the supply schedule and the supply curve show what is true at a given time. Changes in the techniques of production will change our original productivity table (Table 14·4) and therefore alter

Table 14·8. Hypothetical Supply Schedule

Price per unit	Quantity offered for sale	Price per unit	Quantity offered for sale
$1	150	$6	420
2	185	7	560
3	251	8	740
4	294	9	990
5	350		

our results. Similarly, changes in the variable costs will influence our results. As long as the variable costs and the techniques of production remain unchanged, the supply will remain unchanged, but changes in either of these will alter the supply. Obviously changes in the price of the product will not change the supply, since the supply already shows the quantities offered at all different prices.

Although in every one of these ways the ideas of supply are parallel to the ideas of demand, we must note one point of marked difference. The Law of Demand told us that increases in price reduced the quantities that consumers were willing to take. Price and quantity moved in opposite directions. But inspection of Table 14·8 or 14·7 will show that the supply situation runs in the opposite direction. The Law of Supply states that *the quantity of any commodity offered for sale by producers varies directly with the selling price, other things equal.* In other words, raising the selling price raises the quantity offered, while lowering the selling price lowers the quantity offered, if we keep other things (costs and techniques) unchanged. The sup-

ply curve of Figure 14·4 consequently has a positive slope instead of the negative slope characteristic of demand curves.

14·11. Elasticity of Supply. When we were considering problems of demand, we found that the idea of elasticity of demand was of some importance. Since the idea of supply is so much like the idea of demand, it is but natural to raise the question of elasticity of supply. Obviously it is possible to set up a concept in every way parallel to that of elasticity of demand. In this latter case we noted that quantities that consumers stand ready to buy tend to diminish as prices rise, but sometimes the adjustment was great and sometimes small. An elastic demand was one characterized by large consumer adjustments to price change, and an inelastic demand was one characterized by only minor consumer adjustments to price change. Elasticity of supply is defined in exactly parallel terms, except that the quantities used are those offered by producers at various prices instead of the quantities that consumers stand ready to take at various prices. We use the formula of Sec. 7·2, namely,

$$\epsilon = \left(\frac{\text{diff. in } Q}{\text{sum of } Q}\right)\left(\frac{\text{sum of } P}{\text{diff. in } P}\right)$$

The expressions in this definition are interpreted as follows:

"Diff. in Q" means the difference between the two quantities that producers offer in the problem.

"Sum of Q" means the sum of the same two quantities.

"Sum of P" means the sum of the two prices in the problem.

"Diff. in P" means the difference between these prices.

ϵ is the Greek letter epsilon, here used to represent the elasticity of supply. To illustrate the use of the formula, let us suppose that, when the price of tomatoes is \$1.50 per basket, farmers in a given market offer 500 baskets for sale. When the price falls to \$1.25 per basket, they offer 495 baskets for sale. What is the elasticity of supply? The two prices are \$1.50 and \$1.25. Their sum is \$2.75 and their difference is \$0.25. The two quantities are 500 and 495. Their sum is 995 and their difference is 5. Substituting these sums and differences in our formula we get

$$\epsilon = \left(\frac{5}{995}\right)\left(\frac{275}{25}\right) = 0.055$$

As in the case of elasticity of demand,[1] we say that the supply is elastic if ϵ is greater than 1, that it is inelastic if ϵ is smaller than 1; and that there is unit elasticity if ϵ is equal to 1. In the numerical example just given, the supply is very inelastic. This means, of course, that producers offer almost

[1] See Note 6, Mathematical Appendix, p. 892.

the same quantity for sale regardless of the price, making only minor adjustments to price changes.

It seems usually to be true that the elasticity of the supply of a commodity is smaller than the elasticity of the demand for it. Demand tends to be somewhat more elastic than supply. This is true because we have defined supply in an instantaneous sense, as the various quantities that would be offered for sale at various prices *right now*. At any given moment, the quantity available for market is approximately fixed. If for some reason consumers suddenly offered $25 a bushel for apples, farmers could not offer many more apples for sale at the time. They could, to be sure, offer a few that they had intended to eat themselves, or a few that would otherwise have gone to the cider mill, but at the moment they would be unable to increase their offerings very much; so the supply is inelastic. This is especially true, of course, with perishable commodities, since the entire quantity of them has to be sold almost regardless of the price. Businessmen are often interested, however, in another similar idea—not in the changes in quantity offered right now when the price changes, but the changes that will ultimately be made after producers have time to adjust their businesses to the new price. Thus we might ask, if the price of apples rose to $25 a bushel and stayed there, how many apples would be offered after farmers had time to plant new orchards and bring them into bearing—after they had had time to make whatever adjustments they wished to the changed price. In such cases it is certain that the $25 price would stimulate a tremendous increase in output. This relationship between the price and the quantity that would ultimately be offered after producers had adjusted to the price, we call "long-run supply," and long-run supplies of most commodities are very elastic. It is the short-run immediate supply that is inelastic, and it is inelastic because there is little that producers can do to alter their output immediately.

As a matter of fact, this is but another way of stating the principle of fixed costs illustrated earlier. When we consider the quantity of any commodity that is at any moment available for the market, we realize that most of the costs of producing it have already been incurred and cannot now be altered. They have become fixed costs, and fixed costs play no part in determining the quantity that will be offered for sale. Since practically all the costs are fixed in the short run and since fixed costs do not influence the quantity offered for sale, it is natural that there will be but minor variations in the quantity offered for sale when the price changes, and that the short-run supply will be relatively inelastic. If the price of apples rises now to $25 or falls to 25 cents per bushel, the costs of planting the orchard, of cultivating, spraying, thinning, picking, etc., have all been incurred and can no longer be changed whether we sell 1 bushel or 100. These costs will no longer change if we change our volume of business today. They have

become fixed costs. In the long run they are variable costs, since we can decide for another year whether to incur them or not, and we will not incur them unless we think that we can sell for enough to recover them. Costs that have been incurred and cannot now be changed are often called "sunk costs," and the fact that the businessman does not consider them in deciding how much to sell is merely another example of the rule that "there's no use crying over spilt milk."

14·12. Diminishing Returns. In Sec. 14·3 we stated the Law of Diminishing Productivity, which tells us, in effect, that if we hold constant any factor X and increase the quantity of another factor Y, which is associated with it, sooner or later the marginal, average, and total returns will drop. This law was originally discovered, however, in the case where the fixed factor was land, and socially this is still by far the most important case. As far as the individual entrepreneur is concerned, any factor may be the fixed factor. As we have said, the manager's own managerial ability is often the fixed element, and he varies the amounts of the other factors that he associates with himself. But as far as society is concerned, land is the factor that is relatively fixed, as was noted in Sec. 11·2. When we state the Law of Diminishing Productivity for the particular case where land is the fixed factor, we call the statement the "Law of Diminishing Returns," which is stated as follows: *At any given state of the arts, if we add more and more of the other factors of production to a fixed amount of land, we reach points sooner or later after which the marginal, the average, and the total outputs diminish.*

It will be noted that the Law of Diminishing Returns is a special case of the Law of Diminishing Productivity, the case where the fixed factor is land. But, since natural resources are approximately fixed in quantity and since population and the quantity of capital do ordinarily tend to grow larger with the passage of time, it is natural that we should pay particular attention to the case of diminishing returns. We have the conditions under which we would expect, *unless there are changes in the state of the arts,* higher and higher returns to land and lower and lower returns to labor. If population tends to expand, we have to expect lower and lower wages and higher and higher returns to land unless we are ingenious enough to develop new methods which make it possible to get larger outputs than before from given combinations of the productive factors. We have, thus, a social problem of combining the factors of production which is similar to the problem faced by the individual businessman, except that, while the businessman can judge the correctness of his decision by noting their effect on profit, the statesman has no such simple criterion. In order to know whether our decisions on immigration, birth rate, death rate, accumulation of capital, etc., are wise, we have to agree on our social goal—on the end

of life, and what is the good life. Again we find that the problems of ethics begin where those of economics leave off.

14 · 13. Summary. Businessmen usually try to organize their businesses in such a way as to maximize their profit. Their goal is not greatest possible total income, but greatest possible net income. In trying to get the maximum net income, the businessman usually can make variations in his output, but he usually finds that most other things are beyond his control. Particularly, he is apt to find that he cannot vary at will his costs nor the selling price of his product, nor is he is at liberty to alter the amounts of output that can be obtained by using the known methods of production. With these things given, he varies his outputs by using more or less of such factors of production as he can command; but, as he alters the combinations of these factors, he runs up against the fact that he can get added output from any fixed factor only by making his variable factors less efficient. As a result of that, there are certain relationships between his average costs which are common to all business enterprises and which can be depicted by cost curves, showing the changes in average and in marginal cost as the output is varied. As a result of the cost relationships, businessmen find it profitable to hire additional units of variable factors as long as the value of the marginal output exceeds the cost of the factor. In cases where the selling price is fixed, this is equivalent to saying that men find it profitable to expand output until marginal cost equals selling price. If variable costs are increased, other things equal, profit shrinks and it pays to reduce output. If fixed costs increase, profit also shrinks, but it does not pay to alter the amount of the output. If the selling price rises, profits rise and it becomes profitable to expand output. As a result of this last fact, rising prices ordinarily stimulate larger offerings of goods by their owners—the Law of Supply. Supply is in most ways the counterpart of demand, being represented by supply schedules or supply curves. Elasticity of supply, measured like elasticity of demand, is usually very low in the short run and for perishable goods, but somewhat greater for durable goods and usually quite large in the long run. Socially, the problems of production are primarily those of adapting the quantities of the other factors of production to fixed quantities of land; and the general tendency for increased population tends to bring lower marginal productivity to men, and lower levels of living, unless offset by new knowledge—by changes in the state of the arts.

Questions

1. "If a businessman makes a profit of $1000 when he produces 100 units of output, he will make $10,000 of profit if he produces 1000 units of output" This statement is not necessarily true. Why not?

2. Could the number of students in your institution be increased without increasing the number of professors? Would you list the teachers as fixed or as variable factors?

3. In some businesses unskilled labor is a fixed factor and in some a variable factor. In a particular case, how would you decide which it was?

4. How would our productive system be different if the Law of Diminishing Productivity did not work? Assume, for example, that doubling input will always double output, so that average and marginal outputs are constant.

5. Assume that the accompanying table shows the average outputs of a given enterprise when various numbers of men are employed.

No. of men employed.....	1	2	3	4	5	6	7
Average output per man.....	7	9	10	9	8	7	6

a. What is the marginal output when 6 men are employed?

b. If men cost $25 each and if there is fixed cost of $100, what is the total cost of 40 units of output?

c. Under the given conditions of part *b*, what is the average cost of 40 units of output?

d. Under the conditions of par *b*, what is the average marginal cost over the range from 36 to 40 units of output?

6. Draw the cost curves for the table in question 5, making the cost assumptions as in part *b*.

7. Suppose that we have the data of question 5, including the costs of part *b* thereof. Suppose that the product sells for $10 per unit.

a. How many units should the man produce if he is to maximize his profit?

b. How much profit would he make under these circumstances?

8. If Fig. 14 • 5 shows the marginal cost curve and the average cost curve for a given business,

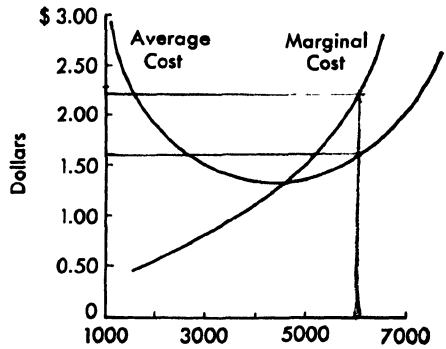


FIG. 14 • 5 —Illustrating question 8

a. What volume of output will it be most profitable to produce if the selling price is fixed at $2.25 per unit?

b. How much profit will be made under these circumstances?

c. Suppose the businessman, instead of producing the quantity that you compute above, does actually produce 7000 units of output. How much profit will he make?

9. Using Fig. 14 • 5, make up a supply schedule showing the numbers of units of output that this firm would offer for sale at each of the following prices: $1.50, $1.75, $2, $2.25, and $2.50.

10. Basing your answer on your work of Question 9, what is the relationship between the firm's supply curve and its marginal cost curve?

11. Suppose that the firm whose cost curves are depicted in Fig. 14 • 5 finds that the selling price of the product is $1 per unit

a. How many units of output should it produce if it produces at all?

b. Do you think that it is wise for the firm to continue production for the time being when the price is $1? Do business firms always go out of business the moment they lose any money?

12. The curves of Fig. 14 • 5 have the correct general shapes for cost curves, but they have not actually been computed from one another. Transfer the average cost curve from Fig. 14 • 5 to cross-section paper, and then compute the actual marginal cost curve that corresponds to it. *Hint:* We see from question 5 that, if the average costs are given, the marginal costs can be computed.

13. In Sec. 14 • 3 the statement is made that "almost any farmer in the United States could get higher yields per acre if he wanted to." If this is true, why do not farmers want to get the higher yields that are possible?

14. When the price of eggs is 50 cents a dozen, farmers offer 300 dozen eggs for sale in a certain city. When the price drops to 40 cents a dozen, they reduce their offerings to 290 dozen. What is the elasticity of supply?

15. Is it correct to say that changes in the fixed cost have no effect on a business? Why or why not?

16. Why is the supply likely to be less elastic in the case of perishable than in the case of nonperishable commodities?

PART FOUR. Exchange

15. EXCHANGE; COMPETITIVE MARKET PRICE

15·1. Nature of Exchange. In a society characterized by private property and by specialization in production, it is obvious that men will have to exchange in order that their wants may be satisfied. In any society that recognizes property rights, there must be rules for determining to whom goods belong and for transferring title to goods. Men may, of course, come into possession of goods by force or fraud, but the transfers of ownership that are a part of what we call "exchange" are transfers made voluntarily. In fact, we can define an exchange as *two voluntary transfers of ownership, each made in consideration of the other.*

Although it is evident that exchange must take place in any society which recognizes property rights and in which production is specialized, it is also evident that, in a society where either private property or specialization is lacking, the whole situation will be changed. It may seem at first sight as though in such cases there could be no exchange—as though exchange was entirely dependent on these two phenomena. Yet, although exchange would surely be very different in nature in societies without private property or division of labor, nevertheless all the problems of exchange would still exist. It is important for the student to understand why this is true if he is to sense the basic importance of problems of exchange in economics.

Let us imagine a Robinson Crusoe on his island before the coming of Man Friday. He cannot specialize in production, since he is entirely dependent on himself for everything. And it appears that he cannot exchange goods, since there is no one with whom to carry on exchange. But Crusoe's time and energy and equipment are limited. The things that he can produce are limited in number, and he has to plan the use of his time and his resources carefully if he is to get along at all. He could, of course, spend his afternoons at target practice on the beach, but if he does he will have used up ammunition that is urgently needed for other purposes and he will have squandered time that might have been given to more productive pursuits. So let us suppose that Crusoe uses his afternoon for hunting and that he

shoots a goat. In a very real sense he has given up leisure and recreation for the goat. He has made an exchange. Or if, instead of hunting for the goat, he had spent his time gathering oysters, he would have had to give up the goat in order to get the oysters. The exchange would be just as real as though he had owned a goat and traded it to someone else for oysters. In either case he could have had a goat, but instead did have oysters.

Exchange is a necessity that comes with limitation of resources—with scarcity. When men cannot satisfy all their wants but have to choose among alternatives, we have exchange. Choice between desirable things is fundamental to all economics, and choice is the basis of exchange. In a world where men could have everything they wanted—where they did not have to choose but could satisfy every want completely—in such a world there would be no exchange. But, whenever men find themselves faced with a situation in which some wants have to go unsatisfied, they have to choose which to satisfy and which not. They have to choose to give up some in order that they may have others. In such cases they have in effect exchanged some for others. Although the student surely has not been in the habit of thinking of such a situation as exchange, he should realize that all the problems of exchange are implicit as soon as men make choices. While we shall spend our time studying exchange as it does take place in a society of private property and specialization, the student should not be so superficial in his appraisal of the situation as to fall into the error of supposing that exchange and its problems go no deeper.

15·2. Barter. In an organized society, where men are actually exchanging goods with each other, the exchange may take many different forms. Perhaps simplest from the standpoint of understanding, and hardest from the standpoint of actual operation, is the direct exchange of goods for goods, which we call "barter." A man who has more eggs than he and his family can use swaps his surplus for another man's surplus of soap. Only the two men and the two products are involved. No money is used in the transaction; no banking system is necessary to facilitate it. Yet simple as barter is to describe, it is difficult to carry on—so difficult that commerce never becomes important in any society that has progressed only to the barter stage.

There are two types of difficulties with barter. In the first place, the commodities exchanged may not be at all of equal value. The chariot builder who has a fine new four-horse chariot to dispose of may find that his egg man wants the chariot, but if he is to get fair value he will have to take far more eggs for it that he can conveniently use. It will not help to try to exchange the chariot one small piece at a time, because the chariot will have no value piecemeal. Chariots and eggs are incommensurable, and if they are to be exchanged directly, the one for the other, at least one of the parties to the exchange will be put to a serious disadvantage.

The other difficulty with barter is that it is so hard to carry it out at all. In order to exchange my own surplus for something else that I want, I not only have to find someone who wants what I have to give up, but he must also have what I need. Any modern businessman knows that it is hard enough to find customers for goods when the only problem is to find someone who wants what you have for sale; but when you add to this the requirement that the other party must also have something that you want, the situation becomes a very difficult one indeed. The economist describes this by saying that barter necessitates a "double coincidence of wants and supplies." Not only must the other man want what you have (this is coincidence number one) but he must also have what you want (this is coincidence number two). Imagine, let us say, the situation that confronts a breeder of pouter pigeons who wants a dish of fried eels for supper. Not only must he find someone with a surplus of eels, but the other man must also want pouter pigeons! To be sure, most cases of exchange would not involve articles quite so unusual as these, yet every barter exchange does involve the difficulty of establishing the double coincidence of wants and supplies, as well as the difficulty of getting commensurable and satisfactory units for exchange after the double coincidence is established.

15·3. Purchase. Men faced with the difficulty of carrying out exchange under a barter system long since realized that it was often worth while to exchange their surpluses for things that they did not really want if they could not find someone at once who had what they did want and wanted what they had. The man who has pouter pigeons and wants eels may not at once find a man who has eels and wants pouter pigeons, but he may find someone who has sugar, wheat, or lamb chops and wants pouter pigeons. Under these circumstances it may pay to exchange the pouter pigeons for wheat now, with the idea later on of exchanging the wheat for eels. After all, there are relatively few people who want pouter pigeons, and almost everyone uses wheat. If we cannot make our direct exchange, we can exchange now for the wheat, and the chances are fairly good that when we do find someone with eels for sale he will be willing to take wheat for them, since almost everyone does use wheat. We want to swap commodity *A* for commodity *B*. Since we cannot do it directly, we swap commodity *A* for commodity *C* with the intention of swapping commodity *C* in turn for commodity *B*. We make two swaps instead of one, which seems to complicate the situation but really makes it easier.

Of course, if we are going to carry out one of these double-barreled exchanges, we ought to make sure that the commodity that we accept temporarily is one in pretty general use—something that almost everyone wants. Then, when we get around to making the other half of the exchange, we can be reasonably certain that the man who has the commodity that we need

will be willing to take our commodity *C*. This commodity *C* the economist calls a "medium of exchange," since it is a commodity used to bring about an exchange of commodities *A* and *B*. We could not exchange *A* and *B* directly, but we did do it through the medum of commodity *C*. We have just pointed out that men ordinarily use as mediums of exchange only things that are in pretty general demand. As exchange becomes more common, men seem slowly to adopt some few commodities as mediums of exchange, and these commodities come into common use for that purpose. Ultimately most exchanges are carried on through these few commodities as intermediates, and men get so that they not only are willing to accept them, but actually prefer them. Then we call these commodities "money." *Money consists of those things which are generally acceptable in any society, and which pass freely from hand to hand as mediums of exchange.* In some countries the money has been gold and silver, in some it has been cattle, and in some it has been beaver skins; but until a commodity is accepted by everyone or almost everyone in the society—until it has become the medium through which all or almost all exchanges are affected—we do not call it money. When a society has progressed to the stage where it does have some sort of money, so that barter has entirely or largely disappeared and all or almost all exchanges are effected through the use of some widely accepted intermediate commodity, we say that the society is in the "purchase" stage, or that it has become a "purchase economy."

It is easy to see that under such circumstances money is actually productive. It enables men to get goods or services with less effort than would otherwise be possible. It enables them to get around the difficulty of double coincidence of wants and supplies, selling their goods always for money and later using the money to buy other goods. Railroads are of no use in themselves, but we speak of them as productive because they make it easier for us to get the goods and services that we want. They bring us desired goods with less expenditure of effort than would otherwise be required. Money is productive in exactly the same sense. It is a laborsaving device. It makes it easier for us to get the goods and services that we want. It brings us desired goods with less expenditure of effort than would otherwise be required. This is often overlooked by those who say that "money is sterile." Money is, in fact, a production good. It is capital.

15·4. Credit. As commerce becomes more common, men often pass beyond the purchase stage into what we call the "credit" stage. Instead of exchanging *A* for *C* with the idea of exchanging *C* for *B* at a later date, men give up commodity *A* now and get nothing back now save a promise of future repayment. Half of the exchange takes place in the present, and the other half in the future. The lapse of time between the two transfers is the essential feature of credit, and this lapse of time raises many troublesome

problems in those societies where credit is common. Yet, as we shall see, credit is convenient, and businessmen or consumers find it greatly to their advantage to use it despite the problems that it raises. Every nation that develops to a position of commercial importance finds it necessary to establish various sorts of institutions to aid in the making of credit transactions.

The student should not assume that any society must be wholly within one or another of the stages just mentioned—wholly within the barter, the purchase, or the credit stage. In modern American society there are still many people who drive to town and barter their surplus eggs for groceries at the country store. There are also many people who sell goods for cash and then use the cash to buy other goods. Likewise there are thousands of transactions every day that involve the use of credit. If we were to say, as we might well do, that the United States had progressed beyond the barter stage, we would not mean that there is no barter at all here, but merely that it is no longer the common or usual method of carrying on exchange. The United States today is either in the purchase or the credit stage, and we shall be in a better position to assign it to one or the other after we have studied somewhat more thoroughly our exchange system, especially our systems of money and banking.

15 · 5. Exchange Requires Evaluation. If men are to exchange things, they must have some method of determining the values of the things to be exchanged. How much of *A* is to be given for how much of *B*? This, of course, would be true even in the pseudo exchange of a Crusoe economy such as that mentioned in Sec. 15 · 1. When Crusoe is trying to decide whether to spend his afternoon hunting goats or gathering oysters, or possibly spending part of the time on one and part on the other, he has to evaluate goats and oysters —he has to decide how badly he wants each, and how many oysters he would have to get before it would pay him to give up the goat that he might otherwise get. In an organized society, where men are dealing with others instead of trying merely to satisfy themselves, the problem is more complicated. Yet it is quite as necessary for value decisions to be made.

It seems especially difficult to arrive at decisions on the matter of values when we remember that usually the producer and the consumer of goods in our society never see each other or know each other. The wheat farmer has no idea who eats the bread made from his wheat. The shoe factory has no way of knowing who wears the shoes that it sends on to the market. Production today is seldom made to order. Most things are made "for the market"—for an impersonal and unknown group of consumers. On the other hand, the ultimate buyer and user of the goods seldom knows the source of supply. Often he does not know the name even of the company that produced the things he uses; in the few cases where he knows the name of the company, as for example when he buys a new automobile, he has no

personal contact. He and the manufacturer do not agree on a value by any process of discussion. In fact, even if the consumer knows what company made the product, usually he does not know personally even a single person who works for that company. How are values to be arrived at for the exchange of goods when exchange takes place in an impersonal something which we call "the market"—when producer and consumer never meet?

The answer is that there are many ways in which such values can be arrived at. They may be determined by government fiat, either on the decision of some autocratic price maker or by some more democratic legal process. Or individual buyers and sellers may be left free to make their own decisions, each following what seems to him to be his own advantage. Each potential seller of goods may be left free to ask as much for his goods as he wishes or to sell for as little as seems to him desirable. Each potential buyer may be left free similarly to offer as high a price as he wishes, or to offer little or nothing. In such a case, where we have a *free market,* the method of setting the price may not be so immediately obvious, but there will still be forces there that determine the price, and the result will by no means be a random one.

No country actually operates its economic system at either of these extremes. In every country there is some government regulation of economic activity, and neither buyer nor seller is left free to follow completely the path along which his own advantage might lead him. On the contrary, in no country are all decisions made by governmental or other authority and no liberty left to buyers and sellers to make decisions of their own. Freedom in the economic sphere, like freedom everywhere else, is a relative matter. We do, however, find very great differences in the degree of freedom of the market in different countries and at different times. In the Middle Ages it was common for prices and qualities of goods to be regulated by law.[1] In modern times both communist and fascist countries have limited closely the freedom of individual buyers and sellers in the making of economic decisions, and the agreements of the Allied Powers for the control of German industry during the period of postwar transition likewise impose many sorts of limitations on the free market. On the contrary, although the market has never been left entirely free in the United States, nevertheless in peacetime a very large proportion of the economic decisions that have been made have been left to buyers and sellers themselves. Our markets have not been entirely free, but they have been relatively so. Under such conditions, the sets of forces that determine market values would be those forces which influence consumer decisions and those which influence producer decisions. These forces we have already studied, and it remains now only to put them together and to ascertain the result of their interaction. This result will

[1] See Sec. 3 • 12.

depend in part on another factor, the degree of competition or monopoly in the market. In the remainder of this chapter we shall turn our attention to the free competitive market, and in the following chapter to price determination under monopoly.

15·6. Nature of Competition. The layman often uses the terms "free market" and "competition" as synonymous. Actually they refer to somewhat different things. As was pointed out in the preceding section, a market is free if buyers and sellers are left free to pursue their own advantage, if they are allowed to determine for themselves what quantities of goods they will buy or sell and at what prices. In order that a market may be competitive, an entirely different set of conditions is necessary. We might have a single monopolistic seller and millions of buyers on the market. If the single seller and the millions of buyers were left free to do what they wished, we would have a free market, but by no means a competitive situation.

In the first place, the word "competition" carries the connotation of rivalry—of people striving against each other. There is no rivalry unless there are rivals, and the single seller of our illustration in the preceding paragraph had none. Even if we have two or three sellers, we do not necessarily have competition. We might have a hundred or a thousand sellers and not have competition. Thus, if the football team of one of the larger universities were to play the football team of a small grammar school, there would be no rivalry—no competition. Taking candy away from a baby is not competition. Competition connotes at least approximate equality among the competitors. In athletic events it is not uncommon, in order to make an event truly competitive, to handicap one of the participants in order to get this approximate equality. Similarly in the economic field, we might have a thousand sellers of a good, but if one of them had certain unusual and important advantages that the others did not share—advantages from the ownership of patents, from the ownership of unique sources of raw materials, from location with respect to market, or from any other source—there would not be competition among the sellers. In order to have competition, the buyers and sellers must be of at least approximately equal strength. They must have common knowledge or at least common access to market information. They must be at least reasonably competent and intelligent, and there must be no discrimination. Sellers must be equally willing to sell to all buyers; buyers must be equally willing to patronize all sellers. This latter stipulation means that the products offered for sale by the different sellers must be identical or substantially so. At least, the differences existing among the products must be differences that are of absolutely no importance to buyers. Thus beet sugar and cane sugar come from different sources, yet from the standpoint of the buyer this difference in source is immaterial. The economist would say that beet sugar and cane sugar are, when they reach the retail grocery,

the same commodity. On the other hand, two manufacturers may produce aspirin tablets of exactly the same weight and purity, alike in all their physical and chemical characteristics, but sold in different bottles. If buyers prefer the bottle with the blue label to the bottle with the red label, and insist on the brand to which they are accustomed, then the two different kinds of aspirin are not really the same commodity, even though the physicist and the chemist can see no distinction between them. From the standpoint of competition, the important thing is the attitudes of buyers and sellers. If they make no distinction between products, then we say that the products are the same. If they distinguish the products, even on the basis of what seem to us foolish distinctions, the products are different. When we say, then, that in order to have competition the buyers and sellers must be purchasing and selling identical commodities, what we really mean is that any differences between the commodities must be matters of indifference both to the buyers and to the sellers, as are the differences between cane and beet sugar.

In addition to having rivalry among equals in the purchase and sale of identical commodities, if we are to have competition, it must be true that no one of the buyers or sellers is important enough in the market so that he has any appreciable influence on the price. This would usually mean in practice that the number of buyers and the number of sellers must be rather large. It is not at all necessary that their purchases or their sales be of equal size as long as each one of them is insignificant when his volume of business is compared with the volume on the entire market. Each seller must be in the position that he can double or treble his output or go out of business entirely without affecting the market price measurably. Each buyer must be in the position that he can double or treble his purchases, or withdraw from the market completely without affecting the price measurably. These last two statements may be put in somewhat different form, which we shall find useful in our later analysis, by saying that the seller must find himself faced by a horizontal demand curve for his own products, while the buyer must find himself faced with a horizontal supply curve as far as his own purchases are concerned.

It may seem strange for us to say that the market demand curve has negative slope while the demand for the product of each individual seller is a horizontal straight line. How can we add together a large number of horizontal demand curves and get a market demand curve of the usual shape? There are two answers to this question. The first is that, although the market demand curve is found by adding together many individual demand curves, those which are added represent the desires of individual consumers, not the demands for the products of individual producers. The other part of the answer is quite as simple when we think of it. Suppose that the data of Table 15·1 represent the demand schedule for

potatoes in the United States. We note at once that the table indicates a decided negative relationship between price and quantity. When the price is high, consumers buy only small quantities; when the price is low, they increase their consumption markedly. At a price of $1 they buy over twice as much as at a price of $1.10, for example. Yet notice what would be true under these circumstances when an individual potato grower varied the amount of his output.

Let us suppose that a particular farmer can grow as many as 100 acres of potatoes, or any smaller amount. By varying his productive practices, he can vary his yield per acre up to a maximum of 300 bushels per acre. His largest possible output under these circumstances would be 100 acres at 300 bushels per acre or 30,000 bushels. He can vary his output at will between 30,000 bushels and nothing at all. Suppose that all the other producers on the market do produce a total of 70 million bushels, which they put on the market. How much difference will it make if our one farmer sells 30,000

Table 15·1. Hypothetical Market Demand for Potatoes

Price per bushel	Quantity demanded, million bushels	Price per bushel	Quantity demanded, million bushels
$1 15	25	$0 95	175
1.10	40	0.90	285
1.05	70	0 85	460
1.00	110		

bushels or refuses to sell any? We can get an approximate answer from Table 15·1. We see that if the farmer sells no potatoes at all the total quantity offered on the market will be the 70 million bushels of the other producers, which will sell at a price of $1.05 per bushel. If 40 million more bushels were offered, the price would drop to $1 per bushel. In other words, it takes 40 million bushels of additional potatoes to reduce the price 5 cents per bushel. Evidently it would take approximately 8 million added bushels to reduce the price 1 cent and 800,000 added bushels would reduce the price 1/10 cent. Eighty thousand additional bushels would reduce the price only 1/100 cent. Even this is more than our farmer can put on the market. If he gets his maximum output and sells it all, he can raise but 30,000 bushels; if he throws them all on the market, they will amount to so little compared with the vast volume already there that there will be no measurable change in price at all—a change of something less than 1/100 cent a bushel. As far as this individual farmer is able to see, the price will not change at all when he varies his output. He can sell 1, 100, 1000, or 10,000 bushels, up to the

maximum that his resources can produce, and the price will stay so close to $1.05 per bushel that he will not be able to see any change at all. To be sure, if a large number of farmers increased their output, so that we got 2 or 3 million added bushels of potatoes on the market, the price would fall noticeably, but no one producer himself would have a noticeable effect. Under these circumstances, each producer finds himself confronted with what appears to be a fixed price of $1.05. Within the limits of his productive ability, he can sell as much or as little as he wishes without any change in the price at all. If he tries to charge 1 cent a bushel higher than $1.05, no one buys from him, since there are plenty of potatoes available from other producers. If, on the other hand, he is willing to sell at ½ cent a bushel below the market price, he can sell as many potatoes as he can lay his hands on. Since there is no discrimination between products in cases of competition, buyers are just as willing to buy from him as from anyone else and, if they can buy from him more cheaply, they all prefer to do so. The smallest conceivable advance in his price above the general market level reduces his sales to nothing, while the smallest conceivable reduction in price below the general market level brings him all the trade he can handle. The demand for his product has infinite elasticity, even though the demand for the commodity on the market as a whole may be inelastic. This is the final characteristic of competition: the demand for the products of each producer must be infinitely elastic, as must the supply of products to any particular consumer.

Seldom in practice do we get the complete competition just described, just as we seldom find the complete monopoly that is described in the next chapter.[1] Even though no business is completely competitive or completely monopolistic, we do find some that are close approximations to complete competition, and some in which the monopolistic elements are significant and important. Perhaps the closest approach to complete competition is in the cases of some staple agricultural products which are carefully graded and sold on an organized market, such, for example, as the grains. In practice, the wheat grown by different farmers is so nearly the same after it had been graded that buyers do not care whether they get the product of one man or of another. In fact, when the grains reach market they are mixed together and the grain from a particular farm is immediately indistinguishable from other grain. Moreover, no wheat farmer is big enough to play any important part in the market. If the biggest wheat farmer in the United States were to go out of business, it would have no appreciable effect on wheat prices. When, on the contrary, we get into the manufacturing fields, we find that the situation is often not quite so competitive. Buyers do distinguish between different makes of automobiles. If one maker raises his

[1] See Sec. 16 • 1.

price slightly, he may lose some trade, but not all. Buyers have preferences for which they are willing to pay.

15·7. Price Setting under Competitive Conditions. It is often helpful in studying a phenomenon to consider it at first under idealized conditions that are not exactly the same as those which one meets in real life. Thus the physicist reasons about the rates at which bodies would fall in a perfect vacuum, even though in practice perfect vacuums do not exist. Similarly, even though we may never get complete competition, it is helpful for us to consider how values would be set if there were complete competition, and we may later make whatever adjustments in our results seem wise to offset the error of our assumptions. Let us assume, then, that there are many buyers and sellers of a homogeneous commodity, and that no one of them is important enough to have any individual effect on the market situation. Each is trying to do what seems most advantageous from his own standpoint, and each is left free to do what he wishes. We have all the conditions, we assume, of competition. The commodity, let us say, is wheat, and we wish to see how its price will be determined under these circumstances.

Each buyer or prospective buyer of wheat will have some given tastes and resources to determine his individual demand for wheat, and these individual demands taken together will compose some market demand. Similarly, the individual supplies of the prospective sellers, determined by their costs and by the techniques of production, will compose when taken together some market supply. For any economic good at any time, there is some market supply and some market demand. The supply and the demand may change from time to time—almost certainly do so—but at any given moment they are fixed. So let us ask what are the supply of and the demand for wheat now. Suppose that they can be represented by the supply schedule and the demand schedule of Table 15·2, which shows at various prices the quantities of wheat that sellers will be willing to sell and the quantities of wheat that buyers will be willing to buy. The former set of quantities increases as the price rises, while the latter set decreases. Thus the figures illustrate the Law of Supply and the Law of Demand. These two laws, we recall, are themselves dependent on the laws of cost and on the Law of Diminishing Utility, respectively.

We see from the table that if the price is $1.70 per bushel, buyers, reluctant to buy at such a high price, are willing to take but 103 million bushels, while ardent sellers, eager to take advantage of the high price, are willing to sell 197 million bushels. How can they sell 197 million bushels to buyers who want but 103 million? Evidently, if they try to sell at this price they cannot do it, and many sellers will have wheat unsold. Yet the table shows that many of them are willing to sell at prices below $1.70 and, if each seeks his own advantage, no one will allow himself to be left with wheat unsold when he

would have been willing to accept lower prices. Someone will cut prices in order to get trade and, since there is no discrimination, the buyers will turn to the men who offer wheat at the low prices. In other words, competition among the sellers, each trying to get rid of his product, will tend to reduce the price below $1.70. This price is too high and cannot continue if there is competition. The price of $1.70 stimulates sellers to offer a quantity far larger than the market will absorb at that price; as a result, the price is lowered.

If, on the other hand, the market price is 90 cents per bushel, the situation is very different. Now we see from Table 15·2 that sellers, dissatisfied with the low price, have reduced their offerings to 108 million bushels, evidently holding the rest in storage for a better market. Buyers, however, are anxious to take advantage of the low price while it lasts, and they are willing to buy 185 million bushels. How can they buy 185 million bushels from

Table 15·2. Supply of and Demand for Wheat

Price per bushel	Quantity offered, million bushels	Quantity demanded, million bushels	Price per bushel	Quantity offered, million bushels	Quantity demanded, million bushels
$1.70	197	103	$1.20	162	136
1.60	191	106	1 10	149	149
1.50	185	111	1 00	133	165
1.40	180	117	0.90	108	185
1.30	172	126	0.80	80	215

men who are willing to sell but 108 million? It is evident that many buyers will be left with wants unsatisfied. Yet the table shows that many buyers are willing to offer more than 90 cents and, if each buyer seeks his own advantage, no one will allow himself to remain with wants unsatisfied at 90 cents when he is willing to pay more. Someone will offer more than 90 cents in order to make sure that he gets the wheat that he desires and, since the sellers do not care to whom they sell, they will all turn to the buyers who offer more. Competition among the buyers, each trying to satisfy his own wants, will raise the price above 90 cents a bushel. This price is too low and cannot continue if there is competition. The price of 90 cents stimulates buyers to seek a quantity far larger than the market offers at that price; as a result the price is raised.

We could take any other point in the table and go through the same sort of reasoning, but we can see already that, whenever the price is so high that the sellers offer more for sale than buyers are willing to take at the high price, competition among the sellers will tend to lower the price; whenever

the price is so low that buyers desire more than sellers are willing to offer at the low price, competition among the buyers will tend to raise the price. In both cases we rely on competition to bring about the adjustment, but in the one case it is competition among sellers and in the other case among buyers. Competition among sellers lowers the price; competition among buyers raises the price. We see at once that there is but one price in the table where one or the other of these types of competition will not spring up; namely, the price of $1.10 per bushel, at which buyers are willing to take 149 million bushels and sellers offer 149 million bushels for sale. At this price, and only at this price, the quantities offered and demanded are equal. At this price, and only at this price, every buyer who is willing to pay the price can find a seller, and every seller who is willing to accept the price can find a buyer. There is no longer reason for buyers to compete with one another and raise the price further, since all buyers who are willing to pay $1.10 can buy as much as they want without raising the price further. Nor is there any reason for sellers to compete with one another and lower the price further, since all sellers who are willing to accept $1.10 can find customers already without reducing their price further. Prices above $1.10 are unstable and will be reduced by competitive selling; those below $1.10 are unstable and will be raised by competitive buying. If there is competition, the market price cannot stay above or below $1.10, and competition will quickly force it to that point, where it will stay. With this supply and demand, the competitive market price will be $1.10 per bushel, and this price will "clear the market"; that is, every buyer who is satisfied with the price can buy, and every seller who is satisfied with the price can sell. To be sure, there are some sellers who do not sell at $1.10, because they were insisting on a price of $1.70 or $1.60. They were not satisfied with the price. There are, also, some buyers who do not buy, because they were waiting for the price to drop to 80 or 90 cents. They were dissatisfied with the price of $1.10. But everyone who was willing to do business at the price of $1.10 could do it, and that is what we mean when we say that the competitive price "clears the market." No other price would do it. For example, if the price is $1 per bushel, some people who are willing to buy at the price cannot do so, because but 133 million bushels are offered and 165 million are desired. People who were quite satisfied with a price of $1 could not find sellers. At a price of $1.40 some sellers who are quite willing to accept $1.40 cannot get it because they cannot find buyers. There are buyers for but 117 million bushels, while sellers are willing to sell 180 million bushels at a price of $1.40. *In a competitive market, the price tends to be set at the point where the quantity offered by sellers and the quantity desired by buyers are equal.* If for any reason the price is actually at some other point, the forces of competition will restore it, lowering it if it is too high and raising it if it is too low.

The situation can be visualized more easily if we show it graphically, as in Fig. 15 • 1, where the supply curve and the demand curve have been taken from Table 15 • 2 by connecting the prices and quantities of that table by smooth curves. We note that the supply curve, as usual, slopes upward toward the right, and the demand curve has the usual negative slope. Toward the upper part of the diagram, the supply curve lies farther to the right than the demand curve, indicating that sellers are offering larger quantities than buyers are willing to take at the price. When the supply curve lies to the right of the demand curve, the price is higher than can be maintained. Toward the lower part of the diagram, the demand curve lies farther to the right than the supply curve, indicating that buyers are willing to buy larger

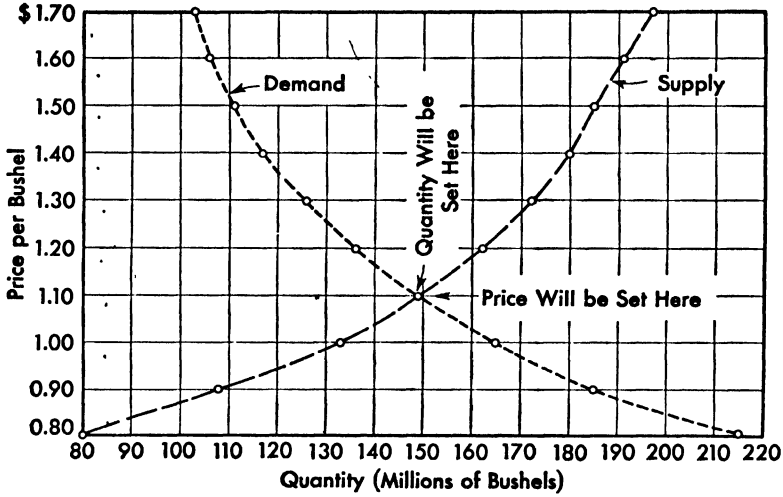


FIG. 15 • 1.—The setting of price under competitive conditions.

quantities than sellers will offer at the price. When the demand curve lies to the right of the supply curve, the price is lower than can be maintained. The equilibrium price—the competitive market price—corresponds to the point where the supply curve and the demand curve cross. Our diagram shows us that the price will be $1.10 per bushel if there is competition and that 149 million bushels of wheat will actually change hands. Always, of course, the quantity sold and the quantity purchased are exactly the same, since each sale is also a purchase. In the present case, the quantity sold is 149 million bushels and the quantity purchased is the same.

The student should not be led to suppose that the forces about which we have just been speaking work themselves out immediately and with complete precision. As in the cases of other economic laws and principles, we are here stating a general tendency. Yet the tendency is a pronounced and powerful one, so much so, in fact, that when the United States government

tried, through its Agricultural Adjustment Administration, to raise the prices of certain farm products that had been forced to low levels by what was loosely called "surplus production," they found that the artificially high prices that they set discouraged consumption and stimulated even greater production, thus aggravating the problem. Finally, they were able to bring about some rise in prices only when they coupled price control with production control, by paying farmers not to produce crops. Price control alone does not work because, when we alter the price, we upset the equilibrium that has been established between the quantities offered and the quantities demanded.

Some economists have summarized the statements that we have just been making about the setting of prices under competitive conditions by means of what they call the "First Law of the Market": *Under competitive conditions, there can be but one price for a given commodity in a given market at a given time, and that will be the price that equates the quantity offered by sellers and the quantity desired by buyers.* This law is correct, of course, only under the assumptions that we have made. These include not only the assumption of complete competition, but the assumption that the commodity is entirely uniform in quality—that no buyer or seller cares which unit of the commodity he gets or gives. When the student notices in practice that different grades of wheat sell at different prices or that different horses or houses or automobiles have decidedly different values, he is apt to think that he has run across an exception to the First Law of the Market, whereas he has merely run across an indication of the artificiality of the assumptions under which the law is formulated. The student who drops a marble into a pail of molasses will find that it does not fall in the same way as the "freely falling bodies" which the physicist imagines falling in a vacuum, yet these laws of falling bodies are very useful as giving first approximations to what does happen in practice. Similarly the First Law of the Market is very useful in giving us a first approximation to what does happen on the market, and especially in helping us to understand the kinds of forces that are at work and the way in which they make themselves felt. Still it is only fair to say that some of the most important and most interesting problems of economics arise in those cases where the commodity is not uniform, but where consumers or producers prefer certain units to others. This common tendency to prefer some grades or qualities of a commodity to others is called "product differentiation," but it would not exist in a strictly competitive situation, since by definition there is no discrimination between products in such a situation. If we really had complete competition, we think that prices would be set exactly as we have described, that there would in fact be but one price on the market at any time, and that that price would "clear the market."

15 · 8. Effects of Changing Demand or Supply. In Chap. 6, when we were considering the characteristics of demand, we stressed the fact that changes in the price of a commodity do not change the demand for it. In Chap. 14, when we were considering the characteristics of supply, we noticed also that changes in the price of a commodity do not change the supply of it. Neither the supply of nor the demand for a commodity is affected by changes in its price. We now turn to the other side of the problem and ask what will happen to the price of a commodity if we change the demand for it or the supply of it. The competitive price, we have seen, depends on the supply and the demand at the time. But supply changes every time there

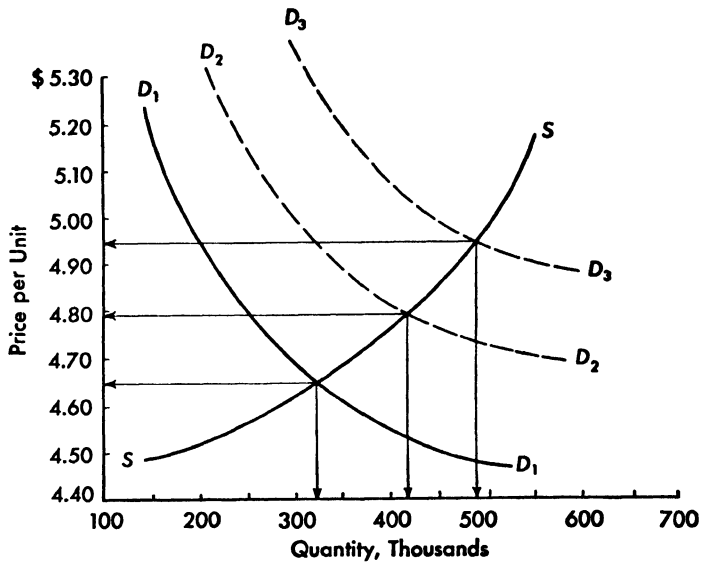


FIG. 15 · 2.—Effect of changes in demand when supply is constant.

are changes in production methods and costs, and demand changes every time there are changes in tastes or incomes. What are the effects of such variations of supply or demand on the competitive market price?

We can understand these effects most easily by a diagram similar to that of Fig. 15 · 1. In Fig. 15 · 2 let us assume that the positively sloping curve labeled S is the supply curve, and the lowest curve of negative slope, labeled D_1, is the demand curve. If there is competition, the price will be set opposite the point where these two curves cross, or at $4.65 approximately; and some 325 thousand units will be exchanged. We remember from Sec. 6 · 12 that a change in demand would be shown by altering the position of the demand curve. Let us suppose that the demand for this commodity increases, so that consumers are willing to buy more at any price than before. Under these

circumstances the old demand curve D_1 disappears and is replaced by a new curve such as D_2 which lies higher and farther to the right. Under the new conditions there are but two curves, S and D_2. The curve labeled D_1 on the diagram is past history. It shows what the demand used to be; the actual present supply and demand are S and D_2. The competitive price will correspond to their crossing point, so the price will be \$4.79 and 420 thousand units will be exchanged. If the demand rises still higher, to the position indicated by curve D_3, the price will rise still higher to \$4.94 and the quantity exchanged will rise to 490 thousand units. The student must remember that at any given time but two of the curves shown in Fig. 15·2 exist: the supply curve and one of the demand curves. We have assumed that the three demand curves of the diagram existed at three different times, and that increases in demand shifted us successively from D_1 to D_2 to D_3. These increases in demand, with supply unchanged, raised the competitive market price from \$4.65 to \$4.79 to \$4.94, and increased the quantity of the commodity that changed hands from 325 thousand to 420 thousand to 490 thousand units. It is easy to see from the diagram that we can lay down the general rule that *changes in demand change the price and the quantity exchanged in the same direction.* If we increase the demand, we increase the price and the quantity; if we decrease the demand, we decrease the price and the quantity.

The amount by which a given change in demand affects the price and the quantity varies with different cases, depending on the elasticity of the supply. This is illustrated in Fig. 15·3, where the upper and the lower half of the figure represent demand curves that are exactly alike, but the upper half represents a supply that is less elastic than that in the lower half. This is at once evidence from the slopes of the two supply curves.[1] The rather steep slope of the supply curve in the upper part of the diagram indicates that it takes a rather large price change to bring about even a rather small change in the quantity offered. The supply is inelastic. The rather gentle slope of the supply curve in the lower half of the diagram indicates that even a rather small price change will bring about a rather large change in the quantity offered. The supply is elastic. In the two diagrams it is assumed that the demand changes from position D_1 to D_2. The changes are identical, but the effects are different. In the upper diagram, with inelastic supply, the change in demand brought a large change in price but only a small one in

[1] The student who has followed the work of the Mathematical Appendix will realize that the elasticity varies from point to point along these linear demand and supply curves, and that the elasticity of a curve cannot be ascertained directly from its slope on arithmetical scales. Yet, if the two diagrams are drawn to the same scale, it will be true of course that the steeper curve is the one of smaller elasticity at any particular quantity or price. Hence, with the proviso that the two diagrams have the same scales, our conclusions above are valid.

quantity. In the lower diagram, with elastic supply, the same change in demand brought a small change in price and a large one in quantity. In each case the increase in demand raised the price and increased the quantity exchanged, but *the more elastic the supply of a commodity, the greater will be the increase in quantity and the smaller will be the increase in price when the demand rises.*

It would be good practice for the student to make another pair of

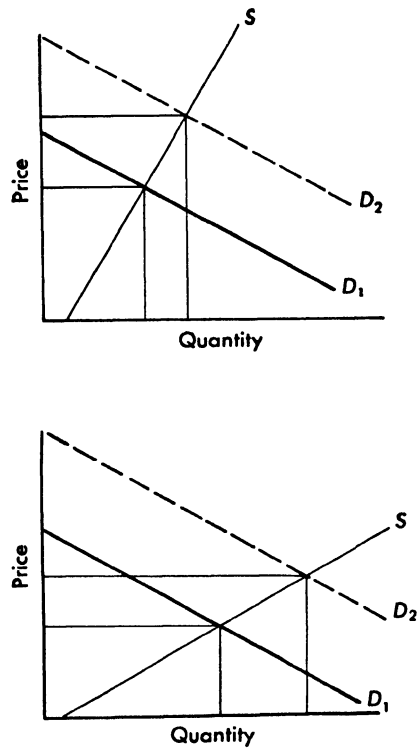


FIG. 15 · 3.—Effect of differences in elasticity of supply on the results of changing demand.

diagrams, similar to those of Fig. 15 · 3, showing the effects of elasticity of demand on changes in supply. Both diagrams should show exactly the same supply curves, but one should have a very steep and inelastic demand curve, while the other has an elastic demand curve that is more nearly horizontal. Such diagrams will demonstrate that *the more elastic the demand for a commodity, the greater will be the increase in quantity and the smaller will be the decrease in price when the supply rises.* These diagrams will also demonstrate that *changes in supply change the quantity exchanged in the same direction but the price in the opposite direction.* If we increase the supply,

we increase the quantity and reduce the price; if we decrease the supply, we decrease the quantity but increase the price.

Finally, we may ask what will happen to the price and quantity if the supply and the demand both change. Naturally if the demand increases and the supply decreases, since both changes tend to raise the price, the result will be a marked increase in price. Likewise, if demand decreases when supply increases, the price will fall radically. The only question arises in case supply and demand both increase or both decrease. Then one of the changes will tend to raise the price, and the other to lower it. Whether the end result is a rise or a fall in price depends on whether the change in demand was greater or smaller than the change in supply, and on the relative elasticities of the two curves. We have said[1] that short-run supply is usually rather inelastic, and as a result changes in demand bring rather large immediate changes in price; but, since long-run supply is more elastic, the long-run price effects are not so great. After producers have had time to readjust their businesses to the changed demand, the ultimate effects will be largely taken up by changes in quantity, even though the immediate effects will be very largely price effects.

In earlier chapters we have emphasized that changing prices do not change either the demand for or the suppply of a commodity. Now we have seen that, on the contrary, changes in either the demand or the supply do change the price. The relationship is unidirectional, from demand and supply to price and not from price to demand or supply. When the layman makes the two loose statements that raising the demand for a commodity will raise its price, but that raising its price will lower the demand for it, he has led himself into circuitous reasoning which forces him to conclude that the ultimate result of raising the demand for a commodity is to lower the demand for it! If he uses the same definitions of demand and supply throughout, one of the statements must be wrong; if he uses the definitions that we have adopted, which define both supply and demand as relationships between price and quantity, the first of his statements is right and the other one is wrong. The confusion is a common one which the student should endeavor to avoid.

15·9. Law of Supply and Demand. Every student who begins the study of elementary economics comes to it with a belief that the subject can be summed up in the Law of Supply and Demand; yet, if he is asked to write out a statement of the law, he usually runs into immediate difficulty. It becomes evident at once that the phrase means one thing to some people and something else to others, and that to most people it means nothing very definite. Some wit once said that, if you could teach a parrot to say "supply and demand," he would make a good economist. This law has been called

[1] See Sec. 14·11.

by others, "the little tin god of the economist." Certain it is that the layman thinks of it as being somehow the keystone of the economic structure.

In fact the Law of Supply and Demand is but part of the set of principles that we have just been investigating—the principles that determine how competitive prices are set. There is no particular reason for stating the law as such, since it is all included among the principles that have appeared in italics earlier in this chapter but, since the student has surely heard people talk about the law before and since he will run across continued references to it later, we can say that the Law of Supply and Demand states that *the price of a commodity varies directly with the demand for it and inversely with the supply of it.* Although public speakers always seem to call it "immutable," it is well for the student to remember that the principle is one which is peculiarly applicable only in cases where there is complete or substantially complete competition. He should also remember that the law is by no means so simple as it appears at first sight, since the words "supply" and "demand" themselves represent rather complicated sets of forces, such as costs, technology, consumer tastes, diminishing utility, diminishing productivity, and consumer incomes. When the economist uses the term "supply" or the term "demand," he is summarizing briefly the effects of forces that are by no means simple.

15 · 10. Summary. Exchange as we see it in our society is the result of a long, slow process of growth and change, which has brought with it increasingly complicated methods of exchange because, paradoxically enough, the more complicated methods make exchange easier. Exchange, as we know it, is largely an effect of our type of social organization; fundamentally, the problems of exchange go back to the fact of scarcity and the necessity for choosing between alternatives. If men are to exchange, it is necessary for them to evaluate the things exchanged. The method by which this evaluation takes place depends on circumstances; particularly, on the presence or absence of what we call "competition." In a free and competitive market, there would be but one price for a commodity, and this would be the price that equated the quantity offered and the quantity demanded—the price that cleared the market. If the price differed from this, it would be restored by the action of competitive forces. Changes in either the supply of a commodity or the demand for it tend to change the price and the quantity that is exchanged. The amount of the effect is determined by the amount of change in supply or in demand, and by the elasticities of the supply and the demand. In general, if one of the curves changes, the effect will be mainly a price effect if the other curve is inelastic; mainly a quantity effect if the other curve is elastic. Competitive market price tends to vary directly with the demand and inversely with the supply; the quantity exchanged on the market tends to vary directly with either demand or supply.

Questions

1. Why would we not have exchange of our present type in a society that did not recognize private property?

2. Why would we not have exchange of our present type in a society in which there was no specialization?

3. Could a society have credit without money? Barter credit?

4. Various societies have used various commodities as mediums of exchange. Does a society select its medium by chance? If not, what factors guide it in the selection of a commodity for this purpose?

5. The text says that no country actually operates with a free market. In what ways are the markets of the United States not free?

6. Why is it necessary that the various buyers and sellers on a market have approximately equal intelligence, knowledge, and power if there is to be real competition? Illustrate.

7. Suppose that two companies manufacture baseballs that are exactly alike except for the fact that each bears the trade-mark of the maker. Under what circumstances would you say that these manufacturers were producing the same commodity, and under what circumstances would you say that the two kinds of baseballs were different commodities?

8. "The demand for wheat in the United States as a whole is rather inelastic, yet the demand for the wheat of every single wheat farmer in the United States is very elastic indeed." Do you think that these statements are probably true? Why or why not? Can they both be true at once? Why or why not?

9. Suppose that the demand for wheat is that shown in Table 15 • 2, but suppose that the supply is so altered that sellers offer 10 million bushels less at any price than the amounts shown in the table. Thus, for example, when the price is $1.70 per bushel, sellers offer 187 million bushels rather than the 197 million as shown. Under these circumstances, what will be the competitive market price? If you cannot give the exact price, what is the most accurate thing you can say about it?

10. Draw a diagram similar to Fig. 15 • 1, showing the facts of question 9. Estimate from your diagram as accurately as you can what the competitive market price will be.

11. Is the demand of question 9 larger or smaller than the demand of Fig. 15 • 1? Is the resulting price higher or lower? Is the quantity exchanged larger or smaller? Give the figures for each case.

12. Explain why price control does not work when it is not accompanied by other controls.

13. State in simple terms the relationship of elasticity of supply to competitive price changes.

14. Suppose that we have two different commodities with exactly the same demand schedules, but their supply schedules are different. They are shown in Fig. 15 • 4.

The supply curve for commodity number 1 is labeled S_1, while the supply curve for commodity number 2 is labeled S_2. The demand for each commodity is labeled D_1. We note that the prices of the two commodities will be the same (what will they

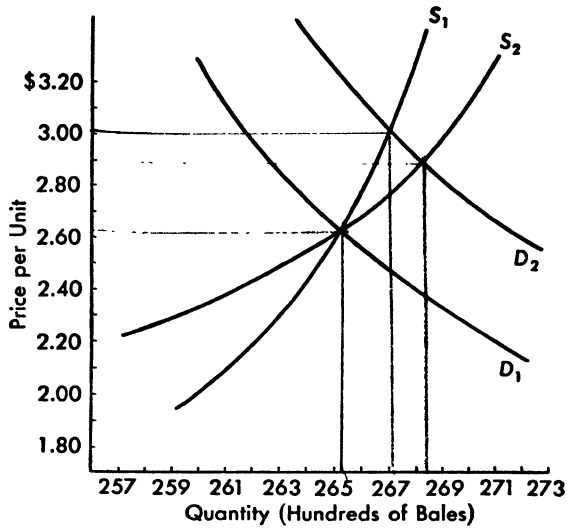


FIG. 15 • 4.—Illustrating question 14.

be?) and the quantities exchanged will be identical (what quantity?). Now suppose that the demand for both commodities shifts from position D_1 to D_2. The prices of the two commodities will no longer be the same, nor will the quantities exchanged be equal. Why not? Which commodity shows the greater change in price? Why? Which shows the greater change in quantity? Why?

15. The mathematically inclined student may like to solve this simple problem with linear demand and supply functions. The supply of a commodity is represented by the equation $Q_1 = 50 + 70P$, and the demand for the commodity by the equation $Q_2 = 550 - 30P$, where Q_1 is the quantity offered for sale in millions of pounds, Q_2 is the quantity in millions of pounds that consumers are willing to buy, and P is the price in dollars per unit. What will be the competitive market price? How many different principles are illustrated either in the given equations or in your manipulations of them? Show the problem graphically. What is the elasticity of demand and what is the elasticity of supply at the competitive market price?

16. MONOPOLY PRICE

16 · 1. Nature of Monopoly. As we saw in Chap. 15, complete competition seldom or never exists in practice. Complete monopoly never exists, being even more unlikely than complete competition. We should define monopoly as the opposite extreme from competition. Under competition, if a seller raises his price slightly, he sells nothing; if he lowers his price slightly, he can increase his sales tremendously. The demand for his product is infinitely elastic. If we really had complete monopoly, a seller would find that the elasticity of the demand for his product was zero. He could raise the price greatly, and buyers would not reduce their purchases at all. If he lowered the price, there would be no increase in sales. In other words, we would have the extreme case of monopoly if one seller had complete control of a commodity that people had to buy in a fixed amount, so that rising prices did not send them to other producers or to other commodities.

In practice, there are more or less satisfactory substitutes for every commodity that men buy or sell. Since Henry Ford is the only man who can make the Ford automobile, one could say that he has a monopoly in their manufacture. But we would not be likely to think of this as such, because other manufacturers make automobiles sufficiently like the Ford to be reasonably good substitutes and, if the price of the Ford is raised too high, buyers will shift to the other makes. Even if one firm bought up control of all the automobile companies, we would not have complete monopoly, since consumers could still ride motorcycles or bicycles or ride in oxcarts. As the substitutes for a commodity become more and more numerous and more and more satisfactory, the element of monopoly becomes less and less important, until if we consider the case of the wheat grower we find that there are thousands of other producers who produce an output that is a completely satisfactory substitute for his, and the monopoly element has disappeared entirely.

As we have said, complete monopoly never exists in practice, and the seller never finds that the elasticity of demand for his output is zero. But the less elastic the demand for his output, the greater is the monopoly element in his business and, as we saw in Sec. 8 · 2, every producer tries by various

means to decrease the elasticity of the demand for his product. This, then, is but another way of saying that he is trying to take his product out of competition with others—to get people to believe that his product is different and superior. The seller tries to get buyers to build up *product differentiation* for his product.

16·2. Legal Definition of Monopoly. Although the economist finds it useful to distinguish the degree of monopoly by means of the elasticity of demand, and to recognize competition only at the extreme where elasticity is infinitely great and monopoly at the other extreme where elasticity is infinitely small (equal to zero), the law has to be more "practical" if it is to recognize as monopolistic any businesses that actually can or do exist. The economist would recognize all shades of difference, grading from real complete competition (which would be very unusual) to complete monopoly (which is nonexistent). But the law insists that a firm be classed as one thing or the other; either it is or it is not a monopoly. Yet the law does not make the decision on the basis of mere size. A firm may be very large indeed, like the Great Atlantic and Pacific Tea Company, yet it may have thousands of competitors and its volume may be but a drop in the bucket compared with the total business done in the field. Most men in the grocery business would probably complain that the business is too competitive, not that it is monopolistic.

The law would hold that *a monopoly is any agency that has sufficient control over the supply of a commodity to influence the price.* When the individual wheat farmer decides to double or quadruple his acreage and output, it makes no difference at all in the market price. But, if a given steel company, let us say, gets so large that its decisions do affect the price of steel, then we would say that the company was a monopoly. If we use this definition, there are actual monopolies in practice. We shall continue to use the term "monopoly" mainly in this legal sense, remembering however that the differences between monopoly and competition are really differences in degree rather than in kind, with most business firms coming somewhere between the two extremes, although certainly in most cases closer to the end that represents complete competition than to that which represents complete monopoly.

In practice, of course, we may not have the entire output in the hands of one seller, but we may have two large firms which share the market between them, or we may have several firms in the business but not enough for real competition. When there are two sellers who divide the market, we have a case of "duopoly" rather than a monopoly; when there are a few firms—more than two, but not enough to get real competition—the case is one that we call "oligopoly." These intermediate stages have some of the attributes of monopoly and some of the attributes of competition.

16 · 3. Monopolies of Demand. We have been talking of monopoly in terms of the number of sellers and their characteristics, which, of course, corresponds with the etymological meaning of the word. But it is possible, even though less common, to have the organization and power on demand side rather than on the supply side of the market. We may have only one buyer for a commodity, we may have two, or we may have several but not enough to get real competition. These cases, analagous to monopoly, duopoly, and oligopoly, go under the names of "monopsony," "duopsony," and "oligopsony." Thus, if we have a firm that is the only employer of a given kind of labor and if all the workers who can do the particular kind of work are organized into one union, then collective bargaining between the employer and the union will be a case of price setting between a monopoly and a monopsony. We shall confine our attention here to the cases of monopoly, omitting those of controlled demand and of oligopoly and duopoly.

16 · 4. Monopoly and the Public Interest. Men have long feared monopolies, especially in the United States. They have felt that when there is competition the public is fairly well protected against unfair and unreasonable prices by the automatic forces of competition, which we described in Chap. 15. If sellers try to set their prices too high, they will find that buyers patronize their competitors and they are themselves merely left with their goods unsold. Of course, if the sellers all agreed to raise their prices together, they might find it advantageous, but such collusion is incompatible with competition, where, by definition, each buyer and each seller is pursuing his own individual advantage rather than group advantage. Such an agreement to raise prices would be called by the lawyers "combination in restraint of trade," and would be illegal in most cases in the United States.

Although unduly high prices bring their own remedy under competition, they do not necessarily do so under conditions of monopoly. If the monopolist sets his price unreasonably high, consumers cannot shift to his competitors, since there are none. The realization that monopoly price may be set and remain unreasonably high has led consumers to fear monopolies and to legislate against them. We shall see later that another fault of monopoly is that it tends to hold productive resources idle and thus to reduce the output of society; but it should be said that this difficulty is one which has received almost no attention from lay consumers and has had almost no influence on legislation. The point is a valid and important one, but it is so technical that it is apt to escape the attention of all save those who have made some study of economics, and the layman is apt to focus his entire attention on the price effects.

It is not our purpose to imply here that monopolies are all bad, nor that their effects are always antisocial. As a matter of fact, not only are some kinds

of monopolies tolerated, but some are encouraged. It is a mistake to assume that all monopolies are illegal in the United States. The situation will perhaps be clearer if we first group monopolies into a number of classes, so that we can see more readily why it is that the law treats them differently.

16 · 5. Classes of Monopolies. Monopolies may be classified under several different headings, in accordance with the following outline:

1. Those which exist by social sufferance.
   a. Those which are considered to be socially beneficial, called "public-welfare" monopolies.
   b. Those which are not of public benefit, but which are either harmless, or, if harmful, have not yet been abolished by society.
2. Those which could not be abolished even if we wished, but which exist because of some peculiarity that we cannot control. These are called "natural monopolies."

Public-welfare monopolies are not uncommon. The Constitution of the United States makes explicit provision, for example, in Art. I, Sec. 8, par. 8, for patent and copyright monopolies. It is assumed that society benefits by them. Not only is invention stimulated, but inventors are encouraged to make their inventions public, so that after a limited time everyone may get the advantage of them. The alternative would be to have inventors keep their processes secret as long as possible, thus sometimes keeping them permanently out of use, and sometimes even encouraging the addition of the process to the "lost arts." Other public-welfare monopolies would include those of a government, which are established in some countries as a method of raising revenue in lieu of or in addition to taxes. For example, monopolies of salt and of matches have been common in some countries and have been fruitful sources of government income. Such monopolies are called "fiscal monopolies." Governments have also sometimes monopolized industries in order to regulate the consumption of a commodity. Thus government monopolies of the liquor traffic or of narcotics have often been primarily sumptuary in nature rather than fiscal. Probably the major purpose of the post-office monopoly in the United States is a sumptuary one, intended to increase communication. These monopolies we call "sumptuary monopolies."

If monopolies are really ever harmless, we do not need to worry about them; when harmful monopolies, which could be abolished or controlled, are allowed to exist as a result of social apathy or private intrigue, it is obvious that the situation is one that calls for correction. Such monopolies may arise because of special advantages given by the public, such as exclusive franchises or tariff favors. The former have been very common, although not by any means always harmful.[1] The latter have been less open and perhaps

[1] See Sec. 26 · 3.

less common, yet the protective tariff has often been called "the mother of trusts" because of the belief that tariff favors to private industries have fostered the growth of monopoly. It seems unlikely that the tariff has played a very important part in this way as an independent and unassisted agency, yet it is likely that, when a tendency toward monopoly already existed on other grounds, tariffs have aided and hurried the trend. Private agencies, such as the railroads, may also give favors that aid and encourage monopoly, as was true in the days when railroad rebates were common. Under this system the railroads usually charged the same rates to all shippers, but entered into agreements with certain ones to return to them a considerable part of the rate, so that the net shipping cost was lower to some firms than to others. The favored firms were naturally able to undersell their competitors and to drive them out of business, establishing a monopoly position for themselves. At present such favoritism on the part of the railroads is illegal in the United States.

If we turn now to the natural monopolies, we find that they spring largely from three kinds of sources. Sometimes they arise because of limitations in the natural supply of materials, making it impossible for these supplies to be distributed among hundreds of different owners. Thus, all the known deposits of anthracite coal lie in one small area in Pennsylvania, and it is natural that they should be controlled by a rather small number of mining companies. In the second place, monopoly may arise from secrecy, as when the inventor of a new process prefers to keep it secret rather than to patent it. As was pointed out earlier in this section, the patent laws are intended to encourage publication of new processes and to substitute a limited patent monopoly for the monopoly through secrecy which would otherwise be likely. In the third place, there are some businesses in which it is possible to make great economies as the volume of business is increased. We have seen[1] that average cost curves are usually U-shaped. If the curve continues to fall long enough before reaching its minimum point, it will be true that there will be economies of large-scale operation. A large firm will be able to produce at lower cost than a small firm, and thus to squeeze the small one out of business. If these economies still continue, a few firms will tend to get larger and larger until all but a few are driven out of business. Possibly only one large firm will remain. At any rate, when the nature of the business is such that large firms are more efficient than small ones, we have a natural tendency toward increased size and ultimate monopoly.

16·6. Factors Influencing Monopoly Price. We have seen that prices are set under conditions of complete competition by two opposing sets of forces which we call "supply" and "demand." When there is monopoly, the supply

[1] See Sec. 14·4.

concept does not have the same validity that it has when there are many competing producers.[1] The supply schedule or supply curve shows us how many units of the commodity will be offered for sale at each price; when a firm has a monopoly, it can set the price or the quantity wherever it wishes. If it likes, it can set the price at $2.50, in which case it will sell as many units as consumers will buy. The monopolist in this case sets the price, and the quantity is determined by the conditions of demand. Or the monopolist may decide to produce 1000 units a month and sell them for whatever he can get. In this case the monopolist determines the rate of output, and the price is determined by the conditions of demand. Instead of having the usual supply situation, where both the price and the quantity on the market are beyond the effective control of any particular producer, we have a situation in which the monopolist may control either the price or the quantity on the market, but not both. He controls either one that he wishes, and the other one is determined by the demand for the commodity.

Although the monopolist does have an advantage that the competitive producer does not enjoy, in that the monopolist is able to determine either the price or the quantity on the market, nevertheless he is by no means a free agent in the setting of his price or in determining the amount of his output. The fact that he is a monopolist does not enable him to determine the demand for his commodity and, as we discovered in Sec. 7 · 4, the elasticity of demand will determine what happens to total consumer expenditures when the monopolist alters his price. But these total consumer expenditures will be the monopolist's income. The consumer who sees the money going out calls it "expenditure," but the monopolist who sees the same money coming in calls it "income" or "revenue." We can easily adapt the principles of Sec. 7 · 4 to our present situation by restating them thus:

1. If the demand for a commodity is elastic, the total amount received by producers will vary inversely with the price, rising when the price falls and falling when the price rises.

2. If the demand for a commodity is inelastic, the total amount received by producers will vary directly as the price, rising when the price rises and falling when the price falls.

3. If the demand for a commodity has unitary elasticity, the total amount received by producers will remain constant as the price fluctuates.

The student who has forgotten why this is true should reread at this time Sec. 7 · 4, where numerical illustrations are given.

The three principles just stated recount what happens to the producer's total income as he alters his price. As we have said, however, the monopolist

[1]These remarks might be expanded by pointing out that the concept of demand does not have the usual validity in cases of monopsony, etc.

has his choice of varying either his price or the volume of his output. Since increases in his output will lower the price,[1] we can easily restate the three principles to recount what happens to the producer's total income as he varies the volume of his output. Because of the fact that changes in volume of output change the price in the opposite direction, we merely recast the three principles accordingly, thus:

1. If the demand for a commodity is elastic, the total amount received by producers will vary directly with the volume of output, rising when the volume rises and falling when the volume falls.

2. If the demand for a commodity is inelastic, the total amount received by producers will vary inversely with the volume of output, rising when the volume falls and falling when the volume rises.

3. If there is unit elasticity of demand for a commodity the total amount received by producers will remain constant as the volume of output fluctuates.

Surely, then, a monopolist cannot just set his price or volume at random. He may have the power to set his price anywhere he likes, but the amount of his total income will be affected by the price that he selects and, if he wishes to make the greatest possible net income or profit, he will have to give some attention to deciding what price will yield him the largest gross income.[2]

The elasticity of the demand for his product is not the only factor to influence the monopolist in deciding on his volume of output or on his price. To be sure, as he changes his price or his volume, there will be changes in his gross revenue (total income); also as he changes the volume of output, there will be changes in his costs. The mere fact that a man is a monopolist does not allow him to fix his costs wherever he wants. His factory will still operate under conditions of diminishing productivity and, after a certain volume of output is reached, he will find that added output can be obtained only at increased marginal and average cost. His cost curves will have exactly the same characteristics as those of competitive producers, and for exactly the same reasons.[3] When a monopolist varies the volume of his output, then, there are alterations in his gross income that depend on the elasticity of the demand for his product, and there are alterations in his costs that depend on the technical methods of production. These changes in gross income and in cost will, of course, determine his net income or profit. Thus, while the forces that influence him are not exactly the same as those which influence

[1]This is, of course, the distinguishing characteristic of monopoly. The competitor can increase his output without affecting the price, but a monopolist is one whose market importance is so great that his variations in output do affect the price.

[2]It should be emphasized that the monopolist does not try to get the largest possible total income, but the largest possible net income or profit. In this respect he is exactly like the man who operates under competition.

[3]These characteristics of cost curves were described and explained in Sec. 14 • 3 to 14 • 5.

the competitive producer, still the monopolist does find that there are forces beyond his control that determine for him some price that will be the most profitable or, if he prefers to set the volume of his output, forces that determine for him what volume will be the most profitable one. The idea that a monopolist is a free agent, who can set his price anywhere he wishes and who can be expected to set as high a price as possible, overlooks the fact that a high price may turn out to be unprofitable.

16 · 7. An Illustration of Monopoly Price Setting. Perhaps we have discussed this matter too long in the abstract. The student should be able to guess from what has just been said that if the demand and the costs are

Table 16·1. Setting of Monoply Price

Price per unit	Quantity consumers will take	Total income	Total expense	Net income
$7	130	$910	$1400	Loss
8	120	960	1300	Loss
9	110	990	1200	Loss
10	100	1000	1100	Loss
11	90	990	1000	Loss
12	80	960	900	$60
13	70	910	800	110
14	60	840	700	140
15	50	750	600	150
16	40	640	500	140
17	30	510	400	110
18	20	360	300	60
19	10	190	200	Loss

known, we can determine what price will be the most profitable for a monopolist. Let us, then, take a definite case where the price and the costs are known and see just what it will pay the monopolist to do under the circumstances.

Suppose that the first two columns of Table 16·1 are the demand schedule for a given product that is produced under monopolistic conditions. Just as in any other case, the Law of Demand is applicable, and at higher prices consumers reduce the volume of their purchases. Of course, under competitive conditions if a producer were to raise the price of his output, he would not be able to sell anything at all, because consumers would patronize his competitors. The monopolist, however, does not have this handicap. He may raise his price if he wishes, merely suffering some reduction in the volume of goods sold. In addition to this demand, let us assume that the

methods of production that this monopolist uses are such that he has fixed costs of $100, regardless of the volume of his business, and variable costs that amount to, let us say, $10 for each unit of output produced. To remind ourselves again what these terms mean, let us say that this monopolist must pay taxes, rents, and other such items amounting to $100 and that these costs will remain exactly the same regardless of the output. He also has to buy raw materials and pay for labor, etc. These latter items vary in total amount as he varies his output. If he increases his volume of production, he must use more raw material and more labor, and we assume for the purposes of our present illustration that these variable costs amount to $10 for every unit of output produced. Now we have been given what information we need to determine what price it will pay the monopolist to charge and what volume of output will be most profitable for him.

From the first two columns of Table 16·1 we can easily compute the total incomes received at each of the various prices that the monopolist might charge. For example, if he charged $10 per unit for his product, the demand schedule tells us that he will sell 100 units, which at $10 per unit will make a total income of $1000. If we multiply each figure in the first column by the corresponding figure in the second column, we get a third column of products showing the total incomes at the various prices.

But we have said that the monopolist does not seek the greatest possible total income or revenue. He wants the greatest possible net income. To find this we must find the costs. If again we illustrate with the case where the monopolist has set his price at $10, we see that under these circumstances he will sell 100 units of output. But his variable cost is $10 per unit of output, which will amount to $1000 total variable cost when his output is 100 units. In addition he will have to add $100 of fixed costs, making his total cost $1100 at this volume. Since his income was but $1000 and his total cost is $1100, he will be assuming a loss of $100 if he operates his business at this selling price. Thus the figures of the fourth column of Table 16·1 are found by multiplying each volume of output in the second column by $10, since the variable expense if $10 per unit of output, and then adding to the product so obtained an amount of $100 to cover fixed expenses. To illustrate once more, if the monopolist sets his price at $19 per unit of product, he will sell but 10 units. His variable expenses will amount to $10 each on these units, or $100 of variable expense. To this he will have to add $100 of fixed expense, making his total cost $200 as shown in the fourth column of the table.

If we have the total incomes in the third column and the total expenses in the fourth column, simple subtraction will yield the net incomes of the last column. Thus, when the price is set at $16, we compute that the total income is $640 and the total expense is $500. No higher mathematics are

necessary to discover that the net income or profit under such circumstances will be the difference, or $140.

When, now, we inspect the figures of the last column of Table 16 • 1, we find that it is by no means a matter of indifference to the monopolist where he sets the price. To be sure, if he wishes, he can set a very high price. He has the power, as a monopolist, to set a price of $19 if he wishes. But if he does so, he will discover that the conditions of demand and of cost, both of which are beyond his control, operate in such a way as to give him a loss. Lower prices will actually be to his advantage. Yet there are limits also beyond which price cutting will not be of advantage in this particular case. If the price is reduced, say, to $8, the monopolist will also lose money. But somewhere in between comes the "happy medium" where profit can be made. In this particular example the cost and demand situations are such that the greatest possible profit will be made if the monopolist sets his price at $15; or, if he wishes to set the volume of his output and let the price take care of itself, he should set his output at 50 units, which amounts to the same thing. The price of $15 per unit of output, or the volume of production of 50 units of output, will yield him the largest possible net income.

16 • 8. Effects of Changing Costs. When we were studying problems of supply, we experimented with various costs to see what their effect was on business profits.[1] We should now try the same sort of experiment in the case of monopoly. If the optimum monopoly price[2] is determined when the costs and demand are given, what will be the effect of changing those costs?

First, since it is simplest, let us ask what will happen if we change the fixed costs in the illustrative case of Table 16 • 1. Suppose that the fixed costs are raised from $100 to $200. There will be no change in the first three columns of the table, since the demand has not been altered. But the cost figures will all be altered. In fact, since the fixed costs have risen from $100 to $200, the total cost will be larger by $100 at every volume of output. Each figure in the fourth column will be larger by $100 than it was before. And, since the expenses are in each case $100 greater, the net income will be in each case $100 less. Each figure in the last column of the table will be $100 lower than before. But, if each figure is reduced by $100, the figure that was largest will still remain largest. If we set our price now at $15, the profit will be only $50 instead of $150; but now $50 will be the best we can do. We thus see at once the first of the principles involved in this illustration; namely, that *changes in the fixed expenses do not change the most profitable*

[1] See Sec. 14 • 8.

[2] Optimum from the standpoint of the monopolist in the sense that he makes the highest possible profit at this price. It may well not be the best price from the standpoint of the general public.

price or volume of output for a monopolist, although the total amount of the profit is changed by exactly the same amount that the fixed expense is changed. It is interesting to note that we come to the same conclusion that we reached before for the competitive case; namely, that fixed costs do not enter into decisions with regard to prices of volumes of business, but they do influence the amount of profit.[1]

Let us now investigate the effects of altering the variable costs in the problem of Table 16 • 1. Let us leave the demand exactly as it is, which means that there will be no changes in the first three columns of the table. Let us leave the fixed cost at $100, as in the original problem, but let us this

Table 16•2. Second Example of Monoply Price Setting

Price per unit	Quantity consumers will take	Total income	Total expense	Net income
$7	130	$910	$1140	Loss
8	120	960	1060	Loss
9	100	990	980	$10
10	100	1000	900	100
11	90	990	820	170
12	80	960	740	220
13	70	910	660	250
14	60	840	580	260
15	50	750	500	250
16	40	640	420	220
17	30	510	340	170
18	20	360	260	100
19	10	190	180	10

time assume that the variable costs are reduced from $10 per unit of output to, say, $8. This will change each figure in the fourth column of the table and it will be best for us to make a new table to see the new results. Table 16 • 2 shows the first three columns exactly like that of Table 16 • 1, but in the fourth column each figure has been found by using a variable expense of $8 per unit and a fixed expense of $100. Thus, for example, if the price is set at $10 so that our volume of output is 100 units, our total variable cost will now be $8 per unit on 100 units, or $800. To this we add the $100 of fixed cost, to get total expenses of $900 as shown in the fourth column. The last column of Table 16 • 2 is found, of course, by subtracting each figure of expense in the fourth column from the corresponding income figure in the third column. In other words, the method of making up this table is

[1]See Sec. 14 • 8.

exactly the same as that used in making up Table 16·1. The only difference is that the variable expense is now $8 instead of $10 and, because of this change, the last two columns of Table 16·2 differ from those of Table 16·1.

Again, however, we note in Table 16·2, as in Table 16·1, that the price that is set is not a matter of indifference. Some prices will yield profits and some will bring losses. This time we see that the largest possible profit will be made by setting the price at $14 per unit and by producing a volume of 60 units.

It should be pointed out, by the way, that neither in Table 16·1 nor now in Table 16·2 did the monopolist select the price that would yield him the greatest possible total income. This would be the price of $10, which yields him a total income of $1000. In either case the monopolist could have selected this price if he had wished. In the first case he would have taken an actual loss by doing so, and in this second case he would have accepted a profit of but $100 in place of the profit of $260 which he makes by setting his price at $14. The raising of his price from $10 to $14 in this second example has cut the gross income from $1000 to $840; but it has cut the expenses from $900 to $580. As the reduction of $160 in gross income is far more than offset by the reduction of $320 in expenses, the monopolist finds it advantageous to set his price at $14 even though he does not thus get the greatest possible total income. *Businessmen, whether monopolists or competitors, are primarily interested in net income rather than in gross income.*

To get back to our illustration, when, in Table 16·1, the variable expense was $10 per unit, it paid the monopolist to set his selling price at $15 per unit and he made a total profit of $150. When the variable expense was reduced to $8 per unit, we find that it pays him to set his price at $14 per unit and his total profit rises to $260. If we were to make up a new table like Tables 16·1 and 16·2, but with a variable expense of $12 per unit,[1] we would find that the highest profit that the monopolist could now make would be a profit of $60, and he would make this by setting his price at $16 per unit of output. He would, of course, prefer the situation of Table 16·2, where he made a profit of $260. But now that the cost situation has changed, he cannot make this profit merely by continuing to produce 60 units of output and selling them for $14 apiece as before. If he does do this, he will still get a total income of $840, but his new expenses will be $820, leaving him a profit of but $20, and he would rather have the profit of $60, which he can get when he sets his price at $16 per unit.

If we assemble the results of our three trials, we get the data of Table 16·3, which show the prices and the outputs at which this monopolist will

[1] The student with a mathematical bent will find this principle discussed somewhat further in Note 7, Mathematical Appendix, p. 894.

maximize his profit at each variable expense, and the amount of profit that he will make. It is easy to see implicit in this table the second of the principles involved in our illustration; namely, that *increases in the variable expense make it profitable for the monopolist to increase his selling price and to reduce the volume of his output, but they reduce the amount of the profit that he makes.*[1]

Perhaps it would be just as well to break into the midst of this chapter, which is to such a considerable extent abstract and theoretical, to point out that the conclusions just reached are of some considerable practical importance—that they can be applied in real life and not merely to hypothetical examples in a textbook. For example, suppose that a state regulatory body, such as a public utilities commission, or even a legislature, is faced with the problem of controlling monopolies. Some member says convincingly that the bigger a monopoly gets the more dangerous it is, and therefore we ought

Table 16·3. Effect of Changes in a Monopolist's Variable Costs

Variable cost per unit	The most profitable		Total profit
	Price	Output	
$8	$14	60	$260
10	15	50	150
12	16	40	60

to tax them more and more as they grow larger and larger. For example, we might lay a tax of $2 per unit of output, and the bigger the firm grows the more tax it will pay. This sounds reasonable, and let us suppose that it is done. If the monopoly so taxed is the one described in Table 16 · 1, which had variable expenses of $10, these variable expenses are now increased by the $2 tax to $12. We have already discovered that, under these circumstances, it will pay the monopolist to raise his price to $16 instead of leaving it at $15, so our tax has raised the price to consumers. These consumers, now that the tax is laid, pay $16 apiece for 40 units which they buy, while they formerly got 50 units and paid but $15 apiece for them. The monopolist's profit is reduced from $150 to $60 (see Table 16 · 3) and the state receives as tax income the $2 tax on each of the 40 units sold, or a tax income of $80.

Let us contrast this with what could have been done by one who understood the principles just enunciated. We know that, if we had increased the monopolist's fixed expense rather than his variable expense, he would not

[1] It would be good practice for the student to make up such a table and to verify the results as given here.

have raised his price. Suppose, then, that instead of levying a sales tax of $2 per unit sold, we lay an annual license fee of $80 on this monopoly. The monopolist will leave his price at $15 per unit, and consumers will still buy 50 units. The monopolist's profit will be reduced by the new fixed tax from $150 to $70. The state will get a tax revenue of $80, just as before. Let us compare the two situations in parallel columns:

If we lay a tax of $2 per unit of output:	If we lay a fixed license fee of $80:
1. Consumers get 40 units of the good.	1. Consumers get 50 units of the good.
2. Consumers pay $16 per unit.	2. Consumers pay $15 per unit.
3. State receives $80 in taxes.	3. State receives $80 in taxes.
4. Monopolist makes profit of $60.	4. Monopolist makes profit of $70.

There can be no question but that the fixed tax is better for everyone concerned, even though the argument that underlay the proposal to lay a variable tax sounded superficially logical. To see the answer to the problem, all that we need is an understanding of the principles that explain the relationship of fixed and of variable expense to monopoly price. The principles sound far removed from real life and practical affairs when the student first runs across them, and when he sees them illustrated with simple arithmetical examples, which were obviously made up out of whole cloth. But it is very important for the student to keep constantly in mind the fact that the principles of economics are descriptions of situations which exist or at least which are approximated in real life. We have just seen, for example, that two different plans for raising public revenue may have entirely different indirect effects even though the amount of revenue raised is the same in both cases. Only one who has understood the general principles involved is likely to realize that the indirect effects are there at all, or to know how they can be used to advantage rather than to disadvantage.

16·9. Relationship of Elasticity to Total Income. In Sec. 16·6 and 16·7 we noted that the most profitable price for a monopolist is determined by his costs and by the elasticity of the demand for his product. In Sec. 16·8 we followed this point somewhat further and found just how changes in the fixed and the variable expense would change the most profitable price. We are still faced with the problem of determining how changes in the elasticity of demand will affect the problem. Before we can go very far in finding an answer to that problem, we must turn aside temporarily to define a few new terms, after which we shall come back to this point and proceed with our study of elasticity of demand and its importance to the monopolist.

Let us first go back to the point made in Sec. 7·4 (see page 90) that the total amount spent by consumers varies according to the elasticity of the demand. We discovered that, when the demand is elastic, rising prices re-

duce total expenditures; when the demand is inelastic, rising prices raise total expenditure; and, when there is unitary elasticity of demand, total expenditure stays constant regardless of price. We illustrated the point at that time with three simple numerical examples in the first paragraph of Sec. 7 • 4. At that time we were interested in consumer problems, and were discussing the factors that determined consumers' expenditures. Now that we have turned to other problems and are for the moment considering the problems of the monopolist, we can put these former facts to good use if we will but remember that consumer expenditures are the same thing as producers' incomes as was pointed out in Sec. 16 • 6. When A buys something from B and pays $10 for it, A sees the $10 as a cost and B sees the same $10 as an income. Just as every sale of goods is at the same time a purchase, depending on how one looks at it, so every sale of goods involves expenditure for one person and income or revenue to another. Since we are now interested in factors affecting the monopolist's income, we may restate the principles of Sec. 7 • 4, putting them in terms of income rather than expenditure: when the demand is elastic, rising prices reduce the monopolist's total income; if the demand is inelastic, rising prices will increase his total income; and if there is unitary elasticity of demand for his products, his income will remain unchanged when prices vary.

These principles have just been stated in terms of the effect of price changes—what happens when we change the price of the commodity. For our purposes, however, we are going to find it useful to state the same principles in terms of the effects of changes in the volume of output: What happens to the monopolist's total income when he alters the volume of his output? Since increases in the volume will lower the price, the effect of raising the volume will be the same as the effect of lowering the price. We can, then, state the effect of changes in volume by reversing the statements at the end of the preceding paragraph, giving us the following general statement which we shall proceed to use at once: *When the demand for a commodity is elastic, increases in the volume put on the market will increase the monopolist's total income; if the demand is inelastic, increased output will reduce the monopolist's total income; and if there is unitary elasticity of demand, the total income will stay constant regardless of output.*

This is really a matter of definition rather than of proof, since it depends on the way in which we defined elasticity of demand. We may illustrate it by reviewing the three simple problems at the beginning of Sec. 7 • 4. There we assumed in the first case that when the price of a commodity was 10 cents, consumers would buy 100 units, but when the price rose to 11 cents, they reduced their purchases to 75 units. We can now reverse the problem and say that, if the monopolist increases his output from 75 to 100 units, the price will drop from 11 to 10 cents, and the total income will rise

from $8.25 to $10. We discovered in Sec. 7 · 4 that this is an elastic demand, and we see that when the output increased the total income rose. In our former example of inelastic demand on page 90, we noted that when the output increased from 95 to 100 units, the price fell from 15 to 10 cents, so the total income dropped from $14.25 to $10 as we increased our output in this case of inelastic demand. Finally, the example of unitary elasticity on page 91 showed that when the output rose from 90 to 120 units, the price fell from $1 to $0.75, so that the total income remained at $90 in both cases. This review will help to explain the meaning of the principles that appear in italics earlier in this paragraph, and the student who has difficulty in following them would do well to reread Sec. 7 · 4 in conjunction with this discussion.

16 · 10. Marginal Revenue. Now that we are discussing the relationship between volume of output and total income we may well raise a question that faces every businessman, whether a monopolist or a competitor; namely, what will be the effect on my income of producing one more unit of output —of expanding my production by just one unit more? We call this concept the marginal revenue, and define it as follows: *The marginal revenue corresponding to any given volume of output is the amount by which the total income would be increased if the output were increased by one unit.* Thus the wheat farmer who is producing 10,000 bushels of wheat and selling them for $1 per bushel is getting a total income or total revenue of $10,000. How much will his total income be increased if he produces and sells one more bushel? Since this is a competitive case, the increase in output will have no effect on price, and the 10,001 bushels will still sell for $1 a bushel or will bring in a total revenue of $10,001. The marginal revenue is the amount by which the total revenue was increased and, since the total revenue increased from $10,000 to $10,001 when the output was increased by 1 bushel, the marginal revenue was $1. In this case the marginal revenue was equal to the price; both were $1. In fact, it is easy enough to see that *under competitive conditions the marginal revenue is always exactly equal to the selling price,* since the added unit can always be sold without lowering the price, and the one added unit will increase the income by whatever it sells for, which will be equal to the selling price.

But let us take the case of a monopolist. If he increases the output, it will reduce the selling price. Suppose, for example, that, when a monopolist increases his output from 500 to 1000 units, the price drops from $1.10 to $1 per unit. What is the marginal revenue? We cannot tell exactly, because we do not know what happened to the total income when we increased the output by one unit. But we see that an increase of 500 units in the output increased our total income from $550 (which we got by selling 500 units at $1.10 each) to $1000 (which we got by selling 1000 units at $1 each). We can find the approximate marginal revenue for this increased output by say-

ing that, if 500 extra units of output increased the income by $450, then one added unit of output would increase our income by $450/500, or by 90 cents. The approximate marginal revenue, then, is 90 cents. To find the marginal revenue corresponding to any particular pair of prices and outputs, we divide the difference in total income by the difference in the quantity of output.

Let us notice now for this monopolistic case the relationship between price and marginal revenue. The prices in our example were $1 and $1.10, but the marginal revenue turned out to be but 90 cents, smaller than either of the prices. In fact, *in all monopolistic cases, marginal revenue is always smaller than the selling price.* The reason is simple. Whenever there is an element of monopoly, increases in output tend to lower the price. Thus, if we increase our output by one unit, we sell that one unit at the price that we can get for it, thus increasing our income by an amount equal to the price; but the addition to our output decreases the price on all the units we sell, thus subtracting something from our income and making the net change smaller than the selling price. This can be illustrated numerically. If, in a particular case, we increase our output from 100 to 101 units, we may reduce the price from $10 to $9.99 per unit. If so, our total income changes from $1000 to $1008.99. Our income increased by $8.99 when we put out one more unit of output, and the marginal revenue was $8.99. Where did this come from? The answer is that we sold one unit that we did not sell before, and got an added income from it of $9.99, which was the price at which we sold it. But in order to do this, we had to accept a reduction in price from $10 to $9.99 on the 100 units which could have been sold at the higher price in case we had not increased out output. These 100 units brought 1 cent less each, or reduced the income by $1. When we throw the two together we find

Increased income from added unit of output	$ 9.99, price
Lost income from reduction of price on original 100 units of output	−1.00
Net change in income, or marginal revenue	$ 8.99

If the increased output has any effect at all on the selling price, it will be to lower it; so if marginal revenue is not exactly equal to selling price, it will always be smaller than the selling price. *In cases of complete competition, the marginal revenue and the selling price are equal; in all other cases, the marginal revenue is smaller than the selling price.*

In fact, the marginal revenue may even be negative. Thus suppose that, when we increase our output from 990 to 1000 units, the price drops from $1.50 to $1. We can summarize these changes in tabular form thus:

Quantity	Price	Total income
990	\$1.50	\$1485
1000	1.00	1000

Increasing the output by 10 units *reduced* the total income by \$485, so we get a negative marginal revenue, as follows:

$$\text{Marginal revenue} = \frac{\text{change in total income}}{\text{change in volume of output}} = -\$485 \div 10 = -\$48.50$$

Of course, we could also have cases where the marginal revenue was equal to zero. In fact, this will obviously be true whenever there is unitary elasticity of demand; for, in these cases, the total income will stay fixed when we vary out output and, since the change in income is zero, the marginal revenue will be zero. We can illustrate this numerically by assuming that, when we produce an output of 800 units, we sell it for \$12 per unit; but if we increase the output to 1200 units, we sell them for but \$8 each. In either case the total income is \$9600, so our summary would be as follows:

Quantity	Price	Total income
800	\$12	\$9600
1200	8	9600

Our formula would give us,

$$\text{Marginal revenue} = \frac{\text{change in total income}}{\text{change in volume of output}} = \$0 \div 400 = 0$$

If now we look back to the principles in italics toward the end of Sec. 16·9, we shall see that the marginal revenue is so connected with the elasticity of demand that, *When the demand is elastic, the marginal revenue is always positive; when the demand is inelastic, the marginal revenue must be negative; and when there is unitary elasticity of demand, the marginal revenue is always equal to zero.* In fact, the relationship between marginal revenue and elasticity of demand is such that we can easily compute one from the other.[1] All that we need to know is that, if η represents the elasticity of demand, then

$$\text{Marginal revenue} = P\left(1 - \frac{1}{\eta}\right)$$

[1] For proof of this relationship between elasticity and marginal revenue, see Mathematical Appendix, Note 8, p. 896.

where P is the average of the prices given in the problem. This formula is so simple, useful, and meaningful that it will pay us to work out one problem with it. Let us assume that, when our output is 500 units, we can sell them at a price of \$1.10 but, if we increase the output to 1000 units, the price will fall to \$1. We can summarize the situation in the following table:

Quantity	Price	Total income
500	\$1.10	\$ 550
1000	1 00	1000

If we use the methods of Sec. 7·2, we find that the elasticity of demand in this case is $\eta = 7$. If we use the method that we have just been using for finding the marginal revenue, we would note that the total income increases by \$450 when we increase the output by 500 units, so the marginal revenue is \$450÷500, or 90 cents. But let us use our new formula, which shows the relationship between marginal revenue and elasticity. The two prices in our problem are \$1.10 and \$1, so the average price is \$1.05. Since the elasticity of demand is 7, we can substitute these values directly in our formula to find that

$$\text{Marginal revenue} = \$1.05\left(1 - \frac{1}{7}\right) = \$0.90$$

The formula yields us at once the same result that we got from our longer calculations. If the elasticity of demand is known at any price, we can find the marginal revenue at once; if the marginal revenue is known, we can find the elasticity.

16·11. Relationship of Elasticity of Demand and Marginal Revenue. The formula of the preceding paragraph is helpful in pointing out the relationship between marginal revenue and elasticity. Remembering that

$$\text{Marginal revenue} = P\left(1 - \frac{1}{\eta}\right)$$

always we can see what will happen as η takes various values. Evidently whenever η is greater than 1 (which means that the demand is elastic), the value in parenthesis will be positive and the marginal revenue will be positive. But, if η is a fraction smaller than 1 (if, for example, we have an inelastic demand with $\eta = 0.5$), then the value in parenthesis will become $(1 - 1/0.5)$ or $(1 - 2)$, or -1. Whenever the elasticity of demand is a fraction smaller than unity (that is, whenever the demand is inelastic), the

marginal revenue will be negative. If we have unitary elasticity of demand, so that $\eta = 1$, the value in parenthesis becomes $(1 - 1/1)$ or zero, and the marginal revenue is zero. We can sum up these general statements in convenient tabular form as follows:

Nature of demand	Elasticity of demand	Marginal revenue
Elastic	$\eta > 1$	Positive
Inelastic	$\eta < 1$	Negative
Unit elasticity	$\eta = 1$	Zero

16·12. Marginal Revenue and Amount of Profit. When a businessman increases his output, there is some resulting change in income and in cost. We have called the increase in cost that arises when we expand our output one unit the marginal cost (see Sec. 14·4), and we have called the increase in total income that arises when we expand our output one unit the marginal revenue (see Sec. 16·10). If we increase our output slightly and, as a result, our income is increased more than our expenses are increased, the expansion in output will add to our profits. If we translate this into more technical terms, we can say that, whenever marginal revenue exceeds marginal cost, it pays to expand output. But if added output increases our costs more than it increases our income, we lose money by expanding and it pays us not to produce the added units. Again we can translate this into the statement that, whenever marginal cost exceeds marginal revenue, it pays us to contract output. And, of course, if the addition of a unit to our output increases our income but increases our cost by the same amount, there is neither gain nor loss by changing the output. We may as well stay right where we are. *Every businessman finds it advantageous to expand his business when marginal revenue exceeds marginal cost, and to contract his business when marginal cost exceeds marginal revenue, reaching finally the equilibrium point where marginal revenue and marginal cost are equal. Business profits are maximized at the output where marginal revenue equals marginal cost.*[1] Since under competitive conditions the elasticity of the demand for a business man's product is infinitely great, our equation, marginal revenue $= P\ (1 - 1/\eta)$, indicates that under such conditions the marginal revenue will be equal to the price, which is a fact noted earlier on the basis of other considerations. If marginal revenue equals selling price, as it does under competition, then when we say that maximum profits are made by making marginal revenue equal to marginal cost, we are saying the same thing that we said toward the end of Sec. 14·6 when we said that "if the

[1] For proof of this statement, see Note 9, Mathematical Appendix, p. 898.

selling price is fixed, a businessman always finds it profitable to expand his output until the marginal cost equals the selling price." Under competitive conditions, we may say either that marginal revenue equals marginal cost or that selling price equals marginal cost at the point of highest profit; but since under monopolistic conditions marginal revenue is always smaller than the selling price, we may make only the former of these statements: *In every business, it pays to expand until marginal revenue equals marginal cost. In competitive businesses, it pays to expand until marginal revenue, marginal cost, and selling price are all equal.*

One further principle to be drawn from the foregoing relationships is this: Since the monopolist always selects an output where marginal cost equals marginal revenue and since marginal cost is always positive, because additions to output will always increase the total cost, it follows that the monopolist always selects an output where the marginal revenue is positive. But marginal revenue is positive, as we have seen, only if the demand is elastic (see Sec. 16 · 10), so we now discover that *the monopolist always contracts his output to some point where the demand for his product is elastic.* If the demand is inelastic, the marginal revenue, being negative, will be smaller than the marginal cost, and it will pay the monopolist to contract, and to keep on contracting until he is either selling the smallest single indivisible unit that he can sell, or until the demand has become elastic with falling output and rising prices. This brings us back, as we promised in the first paragraph of Sec. 16 · 9, to the question of the relationship of elasticity of demand to the setting of monopoly price.

The principles that we have just discussed can be applied to those unusual cases in which a monopoly operates without any variable costs of production. Let us imagine, for example, that someone owns a fine spring yielding mineral waters of unique curative power. The water comes forth from the earth, and the monopolist allows people to come in to drink it on the spot on payment of a fee. (If the water were bottled, there would be variable costs for bottles, and we are trying to imagine a case where there would be no variable costs at all.) Since there are no variable costs involved in the business, there are no marginal costs; the marginal cost is zero. If one more customer drinks, our costs are not increased at all. Since our marginal costs are zero, we will find it most profitable to set our price or our volume at the point where the marginal revenue is zero, which means that there is unit elasticity of demand. Under these circumstances, and only under these circumstances, will it pay the monopolist to charge the price that will yield the largest possible gross income. Where variable costs enter into the picture, he will contract his output still further, until the demand is elastic and the marginal revenue is positive. Then his gross income will be below its maximum, but his net income will be maximized.

16 • 13. Changes in Elasticity of Demand at Different Parts of the Demand Curve. We are now in a position to understand a fact that should have been taken up in Chapter 7 had we had the background for understanding it then. If we vary the price of any commodity over a very wide range, we usually find that the total income is small when prices are very low, and small again when prices are very high, but that there is a peak of total income somewhere in the middle ground, as in Fig. 16 • 1. For example, if the price of potatoes is zero (that is, if potatoes are free), the total income from them is zero. At the other extreme, if potatoes cost a million dollars an ounce the chances are that no one would use them, and the income from them would be zero. At very high and at very low prices there is no income at all. If we raise the price from zero to 1/10 cent a bushel, consumers will want to buy a good many potatoes at this low price, but the price will be so low that the total income will probably be small. As the price rises, it is likely that the total income will rise for a time, reach a peak, and then fall again.

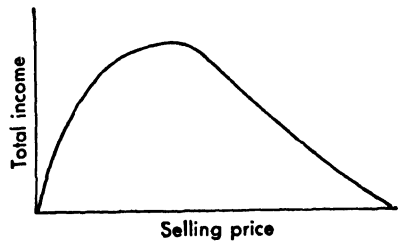


FIG. 16 • 1.—Changes in total income with changes in the selling price.

Since the selling price varies inversely with the quantity of output, we could restate this idea by saying that, if the volume of output is varied over a wide range, we shall find that the total income rises for a time, passes through a maximum, and then falls again. If only 10 bushels of potatoes were raised, they might sell for a great deal per bushel, but the total income would almost certainly be lower than for an ordinary crop. As the size of the crop increased, the total income would increase for a time, finally, we would get to the point where potatoes were almost free goods, and then to the point where they were actually free goods, and the price would drop to zero, as would the income. The relationship between the volume of output and the total income is shown in Fig. 16 • 2.

When increased output increases the total income, the demand is elastic and the marginal revenue is positive; when increased output decreases the total income, the demand is inelastic and the marginal revenue is negative; and when changes in output do not change the total income, there is unitary elasticity of demand and the marginal revenue is equal to zero. Since it is reasonable to suppose that every commodity will exhibit the characteristics

of Fig. 16·1 and 16·2, it is also reasonable to assume that *every demand curve is elastic at low volumes of output and high prices; that it is inelastic at high volumes of output and low prices; and that somewhere in between it passes through a point of unitary elasticity.*[1] It should be emphasized that this is true of all demand curves, not merely for demands for monopolized products. The principle is included in a chapter on monopoly merely because it is here that the necessary background information first became available, but the principle applies to all demand curves.

We saw in the preceding section that, in an industry operating without variable costs, the most profitable output is that which maximizes total income, or the output where marginal revenue is zero. This would be the output corresponding to the highest point on the curve in Fig. 16·2. When there are variable costs involved in production, as there usually are, the most profitable output will always be one that yields a positive marginal revenue, or one where the demand is elastic. In Fig. 16·2 we see that this

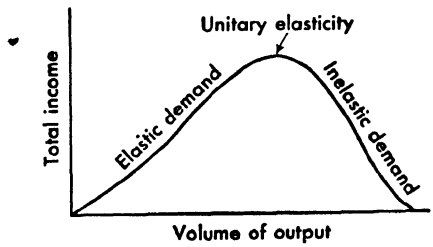


FIG. 16·2.—Changes in total income with changes in the volume of output.

will be a smaller volume of output than would be most profitable in the absence of variable costs. *The effect of variable costs is always to reduce output,* and the introduction of new variable costs always makes it profitable to reduce the output still more. Fixed costs, if really fixed, do not have this effect, since we would have to pay them anyway even if we did no business. Since marginal cost is the increase in total cost accompanying a small increase in output and since fixed costs do not increase at all when there is a small increase in output, it is evident that fixed costs do not contribute to marginal costs and hence play no part in price setting.

16·14. Monopoly Discrimination and Class Prices. We noted in Sec. 15·7 what we called the First Law of the Market, which states that under competitive conditions there will be one price and only one for any given commodity on a given market at a given time. Such is by no means the case, however, with monopoly. The monopolist finds it to his advantage to discriminate among his customers, charging different customers different

[1] For comment on this point, see Note 10 of the Mathematical Appendix on p. 901.

prices for the same article. He makes the greatest possible profit, not by charging each buyer the same amount, but by charging each buyer the particular price that will bring in the greatest possible net income from that particular buyer. Thus it may well pay him to charge John Jones $5 and charge Phil Taylor $2 for the same commodity. Under competitive conditions John Jones, finding that he was being charged a high price, would refuse to buy and would shift to some competing seller; but under monopoly conditions there are no competing sellers. If Jones and Taylor have different demands for the product, it will surely pay the monopolist to discriminate and charge them different prices. A monopoly that follows this practice is called a "discriminating monopoly."

We can illustrate this situation by taking a simple numerical example similar to those which we have used before. Let us assume that a monopolist tries setting various prices for his product and finds that the quantity sold responds as in Table 16·4.

Table 16·4. Demand for a Monopolized Product

Price	Quantity sold	Price	Quantity sold
$1	560	$5	240
2	480	6	160
3	400	7	80
4	320	8	0

If his fixed expense is $100 and his variable expense is $2 per unit of output, the monopolist will make the greatest possible profit by charging a price of $5 if he charges a single price to all buyers.[1] But let us assume that the purchases on Table 16·4 represent purchases by two buyers, who have somewhat different demands for the product. Thus, for example, the quantities bought by each buyer at each price are shown in Table 16·5, where it will be seen at once that the total amount bought by both buyers is the same as that shown in Table 16·4. Assuming that the cost situation is unchanged ($100 fixed cost and variable costs of $2 per unit of output), the monopolist will find it to his advantage to charge Mr. A a price of $6 and Mr. B a price of $4, instead of charging them $5 each. If he charges them $5 each, his total net income from both of them will be $620; if he charges discriminating prices, he can make a profit of $640 on Mr. A and of $160 on Mr. B before deducting his fixed costs, or a total profit on both men of $700 after deducting

[1] This result is discovered by the same numerical process that we have applied in earlier monopoly examples. It would be good practice for the student to carry out the computations and check the result.

the fixed cost. Discrimination raised his total profit from $620 to $700. Since the two men had different demands, it paid to discriminate in prices.[1] In fact, it can be demonstrated that *it always pays a monopolist to discriminate between buyers if their demands differ in elasticity, and always to charge the higher price to the buyer whose demand is less elastic.*

Although the student may not think of illustrations immediately, actually such discrimination is common where there is any element of monopoly power in the situation. Steamship companies charge different prices for different classes of service, physicians ordinarily charge very different fees to patients with different incomes, power companies usually charge quite different rates per kilowatt-hour for electricity sold to different classes of users, etc. In fact, discrimination among customers is generally taken to be one of the typical and characteristic signs of the existence of monopoly.

Table 16·5. Individual Demands for a Monopolized Product (Based on Table 16·4)

Price	Quantity bought by		Total
	Mr. A	Mr. B	
$1	360	200	560
2	320	160	480
3	280	120	400
4	240	80	320
5	200	40	240
6	160	0	160

16 · 15. Summary. Although complete monopoly never actually exists, we consider that there is some element of monopoly whenever an individual producer can in some degree influence the price of the commodity that he sells. Instead of finding that the elasticity of the demand for his output is infinitely great, he finds that it is finite, and the smaller the elasticity of demand the greater is his monopoly power. Because such power may be used to the advantage of the monopolist, yet to the disadvantage of the buying public, men have long feared monopolies and have tried either to eradicate them or to control them. Some monopolies are not eradicable, and some are considered beneficial. In these cases the tendency has been for public control or even public ownership.

The student's first superficial assumption that the monopolist's power to set prices means that he can raise the price without limit overlooks the fact that monopoly profits are determined by the nature of his costs and by

[1] For further discussion of discrimination, see the Mathematical Appendix, Note 11, p. 902.

the elasticity of the demand for his output. If the demand and the costs are given, there will be some one selling price and volume of output that will yield a larger net income than any other. If the demand or the costs are changed, the most profitable price and volume of output may also be changed. Changes in the fixed costs change the amount of the monopolist's profit, but do not change his selling price nor his volume of output. Changes in the variable cost, however, change all three, with increases in the variable cost tending to raise the monopolist's selling price, reduce the volume of his output, and reduce the amount of his profit.

The elasticity of the demand for his product is of importance to the monopolist because it determines the effect on his total income of changes in the volume of his output. If the demand is elastic, the marginal revenue will be positive and increased output will increase income. If the demand is inelastic, the marginal revenue will be negative and increased output will decrease income. Unit elasticity of demand corresponds to a marginal revenue of zero, and in this case changes in output have no effect on income. But total income is not the matter of final importance to the businessman, whose decisions are made in terms of net income.

Under conditions of complete competition, marginal revenue is exactly equal to the selling price; if there is any element of monopoly in the situation, the marginal revenue is always less than the selling price. In any business, competitive or monopolistic, the highest profit will be made when the output is adjusted until the marginal revenue and the marginal cost are equal. In competitive businesses, they will both be equal to the selling price but, since in monopolistic businesses the selling price exceeds the marginal revenue, in monopolistic cases they will both be smaller than the selling price. Because of the fact that profits are maximized when marginal revenue equals marginal cost and because marginal costs are always positive, the monopolist always contracts his output until the marginal revenue is positive, which means that the demand is elastic; or until he has cut down his business to the point where he is selling just one unit of output if the demand stays inelastic down to that point. If a monopolist could operate without variable costs, it would pay him to select that volume of output which yielded the maximum gross income but, if he has any variable costs, they will always operate to reduce his most profitable output below this point. The addition of variable costs always tends to restrict output.

If the market for a product can easily be divided into segments, and if the elasticity of the demand for the product in different segments of the market differs, it pays the monopolist to charge different prices in the different segments of the market and the most profitable prices in the various segments will run in the reverse order of the elasticities of the demands;

that is, the higher the elasticity of the demand, the lower the price that it will be profitable to charge.

Questions

1. If a man really had a complete monopoly in the theoretical sense, would the monopolized commodity be a necessity or a luxury? Why?
2. The elasticity of the demand for a manufacturer's product is equal to 5. Does he operate under conditions of complete competition, complete monopoly, or at some stage in between? How do you know?
3. Are there in practice any monopolies if we use the legal rather than the theoretical definition of the term?
4. What would be the situation if we had a complete monopsony?
5. List several actual monopolies known to you.
6. Classify each of the monopolies named in answer to the preceding question according to the classification of Sec. 16 • 5.
7. How do secret processes create lost arts?
8. How could a protective tariff tend to aid the formation of monopoly?
9. Name any monopolies, other than anthracite coal, that arise from limitations in the natural sources of supply.
10. Explain why we do not have a supply schedule for a monopolized product in the same sense that we do under competition.
11. "Competitive price is set by demand and supply; monopoly price by demand and cost." Would you agree?
12. "If the demand for a commodity is inelastic, the manufacturer who increases his output will reduce his total income. Since the increased output will certainly cost at least as much as and probably more than the smaller output, the reduction in total income will also be a reduction in net income. Therefore the manufacturer would never find it advantageous to increase his output if the demand for his product was inelastic." Do you agree?
13. It seems fairly likely that the demand for wheat is inelastic. If this is so, how does it happen that individual farmers ever increase the volume of their output?
14. We have learned that a monopolist always selects a volume of output at such a point that the demand for his product is elastic. Is this also true of the man who produces under competition? Comment.
15. Explain why it is that the monopolist whose business is described in Table 16 • 1 on page 301 prefers to set his price at $15 and take a total income of $750 rather than set his price at $10 and take a total income of $1000.

16. The accompanying table shows the demand for a monopolized product. The variable expense amounts to \$16 for each unit of output, while fixed expenses amount to \$200. What will be the volume of output that will yield the largest possible net income? What will be the corresponding selling price? What will be the amount of the profit?

Demand for a Monopolized Product

Price per unit	Quantity consumers will take	Price per unit	Quantity consumers will take
\$15	300	\$22	160
16	280	23	140
17	260	24	120
18	240	25	100
19	220	26	80
20	200	27	60
21	180	28	40

17. If a tax of \$4 per unit of output is laid on this monopolist, what will happen to his selling price? To the volume of his output?

18. The student who has done the work of the Mathematical Appendix (other students may skip this question) should use the methods there expounded to test the elasticity of demand and the marginal revenue at the various points in the table on page 317. Is the most profitable price at a point where the demand is elastic or inelastic? Where the marginal revenue is positive or negative?

19. Does the monopolist of question 16 set his price in such a way as to get the greatest possible total income? At what price would he get the greatest possible total income?

20. If the monopolist of question 16 finds that his property taxes have been increased by \$200 a year, what effect will it have on his selling price? On the amount of his profit?

21. If the price rises for the raw material used by a monopolist in manufacturing his product, what will be the effect on his selling price? Why?

22. A man sells 1000 units of a commodity at a price of \$16 per unit. The marginal revenue is \$8. Is the situation competitive or monopolistic? Why?

23. Given the data of question 22, what is the elasticity of the demand for the product? *Hint:* The elasticity can be found easily from the relation between elasticity and marginal revenue.

24. A monopolist finds that he can sell 500 units of a commodity for \$7 per unit but, if he raises the volume of output to 550 units, the price falls to \$6 per unit. What is the marginal revenue? Be careful of signs!

25. What is the elasticity of the demand in question 24? Is the answer consistent with

the rules that you know concerning the relationship of elasticity and marginal revenue?

26. State the rules concerning the relationship of elasticity and marginal revenue.

27. A man finds that when he sells 500 units of a product the price is \$7, but when he raises his output to 550 units the price is still \$7. What is the marginal revenue? What conclusions do you draw about the degree of monopoly in the industry? Why?

28. Compute the elasticity of demand for the case in question 27, using the usual elasticity formula. What do you conclude?

29. Explain briefly and in nonmathematical terms why profits are at a maximum when marginal revenue equals marginal cost.

30. In a given firm the marginal revenue is \$25 and the marginal cost is \$1. Explain exactly why it will pay this firm to expand.

31. In a given firm, the marginal revenue is negative. Explain exactly why it will pay this firm to reduce the volume of its output.

32. What is the elasticity of demand when the marginal revenue is negative? From your answer to question 31, can you lay down any rule as to the sensible treatment of volume of output when demand is inelastic?

33. We have said that, if a monopoly has no *variable* costs, it will pay the management to select the output that maximizes gross income. Suppose that such a firm has no variable costs but does have *fixed* costs. Will the rule still hold? Why or why not?

34. In the problem of Table 16·5 it was found that the monopolist would make a net income of \$700 if he charged Mr. A a price of \$6 and Mr. B a price of \$4. How much net profit would he make if he charged both men a uniform price of \$6? A uniform price of \$4?

35. A monopolized product is sold in two markets, A and B. The elasticity of demand in market A is 5 and in market B is 4. In which market is the price the higher?

36. Those who have read Note 11 of the Mathematical Appendix should be able to tell in the preceding question how many times as high the price will be in the high-priced as in the low-priced market.

37. The last paragraph of Sec. 16·14 seems to imply that physicians have at least some element of monopoly power. Do you think that this is true? If physicians actually competed for business, could they discriminate among customers and charge wealthy patients higher prices than they charged poor people?

38. Using the data of question 18 at the end of Chap. 7, compute the marginal revenue in the cotton industry when the price of cotton is 20 cents per pound. Use the formula in Sec. 16·10. Explain carefully the meaning of your answer.

39. Using the data of question 19 at the end of Chap. 7, compute the marginal revenue

in the wheat industry when the price of wheat is 90 cents per bushel. Use the formula in Sec. 16 • 10. Explain carefully the meaning of your answer.

40. Supposing the data of the two preceding questions to be correct, if you were told that you could have a complete monopoly of either wheat production or cotton production, which would you choose and why? Assume that you are motivated by selfishness only.

17. THE RELATION OF PRICE TO COST; NORMAL PRICE

17 · 1. Review. In our earlier study we have already run across two ways in which the price of commodities is related to their cost of production. First, we saw[1] that in cases of competition businessmen find it profitable to select the volume of output that makes the marginal cost equal to the selling price. Then we saw[2] that in any business, competitive or otherwise, the most profitable output is that which makes the marginal revenue and the marginal cost equal. But we discovered that marginal revenue, which is equal to selling price in competitive cases, is always smaller than the selling price in noncompetitive cases. Therefore, if marginal cost is made equal to marginal revenue, marginal cost will be equal to selling price in competitive cases, and lower than the selling price in noncompetitive cases. The businessman may not always succeed in finding the point of highest profit and in adjusting his output accurately at the most profitable point, but insofar as he is able to do so, we can see that there are two statements that we can already make about the relationship of the prices of commodities to their cost:

1. In competitive situations, the price of a commodity tends at any time to equal the marginal cost of the commodity.

2. In noncompetitive situations, the price of a commodity tends always to be somewhat higher than the marginal cost of the commodity.

17 · 2. Normal Price. Let us ask, however, not what the price of the commodity is at any moment, but what price we can expect in the long run. The price at the moment—the actual market price—is, as we have seen, subject to the vagaries of the momentary supply and demand. A sudden and unexpected frost, through its effect on supply, may alter the price of wheat for the time being until the market has had time to readjust itself. But is there no "typical" price of wheat—no "expected," "correct," or "proper" price of wheat? If you were told that the price of wheat was $800 a bushel, would

[1] In Sec. 14 · 6.

[2] In Sec. 16 · 12.

it not affect you differently than to be told that the price of an automobile was $800? Do you not have the idea that automobiles "ought" to cost more than a bushel of wheat, that such a relationship between wheat and automobiles is to be expected? Would you not think that a price of $800 a bushel was unreasonably high for wheat?

These questions take us into a new field of price relationships, away from a study of market prices to a study of what economists call "normal" prices. Yet when they use the term "normal price," they are not thinking of the price which is good or proper in any moral sense. They realize that the price of wheat might conceivably rise to $10 a bushel at some given time if the climatic conditions were such as to yield an extremely short crop. But they also realize that a price of $10 could not be maintained very long, because it would stimulate a tremendous output of wheat. At such an unusually high price it would be profitable to grow wheat on hundreds of acres of land ordinarily used for other crops, and the high price would bring forth unheard-of quantities. Yet at the high prices consumers would be be tempted to turn to other grains, and the consumption of wheat would drop. The high price would stimulate a rate of production far above the rate of consumption, and consequently the price could not be maintained. If, on the other hand, the price of automobiles were reduced by some chance to $10 apiece, hundreds of people would want to buy automobiles who had not had them before, and many families that did have one would buy a second or a third. The rate of consumption would be greatly increased. Yet this low price would discourage producers, who would be losing money, and they would reduce their output. A price of $10 for automobiles would stimulate a rate of consumption far above the rate of production, and the price could not be maintained.

We say that *the normal price of any commodity is the price which, if maintained for an indefinite period, would stimulate a rate of production exactly equal to the rate of consumption.* If the economist says that the normal price of eggs is 60 cents a dozen, he does not mean that this is a desirable price nor that he likes it; he means merely that, if the price were set at 60 cents a dozen and left there, the number of eggs produced would, in the long run, equal the number consumed.

17 · 3. Average Variable Cost. In order to determine what forces influence the normal price under competitive conditions, we shall return to the cost curves used in Chap. 14.[1] Let us assume that the curves in Fig. 17 · 1 are the cost curves of a particular competitive business firm. Shown in the diagram are the marginal cost and the average cost curves with which we are familiar, and a third curve labeled "average variable cost." The meaning of the latter curve can be made apparent from a simple example. Look back

[1] See Sec. 14 · 4 to 14 · 7.

for the moment to Table 14 • 2. We see in that table that when the producer hired 6 men and got 570 bushels of output, his total cost was $400. This was composed of two parts: a fixed cost of $100 and variable costs (wages in that example) of $300, being the wages of 6 men at $50 each. We found the average cost by dividing the total cost by the volume of output—by dividing $400 by 570. This gave us an average cost of 70 cents. If we had been interested in the variable cost alone, we would have noted that the entire variable cost was $300, and the average variable cost was $300÷570, or 53 cents. Since

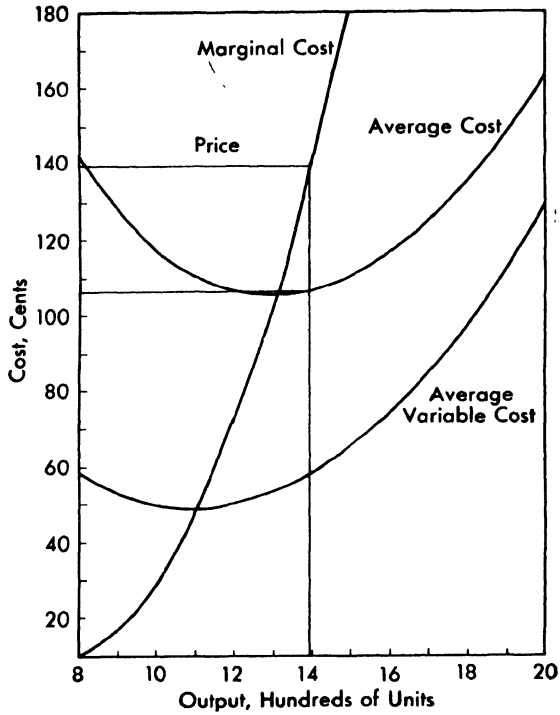


FIG. 17 • 1.—Cost curves of a hypothetical business firm.

the total cost is the sum of the variable and the fixed cost, the variable cost will always be lower than the total cost, and the average variable cost will always be lower than the average cost. The average variable cost is always found by dividing the variable cost by the amount of the output. Taking one more place in Table 14 • 2 for illustration, when 8 men are hired, the output is 720 units, the total cost is $500, and the variable cost is $400. (The difference is the fixed cost of $100.) The average cost is $500÷720, or $0.694, while the average variable cost is $400÷720, or 56 cents. On the basis of the same reasoning by which we showed in Sec. 14 • 4 and 14 • 5 that the average cost curve is U shaped and cut at its lowest point by the marginal cost curve,

we can show that the average variable cost curve is also U shaped and cut at its lowest point by the marginal cost curve [1] as shown in Fig. 17 • 1.

17 • 4. Normal Price under Competitive Conditions. When we were considering the setting of competitive prices in Sec. 14 • 7, we found that it was possible to use a diagram such as that of Fig. 17 • 1 to find the most profitable output corresponding to any given selling price. For example, supposing that the price of the commodity produced by the firm whose cost curves are shown in Fig. 17 • 1 is $1.40 per unit, the firm will expand its output until its marginal cost is $1.40 per unit. Although one cannot tell by inspection of the diagram at exactly what output the marginal cost will be $1.40, it appears that the desired output would be approximately 1390 units. The average cost when the output is 1390 units appears to be approximately $1.06. Therefore the firm will sell 1390 units at $1.40, getting a total income of $1946; the 1390 units will cost $1.06 apiece, so the total cost will be $1473.40; and the profit will be the difference between $1946 and $1473.40, or $572.60. This will be a larger profit than the firm can make by selecting any other output as long as its cost curves remain as they are and as long as the selling price remains at $1.40.[2]

Can the selling price stay at $1.40? Is this a normal price, or will it stimulate a rate of production that differs from the rate of consumption? We have seen that this firm was able to pay all its fixed costs and all its variable costs and still have roughly $570 of clear profit left. Everyone likes a chance to pick up $570, including this firm's competitors. If they are able to do so, they, too, will start producing to get such a profit.[3] And of course, if many other firms are able to produce at the same cost and see the promise of the same profit, their added production will increase the supply and reduce the price. If other firms are able to do the same thing, the price of $1.40 is not a stable price—not a normal price. It is higher than can be maintained and will stimulate a rate of output too large to be absorbed at that price. If, on the other hand, other firms are unable to produce at as low costs as can the firm depicted in Fig. 17 • 1, it must be that this firm has some unusual advantage—unusually good land, exclusive patent rights, a manager of unusual skill, or something of the kind. If this is the fact, the source of advantage has value and should be recompensed accordingly. If the firm has unusually good land, it should be charging enough for it to pay for its value. If the management is unusually skillful and able to get

[1] For proof see Note 12, Mathematical Appendix, p. 903.

[2] These figures are not exact but have been read from the diagram by inspection. They do, however, illustrate the principles involved, just as the same principles were illustrated in Sec. 14 • 7.

[3] Of course, other firms will be attracted by the profit only if they know about it; but it must be remembered that the concept of competition involves complete knowledge on the part of both buyers and sellers.

more output than the ordinary firm, then the wages of the management should be appraised at their proper value. If this is done it will, of course, raise our costs. We will pay more for management, for land, or for patents to cover their true value, and the average cost curve will rise accordingly until its lowest point, where it crosses the marginal cost curve, is at the price of \$1.40. This we see in Fig. 17 • 2. The firm is still producing 1390 units and selling them at a price of \$1.40, but now that it is charging the proper

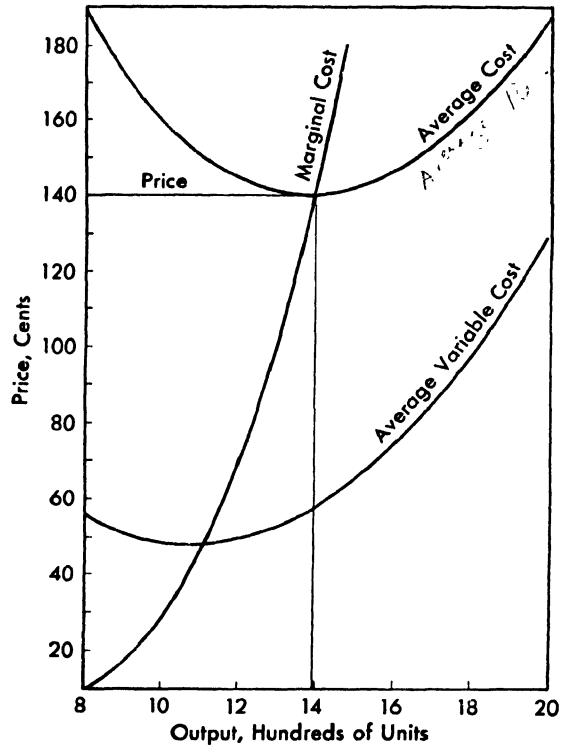


FIG. 17 • 2.—Price and cost conditions for long-run competitive equilibrium.

costs it has an average cost of \$1.40 also. Its total income is \$1946 and its total expense is \$1946, so that it is making no profit at all.

Will a firm stay in business when it makes no profit? What will the owner live on? Do not forget that we have paid all the costs, both the fixed and the variable expenses. In fact, we have paid \$1946 to various people, some for their raw materials, some for their labor, some for their management, and some for the use of their capital. Everyone has been paid and, of course, when we stop to think about it, we realize that in the long run everything that the firm takes in is paid out to someone—to the workers, the

landowners, the government in taxes, the power companies, the sellers of materials, the stockholders, the bondholders, the bankers, or the management. It all goes somewhere in time. That is the peculiarity of long-run equilibrium—the total income must be distributed, all of it, to various people. In our case we found originally that some of it was not being distributed. We paid the various people who had claims and about $570 was left over. Of course, someone has a claim to it. If the firm can make $570 year after year, it is because of contributions that someone is making to the productive process. When we evaluate that person's contribution correctly, there is nothing "left over." Profit, as we have here defined it, is always a temporary, chance thing. *In the long run, there is no profit.* The income is all paid out to various claimants, and from the standpoint of the business firm these payments are costs. In the long run all income goes out to meet various sorts of costs, one of which, of course, will be wages of the managers.

17 · 5. **Price below Lowest Average Cost.** Let us notice, however, how the problem would have been altered if we had set our first price somewhat below $1.40. Suppose, for example, that in our original problem the price had been set at $1 per unit of output. Figure 17 · 3 illustrates the situation, with the cost curves all being identical with those of Fig. 17 · 1, but with the price set at $1. As nearly as one can estimate by eye from Fig. 17 · 3, the output must be set at 1290 units to make the marginal cost equal to $1. We will sell 1290 units at $1 each and get a total income of $1290. But the average cost with this volume of output appears from the diagram to be about $1.05. Our total cost will be 1290 units at $1.05 apiece or $1354.50. If the price is as low as $1, we shall lose the difference between $1354.50 and $1290, or approximately $65 of loss. This is the very best our firm can do at this price. If it selects any other volume of output, it will lose even more. If the costs stay as they are and if the selling price stays at $1, the firm is certain to lose at least $65 if it stays in business.

Under these circumstances will it pay the firm to stay in business, or should it shut up shop? That, of course, depends on the outlook for the future. A business firm does not close down the first day it makes a loss, nor the first week, nor even necessarily the first year. Many a farmer, storekeeper, or manufacturer has continued hopefully and persistently in business for several years through periods of depression in which he was taking losses, counting on a change in the general situation to rescue him before his continued losses finally exhausted his resources and sent him into bankruptcy. Let us see what it will pay the firm to do in the case depicted in Fig. 17 · 3.

We have already computed that the total expense of this firm will be $1354.50. How will this be divided between fixed and variable expenses? To answer this question we turn for the first time to the curve of average

variable costs. We see, as nearly as we can read the diagram of Fig. 17 • 3, that the average variable cost when the output is 1290 units will be about 53 cents per unit of output. The entire variable cost will be 53 cents apiece on the 1290 units of output, or a total variable cost of $683.70. If the total cost is $1354.50 and the variable cost makes up but $683.70 of it, the fixed cost must make up the remaining $670.80. In other words, there are approximately $670 of fixed expenses to be paid anyway, whether we put out the 1290 units or not. We shall owe interest on bonds, taxes, etc., of $670, which must

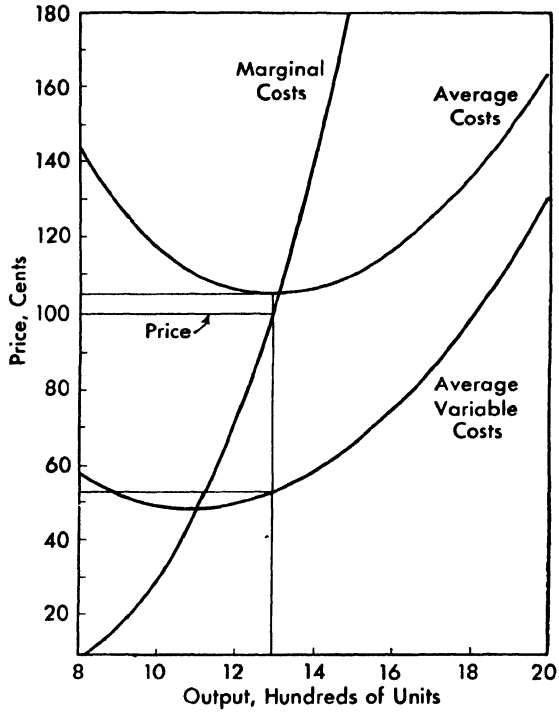


FIG. 17 • 3.—Cost curves of a hypothetical business firm that is losing money.

be paid whether we shut down or keep going.[1] If we shut up shop, we shall have no income and lose the entire $670 of fixed expense. If we stay in business and do the best we can, with an output of 1290 units, we shall lose $65. It is unfortunate to have to make such a decision, but after all one often has to make a choice between evils, and it should not be difficult to decide whether to take a loss of $65 or $670. It is better to stay in business temporarily and take the small loss than to close the plant and take the large loss.

[1] As an indication of the approximate accuracy with which one can read these curves by eye, it may be a matter of interest that the curves of the diagram were computed from formulas in which the fixed expense was actually $676.

Evidently, since the fixed costs have to be paid anyway whether we do business or not, *we are better off not to close the plant if we can do anything better than cover the variable expenses.* If we can pay all the variable expenses and have something, no matter how small, to pay toward the fixed expenses, our loss will be less than if we close the plant and lose our entire fixed expense. If we put this in terms of the diagram of Fig. 17 • 3, we can say that a price that lies above the point where the average cost curve and the marginal cost curve intersect will yield a profit, while any price below this will yield a loss; as long as the price is above the point where the marginal cost curve and the average variable cost intersect, it will be better to continue production at a loss than to close up shop. If worst comes to worst and if the price drops below the point where the marginal cost curve and the average variable cost curve intersect, it will pay us to cease production entirely.

Of course, in the long run a firm must cover all its costs, both fixed and variable, if it is to remain in business. Only temporarily can it neglect them. If in our last example the firm loses $65 and if it continued to lose $65 over and over again, ultimately it will be forced out of business. The situation is not a stable one—not one that can be maintained indefinitely. The price of $1 is evidently not a normal price if the costs are those shown in the diagram. It is too low a price to be continued permanently. This firm and others like it will be forced out of business in time, and the output of the product will be reduced. A price of $1 may bring forth the output for the time being, but in the long run it will stimulate a rate of output that is below the rate of consumption.

We have seen, then, that the market price cannot stay permanently above the point of lowest average cost. If it is temporarily above that point, it will either stimulate added output, bringing the price down, or it is an indication that the firm's accounting is bad—that the firm has not evaluated some of its productive factors at a high enough rate. In this latter case, the correct evaluation of the factors will raise the average cost curve to the point where the price and the lowest average cost are equal. But we have also seen that the market price cannot stay permanently below the point of lowest average cost. If it is temporarily below that point, firms may continue temporarily to do business at a loss, but ultimately they will be driven out of business and the reduced supply will raise the price, or else the firm will decide that it has valued some of its assets at too high an amount and will reduce its costs, bringing the average cost curve down to the point where the lowest average cost coincides with the price. Temporarily the market price may be far above or far below the lowest average cost. In such cases, the firm will be making temporary profits or suffering temporary losses. But *profits and losses are both temporary phenomena in business.* In the long run, the prices and costs are so adjusted that the firm's entire gross

income is paid out to various claimants with nothing left over and with no claimants left unpaid. This means that we can state a new long-run principle of normal price to go with our former principle of temporary or market price, as follows:

Temporary or market price: *At any given time the tendency is for every competitive business firm to adjust its output in such a way that the marginal cost and the selling price are equal.*

Long-run or normal price: *In order that an industry be in long-run com-*

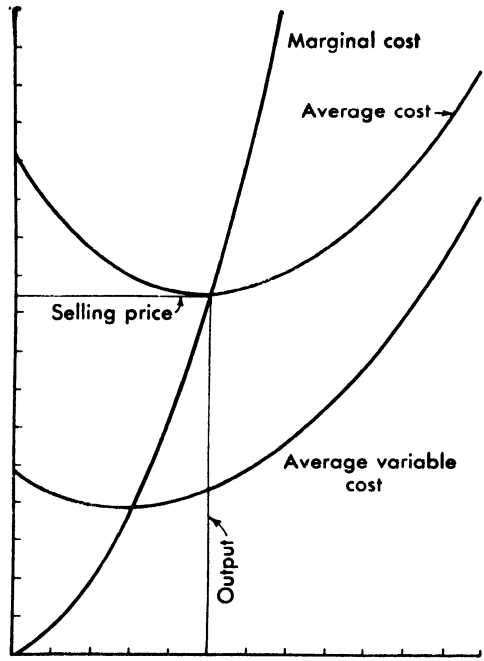


FIG. 17 • 4.—Relationship between costs and selling price for a firm in long-run equilibrium.

petitive equilibrium, it is necessary that each firm in the industry adjust its output and its costs in such a way that the marginal cost, the average cost, and the selling price are all equal.

The situation of long-run competitive equilibrium is that depicted in Fig. 17 • 4. The firm is now covering all its costs and paying off all claims against it, with nothing left over. It is just breaking even.

The tendency is for market price and marginal cost to be equal all the time. Individual firms may differ greatly in their average costs at any time but, insofar as they make the most profitable adjustment to prices, their marginal costs will all be the same, and all equal to the selling price. Since their average cost curves may be different, some of them may be making

temporary profits or incurring temporary losses, but they will all be breaking even "at the margin." If, however, one of the firms is able to continue to show a "profit" year after year under competition, it shows merely that its accountants have made a mistake. This thing that they have called "profit" indicates, perhaps, that the firm has unusually capable management, and this income is really properly chargeable as wages of the manager. It should be included among the costs just like any other wages. The manager is worth it if he can continue to bring in this extra income. Or possibly the continued "profit" is really a reflection of the fact that the accountant has not charged enough for the use of the firm's assets—its land, equipment, or patent rights.

17 · 6. Temporary Neglect of Fixed Costs. Many ordinary familiar phenomena can be explained in terms of the principle that businessmen find it to their temporary advantage to continue in business even at a loss as long as they cover their variable expenses—that they will, if hard pressed by competition or bad luck, neglect their fixed costs for the time being and sell at any price that covers the variable costs. For example, a merchant who has bought goods and paid $10 each for them, hopes to sell them at a profit, perhaps at a price of $15. But if the goods suddenly go out of style and he is unable to sell them for $15, he is better off to reduce his price than not to sell them at all. To be sure, he will cut the price no further than he finds necessary; but if he gets the price down to $10 and the goods still do not sell, should he refuse to lower the price further because it will involve a loss, because he actually paid $10 for them? If he keeps his price at $10 and does not sell the goods, he has lost the entire $10 that he invested. It is now fixed cost. He cannot retrieve it except by the sale of the goods. If he sells the goods for $8 each, he will take a loss of $2, but this is better than taking a loss of $10. If it becomes necessary, a sensible businessman will sell goods at below his actual cost rather than not sell them at all. His investment is now a fixed cost and, if he has to, he will neglect it.

Under the circumstances just described, how low a price can the merchant set rather than not sell the goods at all? That depends on the variable cost, if there is any. If the sale of goods involves such additional costs as putting them in packages, and mailing or delivering them, then the merchant cannot afford to set any price below that which will pay such costs. Suppose that the actual figures are

Original cost of the goods to the merchant	$10
Price originally set for resale of the goods	15
Costs that must still be incurred if goods are sold	2

The merchant will, of course, get the full $15, or even more, if he can. If he cannot, he will take anything that more than covers the variable costs. But the only costs that can still be altered are the $2 per unit for packaging

and shipping. If the merchant does not sell, he loses $10. If he sells the goods at $3 each, his results will be

Original outlay	$10
Cost of packaging and shipping	2
Total outlay	12
Selling price	3
Amount of loss	9

He loses $9 if he does sell, but $10 if he does not. The sale price of $3 more than covers his variable cost of $2, and so it is better to accept this price than to accept nothing. Sales of out-of-season style goods are often illustrations of this sort of thing, as are sales of back-model refrigerators or automobiles, which have been replaced by improved lines. It is no longer necessary or possible to retrieve all the costs that have been sunk in these goods. They will be sold, if necessary, for any price as long as it covers the costs that are still to be incurred. *The only costs that influence business decisions are costs that have not yet been incurred.*

Another illustration of the same sort of thing is one that has raised problems over and over again in the field of international relations. Suppose that a firm is doing business in the United States and that it has fixed costs of $500,000 and variable costs of $2 per unit of output. In its regular domestic market it is able to sell 100,000 units of output. Its total cost is $200,000 of variable expense and $500,000 of fixed expense, or a cost of $700,000 for the 100,000 units. If it sells these goods for $7 per unit, it will just break even. Let us suppose that the firm gets an opportunity to bid on a contract for 1000 units of these goods to be shipped to Brazil, England, or New Guinea—goods that are to be sent far out of the ordinary market, where they will not compete with the goods already being sold. Is it still necessary to charge $7 per unit to cover the costs? Obviously not, since the fixed costs have already been paid by the goods sold in the regular market. The variable cost is but $2 per unit and, if we can get any price above that, we are better off than we would be not to take the new business. For example, suppose we offer to sell the goods in Brazil at $3 each. Our financial picture will now be as follows:

Total fixed cost	$500,000	
Variable cost on 100,000 units sold at home	200,000	
Variable cost on 1000 units sold in Brazil	2,000	
Income:		
From 100,000 units sold at home at $7 each	700,000	
From 1000 units sold abroad at $3 each	3,000	
Total income		$703,000
Total expense		702,000
Profit		$ 1,000

The firm which was just breaking even when it sold the product at home for $7 per unit is now making a profit by selling an additional 1000 units abroad at $3 per unit. This is because the fixed expenses had been paid already, and only the variable expenses needed to be covered on new business. The practice that this firm followed is called "dumping on a foreign market," and it naturally is not looked on with favor by producers in the country to which the goods are shipped—producers who have to cover all their expenses in their own market, but who are facing competition from foreigners having to cover only their variable expenses. *Dumping on a foreign market is the practice of selling goods delivered in a foreign market at a price lower than that which is charged by the same seller in his domestic market.*

There are many other examples of our general principle that *variable costs are the only costs that are considered by sensible men in determining their selling price.* Thus, if a farmer is shipping fresh asparagus by truck from central Massachusetts to Boston every morning and the trucks are coming back empty, how much would he have to charge a Boston wholesaler for carrying his products back on the homeward journey? The trucks are coming back anyway. The drivers will be paid anyway. The gasoline will be burned and the license fees incurred anyway. These costs are already fixed. If the trucks have to go out of their way on the return trip or if the wear and tear on the trucks is greater, these costs will be variable. They will vary according to whether he takes the goods on the return trip or not. The farmer can afford to bring the goods back with him if he gets enough to cover the variable costs; if he can get more than that, as he surely will be happy to do if he can, he may well consider the return trucking business a profitable one. He will say that the main business of asparagus trucking "carries the overhead," by which he means that he gets enough out of his main business to cover the fixed costs. These costs are often called "overhead costs" by businessmen. Any business that has a main product and one or more by-products is apt to set prices for the main product which cover all the expenses, both fixed and variable; under such circumstances, the by-products need bring in only enough to cover the variable expense. In other words, we can point out again that the seemingly prosaic and abstract principles of economics really do describe human behavior and help us to understand phenomena of daily life, which would otherwise be impossible to explain.

17·7. Joint Costs. The case of products and their by-products about which we have just been speaking brings us to another important cost relationship, which the economist calls "joint costs." *When two or more commodities are inseparably produced together, so that changes in the output of one of them are necessarily associated with changes in the output of the other, we say that the two commodities are joint commodities, or that they are produced under conditions of joint cost or joint supply.* Thus the cotton farmer spends

money for fertilizer and labor, and his products are both the cotton fiber used for making cloth and the cotton seed used for meal and oil. He has two products that are produced together. If he decides to grow more cotton fiber, he finds that he has grown more seed along with it. The lumber mill produces boards, sawdust, and slabwood all by the same processes. The wheat farmer produces grain and straw. It is not uncommon for the smelting process to extract both gold and silver from the same ore. In the meat-packing industry hundreds of products come from the same animal: many cuts of meat, fertilizer, glue, buttons, medicinal preparations, and other things too numerous to list here. These are all cases of joint cost.

The inseparability of the outputs of the products jointly produced brings about an important relationship in the prices of these products. Let us illustrate with the case of cotton fiber and cotton seed. Suppose that a farmer finds that he grows 2 pounds of cotton seed with every pound of fiber. Let us also suppose that the seed sells for 2 cents per pound and the fiber for 10 cents per pound, and that these are normal prices; that is, these prices could be continued indefinitely without bringing about any maladjustment of production and consumption.

First, we notice that it is impossible to find any "normal price" of either cotton or cotton seed in the usual sense; it is impossible to draw any cost curves for them, because it is impossible to separate their costs. We cannot say that a given expenditure for fertilizer was a cost of fiber, nor that it was a cost of seed. It was a cost of both jointly. Therefore, we would be in trouble if we tried to say that at any time the price of cotton will equal its marginal cost and, in addition, that in the long run it will equal the average cost. We cannot find the cost of cotton fiber. It is inextricably interwoven with the cost of the seed.

We can find the joint cost of producing 2 pounds of seed and 1 pound of fiber. This combination evidently costs about 14 cents, because we note that, when we sell 2 pounds of seed at 2 cents and 1 pound of fiber at 10 cents, we are getting 14 cents income. If this were lower than the cost, producers would ultimately be driven out of business and the prices would not be normal. If it were much above the cost, it would stimulate expansion of output and the prices would not be normal. If cotton fiber and cotton seed can sell indefinitely at these two prices, it must be that they just cover the costs of production, and

2 pounds of seed at 2 cents	4 cents
1 pound of fiber at 10 cents	10
Income of the combination	14 cents

Now let us see what would happen if there were a sudden change in the demand for one of these products. Suppose, for example, that a new synthetic

fabric were invented which could compete with cotton and, that, as a result, the demand for cotton fiber were reduced somewhat. The reduction in demand would tend to lower both the price and the quantity sold, as we noted in Sec. 15 · 8. But we cannot reduce the quantity of cotton fiber without reducing the quantity of cotton seed. This is the key fact in cases of joint cost—the fact that the products are inseparably produced together and that we cannot alter the quantity of one without altering the quantity of the other. So, in this case, we reduce the quantity of fiber in answer to the reduction in demand and, at the same time, the quantity of cotton seed even though there has been no change in the demand for it. This will, of course, tend to raise the price of cotton seed, until some new equilibrium is established. If the changes in output do not change the cost situation, it will still be necessary to get enough out of the fiber and seed together to cover the 14-cent expenditure. If the new competing fabric forces the price of cotton fiber down to 4 cents per pound and, if the competition is such that the price of fiber has to stay there, farmers would not cover their costs if cotton seed stayed at the original price of 2 cents per pound. This would give them

2 pounds of seed at old price of 2 cents	4 cents
1 pound of fiber at new price of 4 cents	4
Total income now	8 cents

The fact that a new textile has been invented does not alter the cost of growing cotton. It used to cost 14 cents, and it still cost 14 cents, to produce 2 pounds of seed and 1 pound of fiber. Evidently the price of seed will have to rise if the price of fiber is to fall, and we have already seen that the price of seed would tend to rise anyway because of a reduction in the quantity produced. If we are to restore equilibrium, we must get a new price of 5 cents per pound for the seed, which will give us the following:

2 pounds of seed at new price of 5 cents	10 cents
1 pound of fiber at new price of 4 cents	4
Total income now	14 cents

We are now again getting 14 cents, which is enough to cover the cost. One of the products has fallen in price and the other has risen. And this, we see, is just what can be expected in such cases: *In the long run, the prices of joint products tend to move in opposite directions.* In a particular year the price of cotton fiber may be high or low as the result of temporary supply or demand situations; but if something happens to make a long-run change in the situation, it will affect both products and not just one of them. Both products together must bear their joint cost, and the way in which they share it will depend on the demands for them. Sometimes a commodity that had

been a by-product suddenly becomes the main product as new uses are discovered or as old uses for the former main product disappear. But all the products jointly produced must, between them, cover the cost of production in the long run and, if one of them bears more, the others will tend to bear less. In these cases, by the way, we would have a "cross-elasticity of supply" analagous to the cross-elasticities of demand discussed in Sec. 7 · 8. A change in the price of cotton fibre will tend to influence the quantity of cotton seed offered for sale. But this will be only a long-run tendency, not an immediate or short-run tendency such as that involved in cross-elasticities of demand.

17 · 8. Time Element in Economic Problems. Throughout this chapter we have fallen into the easy habit of saying that some things will be true "in the long run" or "in the short run." What do these phrases mean? How long is a long run, and how short is a short one? The answer is that the time involved is an unusual sort of time, not measured with stop watches, clocks, or calendars. When an economist says that something is true "in the long run," he is not thinking about an amount of time in the usual sense at all. He is thinking, rather, about what happens in the economic system—what things are given and fixed and what things are still variable. If we ask what factors determine the price of apples, it makes a difference whether we are thinking of the present apple crop, which is already produced and picked and ready for market, or of the price of apples on the average over a period of 100 years or so. In the former case, the costs are all fixed. There is no way we can now get back what we paid last summer for thinning and spraying the crop, or 10 years ago for fitting the land and planting the trees. In the short run, the goods have already been produced, and our problem is that of finding who will get them and what they will sell for. *The economist's short-run problems are problems concerning existing quantities of goods.* But these quantities can be changed from year to year. More or less land may be put into orchards, and more or less labor and capital may be attracted to the apple business. The short-run problems of distributing and pricing a given apple crop change, in the long run, into problems of determining how the factors of production will be allotted to this and to other competing enterprises. In the long run, the size of the crop is not fixed but variable. Yet in the economist's long run some things are fixed. In considering what will happen in the long run, he assumes that human wants will stay as they are, that the resources available for production will remain as they are, and that the known methods of production will stay as they are. In other words, in the short run, there are fixed supplies of goods to be distributed; in the long run, the supplies of goods are variable, but the wants, resources ,and technology are fixed. When an economist says that the normal or long-run price of milk is 14 cents a quart, he means that that price would stimulate a rate of output

equal to the rate of consumption *if wants, resources, and technology remain unchanged.* It may be that this 14-cent price will bring about some change in the actual distribution of resources, drawing some additional resources into the milk business or squeezing some out before equilibrium is reached.

To be sure, wants, resources, and technology are all changeable. If we consider past history, we can see radical changes in all of them. They do not seem to change in any predictable pattern, but they do nevertheless seem to be changing most of the time. The economist has no basis to go on if we ask him, "What price would you expect for milk if human wants and resources change and if technology changes?" One can predict what will happen only if he can predict it within some known framework. Thus the physicist can tell you how rapidly a body will fall in a vacuum, in air at normal pressure, in water, or in mercury; but if you say to him, "I'm going to drop a body and I want you to tell me how rapidly it will fall, but I won't tell you into what medium I shall drop it," then the physicist has no way of answering. If the economist can be told the wants, resources, and technology, then he can tell what adjustments will take place in the long run; but, if he is not told these things, he has no basis for a prediction. If we allow all these things to change, we say that we are dealing with problems of the "historical long run," and to deal with these problems one has to predict, by ways unknown to the ordinary economist, what will happen to the institutions that govern human wants, resources, and technology. This field is sometimes called the field of "institutional economics," and is one of the places where formal economics comes most closely in contact with the problems of the other fields of sociological thought.

To summarize, then, we can classify the problems faced by the economist into three groups, depending on the things that are given or fixed for the purposes of our problem:

1. The short-run problems, which deal with given supplies of goods and their pricing and distribution.

2. The long-run problems, which deal with the allocation of the various factors of production under conditions of fixed wants, resources, and technology. The supplies of particular goods are no longer fixed, since factors of production can be moved back and forth from one field of production to another if it is profitable to do so.

3. The problems of the historical long run, in which nothing is fixed and in which the economist cannot at present predict what will happen except by some such dodge, as by saying that, *if* such and such changes take place in wants, resources, and technology, then we can expect such and such results.

To illustrate all three of these types of problems we can go back to the

case of apple production. A short-run problem in the field would be one of deciding what the price of apples is going to be under given supply and demand conditions, and what forces will affect it—what consumers will get the existing apples, and how much they will have to pay for them. The long-run problems would be those of deciding how many apples will be produced and how much labor and capital will be employed in the apple business after we give orchardists time to adjust their business to the given conditions of wants, resources, and technology. No business is ever completely adjusted to these things. Wants and resources and technology are changing from day to day, and even the most up-to-date producers are ever somewhat out of adjustment to them. If we could freeze these things—stop all changes in them—then producers might ultimately get their businesses adjusted, and these adjustments are the problems of the long run. But the problems of the historical long run are those of knowing what is really going to happen, considering the fact that wants, resources, and technology can safely be expected to change. New varieties of apples will be found, new sprays and fertilizers and storage methods will be discovered, consumers will change their tastes, etc. Under these conditions, what is going to happen to the apple business? The answer is entirely in the field of speculation except to the extent that the sociologist can predict from his knowledge of human institutions and the forces affecting them what is likely to happen to human wants or to the trend of invention and discovery. To be sure some interesting sociological theories have been propounded in these fields and, although there seems little likelihood of any immediate success in solving the extremely difficult problems involved, even meager knowledge in the field is better than none. In practice, most economists confine their attention to problems of the short and of the long run. This does not mean that the other problems are unimportant, but merely that they are difficult and elusive and usually disappointing.

17 · 9. Average Revenue Curves and Marginal Revenue Curves. In Sec. 16 · 10 we defined marginal revenue and discovered how it is computed. We also noted that it is directly related to elasticity of demand, so that there is a definite mathematical relationship between elasticity and marginal revenue. We found that the marginal revenue, which is exactly equal to the price under conditions of complete competition, is always less than the price when the conditions are noncompetitive. Although we have not used the phrase before, it is at once evident that the price received for the goods that we sell is the *average revenue,* related to the marginal revenue in just the same way that average cost is related to marginal cost. We shall find it convenient, in order that the relationship between average and marginal revenue may be kept in mind, to use the term "average revenue" often in the remainder of this chapter, and the student may remember that it means nothing more

complicated than "selling price." We said that the average cost of a given output of goods was the total cost divided by the quantity of output. Similarly, the average revenue is the total revenue or total income divided by the amount of the output. If we get $5000 total revenue by selling 1000 pairs of shoes, the average revenue is $5 per pair; that is, the price is $5 per pair.

It is now important for us to see the relationship between marginal revenue and average revenue as output changes. We have already said that, if the price is fixed (the average revenue, that is, is fixed), the marginal revenue and the average revenue are equal. Thus in Fig. 17 • 5, if the average revenue or selling price stays constant at all levels of output, the average revenue curve will be a horizontal straight line, and the marginal revenue curve will coincide with it. If the price of wheat is $1.30 per bushel and stays stubbornly and persistently at that price no matter how much or how little we offer for sale, then the marginal revenue curve and the average revenue curve appear as coincident horizontal straight lines, as in Fig. 17 • 5. Of course, we recognize this as a case of complete competition, where the producer is too small to influence the price on the market and where the marginal revenue is equal to the selling price. *In strictly competitive cases, the average revenue curve and the marginal revenue curve are combined in the same horizontal straight line.*

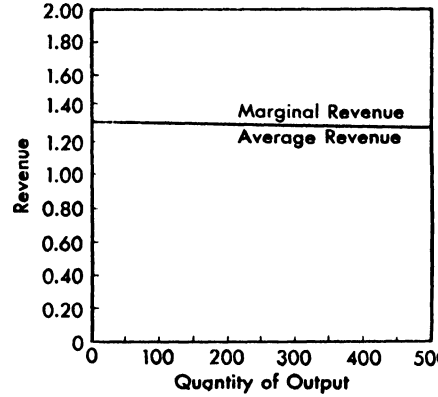


FIG. 17 • 5.—Average and marginal revenue curves under complete competition.

In cases that are partly or completely monopolistic, on the contrary, increased output will lower the price or average revenue. The producer is important enough in the market so that, as he alters his output, the market price responds, rising when he lowers his output and falling when he increases his output. This is one of the identifying peculiarities of noncompetitive situations. Under such circumstances, the average revenue curve will fall as the output increases, as in Fig. 17 • 6. We note, by the way, that the average revenue curve shows the price or average revenue that the producer will realize at various quantities of output. But that is what the demand curve also shows. If we wished to depict the demand for the output of this particular producer, we would show the quantities that consumers would buy at various prices; the average revenue curve shows the prices that can be realized for various quantities of output. The curve showing the

demand for a producer's output and the curve showing his average revenue are really identical. In the case of complete competition, the marginal revenue curve becomes a horizontal straight line because under such circumstances, as we have seen, the elasticity of demand is infinitely great. In monopolistic cases, there is some finite elasticity of demand, and the demand curve, or marginal revenue curve, takes on a negative slope, with its steepness increasing as the elasticity decreases.

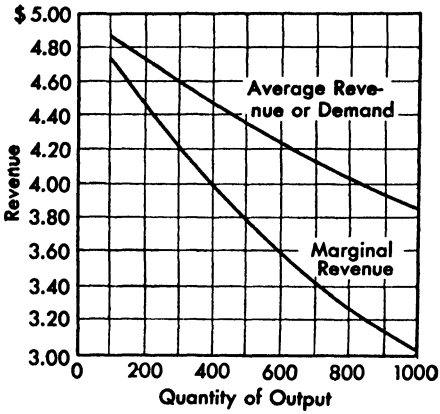


FIG. 17 • 6.—Average and marginal revenue curves of a firm operating under noncompetitive conditions.

We have seen, however, that in noncompetitive cases the marginal revenue is always lower than the selling price. This is indicated in Fig. 17 • 6 by the fact that the marginal revenue curve lies below the average revenue curve. If, for example, the firm whose curves are depicted in the diagram decides to produce and sell 600 units of output, we see by inspection of the graph that the selling price will be approximately $4.25, but the marginal revenue will be approximately $3.60. Thus, when we draw the average revenue curve (or demand curve) and the marginal revenue curve for a firm, we have to follow one or the other of the two rules:

1. If the firm is operating under competitive conditions, the two curves coincide and form a single horizontal straight line.

2. If the firm is operating under conditions that are completely or partly monopolistic, the average revenue curve and the marginal revenue curve will both be represented by sloping lines, and the marginal revenue curve will lie below the average revenue curve.

While we are still on the subject of Fig. 17 • 6, it will pay the student to recall what is meant by the two figures just read from the curves. We said that, at the output of 600 units, the average revenue would be $4.25 and the marginal revenue would be $3.60. This means that with this output we are selling the 600 units at an average price of $4.25; but, if we expand our output, we shall lower the price enough on the 600 units so that the added income from another unit of output would be but $3.60. Also we can see that the demand curve depicted in Fig. 17 • 6 must represent an elastic demand throughout the entire part shown in the diagram, because if the demand were inelastic the marginal revenue would be negative, and if there were

unitary elasticity of demand the marginal revenue would be zero. The lowest marginal revenue shown in the diagram is a little over $3 and, since all the values are positive in the diagram, we know that, at least over the range of prices and outputs shown, the demand is elastic.

17·10. Price Setting under Noncompetitive Conditions. Let us now combine the revenue curves just studied with cost curves of the types that we have used before. Every firm, whether monopolistic or not, will have some cost curves, and their shape will be of the kind that has become familiar to us. In competitive cases we assumed that there was a given market price, and we asked what it would pay the firm to do under the circumstances. But the monopolist may set his price where he will, subject to the forces discussed in the preceding chapter. Consequently, instead of having a given fixed price, we have the possibility of selecting the particular price that seems to hold the greatest promise of profit.

Suppose that Fig. 17·7 shows the marginal cost curve *MC*, the average cost curve *AC*, the average revenue curve *AR*, and the marginal revenue curve *MR* for a particular firm. A glance at the diagram shows us that the firm operates under noncompetitive conditions, since the revenue curves have negative slope, with the marginal revenue curve below the average revenue curve. If these curves represent the cost and income facts for the firm, what will it pay the firm to do? We already know the principle involved; namely, that the firm will maximize its profit if it selects the particular output that makes the marginal revenue and the marginal cost equal. We look at the figure and find the place where the marginal cost curve and the marginal revenue curve cross. As nearly as we can judge by inspection, this is at the point where the output is 575 units. Therefore an output of 575 units is the most profitable one for us. But how profitable is it? The average revenue curve tells us that, if we sell 575 units, we shall get an average revenue for them of about $4.30 each, and the total income from 575 units at $4.30 each will be $2472.50. The average cost of an output of 575 units, read from the average cost curve, is $4.05. The total cost of 575 units at $4.05 will be $2328.75, leaving a net income of $143.75. This is the largest net income that the firm can make from any output. Always in analyzing a noncompetitive situation, we start with our cost curves and revenue curves. First we discover what output will be most profitable. This is the output where marginal cost and marginal revenue are equal. We then discover from the demand curve (the average revenue curve) the price that can be obtained for this output, and the total income that will be realized. Also, from the average cost curve we discover the average cost of the given output, and the total cost. The difference between the total cost and the total income is the net income or profit.

In Fig. 17 • 7, the marginal revenue and marginal cost curves cross at *A*, corresponding to an output of 575 units at *F* on the horizontal scale. But an output of *OF* will yield a price of *FC*, as shown by the average revenue curve. This is a price of $4.30. The average cost will be *FB*, or $4.05. The rectangle *DCFO* will represent the total income. The rectangle *EBFO* will represent the total cost. The difference, or the rectangle *DCBE*, will represent the net income, and this net income rectangle will be larger than any other corresponding net income rectangle that we can get.

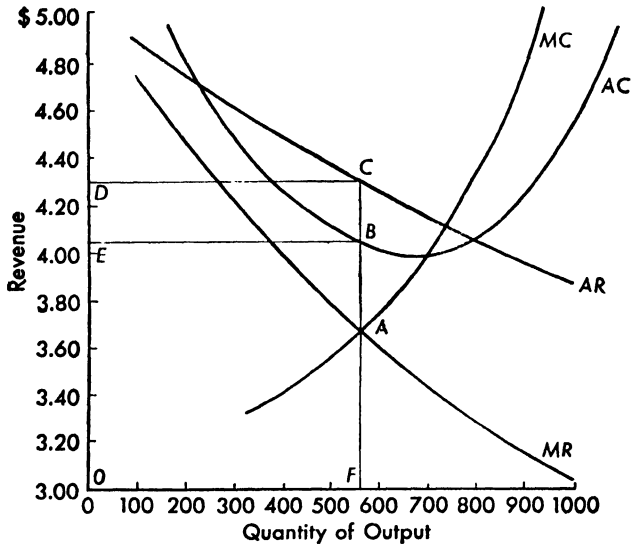


FIG. 17 • 7.—The setting of monopoly price.

17 • 11. Normal Price under Noncompetitive Conditions. We discovered in Fig. 17 • 7 that it would pay the producer to set his price at $4.30. Since he has at least some element of monopoly power (as shown by the fact that his average and marginal revenue curves have negative slope), he can set the price where he wishes, but the conditions of demand and cost are such that it pays him to set the price at $4.30. Is this a normal price? Can it continue without setting up a production rate that exceeds the rate of consumption? That depends on the degree of monopoly control exercised by the firm. If they have a complete monopoly and no competitors can get into the business, then, of course, they can continue to hold their price at $4.30 as long as they wish, and they will find it to their advantage to do so as long as the demand and cost conditions stay unchanged. If this is a case of partial monopoly, so that other producers can get into the business in time, then the profit that is being made will stimulate the growth of competition. Slowly other firms will increase their output and take away some of this firm's customers.

The demand for its output will fall as former customers shift to competitors, and the demand curve will drop to a lower position. Thus the long-run tendency would be, in a case of partial monopoly, for the average revenue curve and the marginal revenue curve to fall. Evidently, however, as long as the average revenue curve cuts across the U of the average cost curve, as it does in Fig. 17 • 7, there will be some point where the average revenue curve is higher than the average cost curve, some point where the firm's average revenue exceeds its average cost, some point at which the firm can make

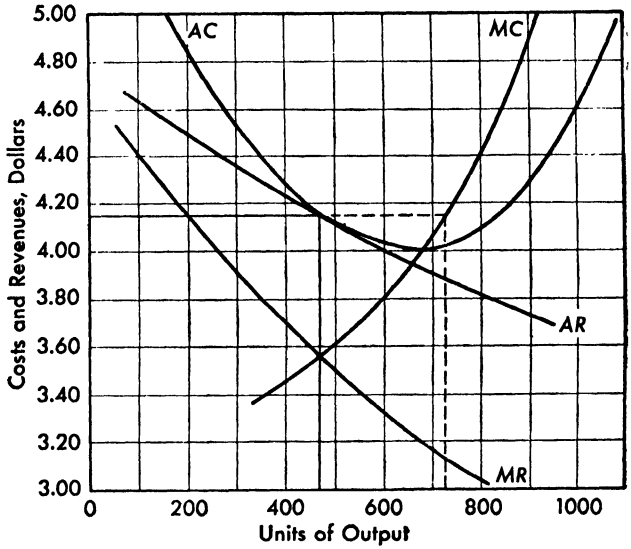


FIG. 17 • 8.—Long run equilibrium for imperfect competition.

a profit. Therefore, as the average revenue curve falls lower, reflecting the loss of business to competitors, the output of our firm will shrink, but it will still be making a profit and attracting competition. This will continue until finally the average revenue curve assumes the position shown in Fig. 17 • 8, tangent to the average cost curve. Now the marginal revenue curve crosses the marginal cost curve at an output of about 475 units. At this figure, the selling price and the average cost are both $4.15, so that the firm is no longer making a profit nor taking a loss. It is just breaking even. It makes enough, to be sure, to pay all its expenses, but it has no excess left over in the form of monopoly profits. *Profits are temporary phenomena under conditions of partial monopoly, just as they are under conditions of complete competition. Only with complete monopoly can profits continue indefinitely.*

One point of interest in connection with Fig. 17 • 8 is immediately apparent. The long-run or normal price of this commodity under conditions

of partial monopoly was $4.15, and at this price the firm produced 475 units of output. How many units of output would the public have obtained from a competitive firm at this same price? We recall that a competitive firm expands until its marginal cost equals the selling price. In Fig. 17·8 the marginal cost equals $4.15 at an output of approximately 710 units, as shown by the dotted lines of the diagram. Competition would have produced 710 units at this price, while the existence of some monopoly control has reduced the output at the price to 475 units. *There is always a tendency for uncontrolled monopoly to restrict output below that which would be produced under competitive conditions, and this tendency for restriction continues even if the monopoly power is not complete so long as the firm is able to influence the selling price of the product.*

17·12. Summary of Normal Price Relationships. If we may now draw together the tendencies outlined for long-run normal price under various sorts of conditions of production, we have discovered the following:

1. Under competitive conditions, a price is not normal and cannot continue indefinitely, unless the marginal cost, the average cost, the marginal revenue, and the average revenue (or selling price) are all equal (see Fig. 17·4).

2. Under conditions of complete monopoly control, when competitors are barred from the business, we do not have normal price unless marginal revenue and marginal cost are equal. This will give both short-run and long-run equilibrium for complete monopoly (see Fig. 17·7).

3. Under conditions of partial monopoly or incomplete competition, where a firm can influence the price of a commodity but does have competitors, the price is not normal unless marginal revenue equals marginal cost and, in addition, average revenue (or price) equals average cost (see Fig. 17·8). A firm operating under partial monopoly is in temporary equilibrium, doing the best for itself that it can at the time, if marginal revenue and marginal cost are equal, as in Fig. 17·7; but such a condition cannot last unless also average revenue and average cost are equal, as in Fig. 17·8. For short-run equilibrium it is merely necessary that marginal cost equal marginal revenue regardless of the presence or absence of monopoly or competition.

17·13. Cost Is Not the Cause of Price. Finally, let us point out, after having discussed throughout this chapter the relationship of costs and prices, that the cost of a commodity is not the cause of its price in any ultimate sense. It might cost $100 a quart to get fresh milk from Siberia to New York City, but that does not mean that fresh Siberian milk would be worth $100 a quart in New York City. It does mean, to be sure, that if New York consumers are to get a continued supply of Siberian fresh milk, they will have to pay the cost, and the price will have to be $100 a quart; but, if the demand is not

there, the value will not be there. Value depends on both demand and supply. We noted that, in the short run under competition, prices were determined by the interaction of demand and supply. Then we noticed that, in the long run under competition, the price must be enough to cover average cost if the supply was to continue; but we also saw that outmoded radios or refrigerators might sell below their total cost. Only the variable cost had to be paid if consumers were not demanding a continuation of the supply. More recently, in the cases of monopoly and partial monopoly, we have seen that the marginal revenue and average revenue played a part in the determination of normal price. But average revenue and marginal revenue are derived from demand. Therefore, although we have found it convenient to approach the problem of normal price in terms of cost, it must be remembered that the cost alone is but part of the picture, and that demand plays a coordinate part. Both the interests of producers on the cost side and those of consumers on the demand side interact to influence the price.

17·14. Summary. The economist is interested in discovering the forces that operate in the economic system in the short run, in the long run, and in the historical long run. These periods differ in the things that are thought of as given or fixed. Short-run problems are those which treat of fixed supplies of goods. Long-run problems are those which treat of fixed wants, resources, and technology. The historical long run considers that nothing is fixed. The economist deals mainly with the short run and the long run. We have heretofore studied the forces that operate to determine prices in the short run. We now turn to the forces that influence long-run price. The normal price is the price that would, in the long run, bring about a rate of consumption and a rate of production that were equal.

A firm finds that it is wise, if necessary, to continue in business even at a loss if the price is sufficient to cover the average variable costs. The only costs that influence decisions as to selling price and volume of output are those which have not yet been incurred. Since in the long run all costs are variable, so in the long run it is necessary to cover all costs. Competitive firms are unable to get more than this in the long run, and are unwilling to accept less. Therefore the normal price is equal to both the marginal and the average cost under competition. We have a special case in joint cost situations, since under such circumstances the costs of the individual products can not be separated and measured. In such cases, the commodities have a joint normal price, equal to their joint cost of production, and the prices of the two commodities tend to vary in opposite directions.

In cases where there are no competitors at all, normal prices are those which make the marginal revenue equal to the marginal cost. Under such circumstances, there can be permanent profits, and the long-run adjustment is exactly the same as the short-run adjustment. In all other cases, whether

of complete competition or partial monopoly, both profits and losses are phenomena of the short run only, being impossible in the long run. The normal price for a firm which has partial monopoly power, but which does have competitors, is one that equates the marginal revenue and the marginal cost, and in addition equates the average revenue and the average cost. Either complete or partial monopoly, unless its policies are determined by the government, finds it to its advantage to restrict output below what it would be at the same price under competitive conditions; or, of course, to restate it in other words, to charge a higher price for a given quantity than would have been charged under competition. This tendency of monopoly to restrict output and to raise prices is the feature of monopoly against which the public tries to protect itself in ways that we shall discuss in Chap. 24.

Although prices of goods are related to their costs, at least in the long run, it is important to remember that prices are also influenced by demand. Both producers and consumers play a part, although often an indirect one, in determining the prices of the goods that they exchange.

Questions

1. If we were to find the average price of eggs for the past 20 years, would this be their normal price? How long a period would be have to take, approximately, to have the average price equal the normal price, or would it ever be equal to the normal price?

2. The market price of a commodity changes from day to day with changes in the supply of or the demand for it. Does the normal price ever change? If so, what things change it?

3. A man raises 500 barrels of apples with a fixed cost of $1000; the total cost is $1500. What is his average variable cost?

4. What is the shape of the average fixed cost curve? This would be the curve obtained by dividing the fixed cost by the various volumes of output.

5. What is the shape of the average variable cost curve? Why?

6. As businessmen use the term, there can, of course, be profits year after year. Some competitive business corporations have continud to earn "undivided profits" and to pay dividends, for decades on end. How can you make this jibe with the statement in the text that profits are temporary phenomena?

7. If a man owned a small retail store and worked in it himself, how could he live if he made no profit?

8. Suppose that Fig. 17·9 shows the cost curves for a competitive firm. What will be the lowest price at which it can make a profit?

9. What will be the lowest price that the firm of Fig. 17 • 9 can accept temporarily rather than go out of business?

10. Why do the average variable cost curve and the average cost curve draw closer and closer together as output grows larger in Fig. 17 • 9? *Hint:* Average cost is the sum of average variable cost and average fixed cost. What happens to average fixed cost as output grows? Why?

11. If the price of the product is $9.50, how many units will the firm of Fig. 17 • 9 find it most profitable to produce? How much profit will it make?

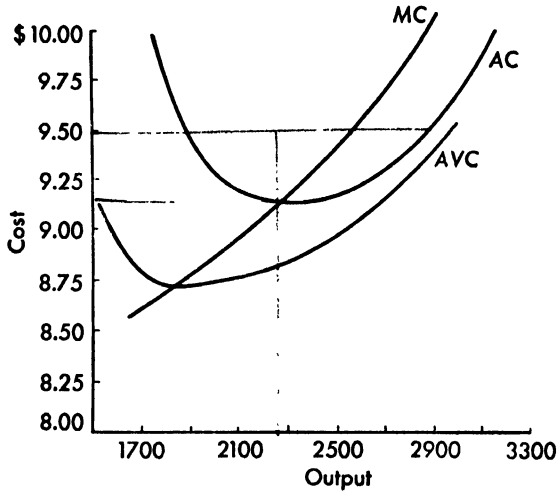


FIG. 17 • 9.—In this diagram *AC* is average cost, *AVC* is average variable cost, and *MC* is marginal cost.

12. What is the competitive normal price of the product of Fig. 17 • 9?

13. If the price of the product in Fig. 17 • 9 is $9, what would you advise the firm to do? Why?

14. If the price of the product is $8.25 in Fig. 17 • 9, what would you advise the firm to do? Why?

15. How can you tell that question 11 deals with a competitive case?

16. Figure 17 • 10 shows the cost and revenue curves for a given business firm. Is it a competitive business? How do you know?

17. What price will it pay the firm of Fig. 17 • 10 to set, and what volume of output?

18. How much profit will the firm of Fig. 17 • 10 make?

19. Under what circumstances can the firm continue to make this profit permanently?

20. If the firm of Fig. 17 • 10 produces an output of about 9400 units, the marginal cost, the marginal revenue, and the average cost will all be equal at about $7.45. Does this mean that the price is a normal price? Comment.

21. Figure 17 • 11 shows the cost and revenue curves for a business firm. The firm will lose money if it stays in business. How much will it lose? Should it continue to operate, or should it close down? Why?

22. If the firm of Fig. 17 • 11 does operate, what price should it charge and what volume of output should it maintain?

23. A firm is producing 500 units of a commodity for the home market at a price of $8 each and, at this price, is just breaking even. The firm's fixed costs are $3000. What is the lowest price per unit that this firm can accept for a foreign order of 100 units of the commodity if its average variable cost is constant? Why?

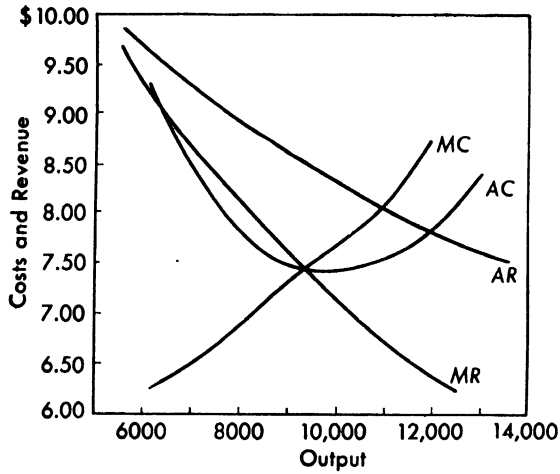


FIG. 17 • 10.—In this diagram, *MC* is marginal cost, *AC* is average cost, *AR* is average revenue, and *MR* is marginal revenue.

24. Two products, *A* and *B*, are produced under conditions of joint cost. The production process is such that 3 units of *A* are produced with each 4 units of *B*. These commodities have been selling at prices of 36 cents per unit of *A* and 8 cents per unit of *B*. A change in the demand for *A* lowers its price to 8 cents per unit. If the previous prices were normal, what will the new price of *B* tend to be?

25. Suppose a woodworking shop produces both cradles and bookcases. Are these products produced under conditions of joint cost? Why or why not?

26. What is the difference between joint costs and overhead costs?

27. Would a study of the economic effects of immigration be a study of short-run or long-run phenomena? Why?

28. We can tell from Fig. 17 • 10 how much output would be produced and at what price. How much more output would have been produced by this firm at the same price if the situation had been competitive?

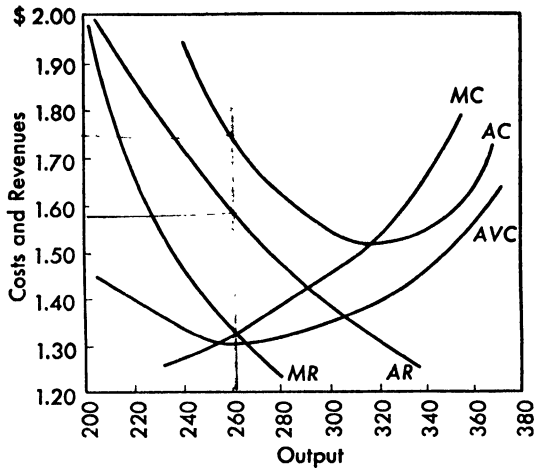


FIG. 17 • 11.—In this diagram *AC* is average cost, *MC* is marginal cost, *AVC* is average variable cost, *AR* is average revenue, and *MR* is marginal revenue.

18. MONEY

18 · 1. Nature of Money. In the early sections of Chap. 15 we learned of the difficulties of barter and of the reasons for the adoption of a medium of exchange to facilitate business transactions. We learned that it is often easier to make two exchanges than to make one, exchanging one's surplus commodities for money, and exchanging the money for other commodities, rather than bartering directly. We also learned that in most societies people sooner or later settle, consciously or unconsciously, on some one medium of exchange, or on a few commodities which they use as mediums, and that within the particular society almost all exchanges are carried on by means of the adopted mediums. We then call the medium "money." *Money consists of those things which, within a society, are of general acceptability, passing freely from hand to hand as mediums of exchange.* No commodity is acceptable everywhere, and in this sense money is always local; it is money in some places and in other places is not acceptable. Thus an American who tried to buy goods in an English store with an American five-dollar bill would find that it was not acceptable there, and it would not be money there, despite the fact that it is money in the United States.

Many diverse sorts of things have been used as money historically. In Bible times some tribes used grain and some used cattle as money. Certain African tribes have used ivory as money. In early North America men used beaver skins and blankets, while farther south the Virginians used tobacco. The Indians had used wampum as money, and in several of the early American colonies the use of wampum was legalized and the white men used it as such. The ancient Aztecs used cacao as money, and certain of the Abyssinians used thin bricks of salt. The province of North Carolina legalized the use of Indian corn, tallow, beaver and otter skins, butter, deer skins, and feathers as money. Marco Polo found the Chinese of his day using a sort of paper money manufactured from the inner bark of a tree. Our word "pecuniary" is said to go back in its derivation to the use of cattle as money among the early pastoral tribes.

Ernest Thompson Seton quotes the following example from early American history:[1]

> In 1788, it appears that in the mountain section of East Tennessee, the inhabitants formed themselves into a separate State, and called that section the "State of Franklin." At that time, money was scarce with which to pay the officers of the new State, so in October of that year [Oct. 18, 1788], the legislature of the "State of Franklin" enacted the following law to provide for the compensation of their officers:
>
> "Be it enacted by the General Assembly of the State of Franklin, and it is hereby enacted by the authority of the same, That, from the first day of January, A. D. 1789, the salaries of the civil officers of this Commonwealth be as follows, to-wit:
>
> "His excellency, the governor, per annum, 1000 Deer skins; his honour, the chief justice, 500 do. do.; the attorney-general, 500 do. do.; secretary to his excellency, the governor, 500 Raccoon, do.; the treasurer of the State, 450 Otter, do.; each county clerk, 300 Beaver, do.; clerk of the house of commons, 200 Raccoon, do.; members of the assembly, per diem, 3 do. do.; justice fee for signing a warant, 1 Muskrat do.; to the constable for serving a warrant, 1 Mink do."

When the student reads of men using tobacco, wheat, or other commodities of that sort, he is likely to think that men used these things *instead of* money. On the contrary, these things *were* money. They seemed just as natural to the people who used them as gold, silver, or paper money seems to us.

18 · 2. Qualities of a Good Money Material. After reading the foregoing list of things that have been used as money, one feels that almost anything might be used. Yet he should not get the idea that all these things were equally good as money. Some of them worked poorly, and lasted but a relatively short time. Others worked better. None of them worked perfectly, nor have men yet discovered anything that even closely approximates perfection as a monetary material. Some things have some advantages and other things have others, but nothing yet tried combines all the advantages that we desire in the perfect monetary material.

It is impossible to list the desirable qualities in any order of importance, but surely one of the things that we would want if we were selecting a substance to use as money would be *scarcity*. We would not want to choose something that was too plentiful. The reason for this is obvious. If we choose something that is too common, it will be cheap, a large amount of it will be required to make a purchase. We want our money to have *large value in small bulk*, so that it will be easily *portable*. On the island of Yap in the Pacific

[1] "Lives of Game Animals," vol. II, Part I, p. 236, Doubleday, Doran & Company, Inc., New York, 1929. The quotation is credited to *Forest and Stream*, Oct. 7, 1911.

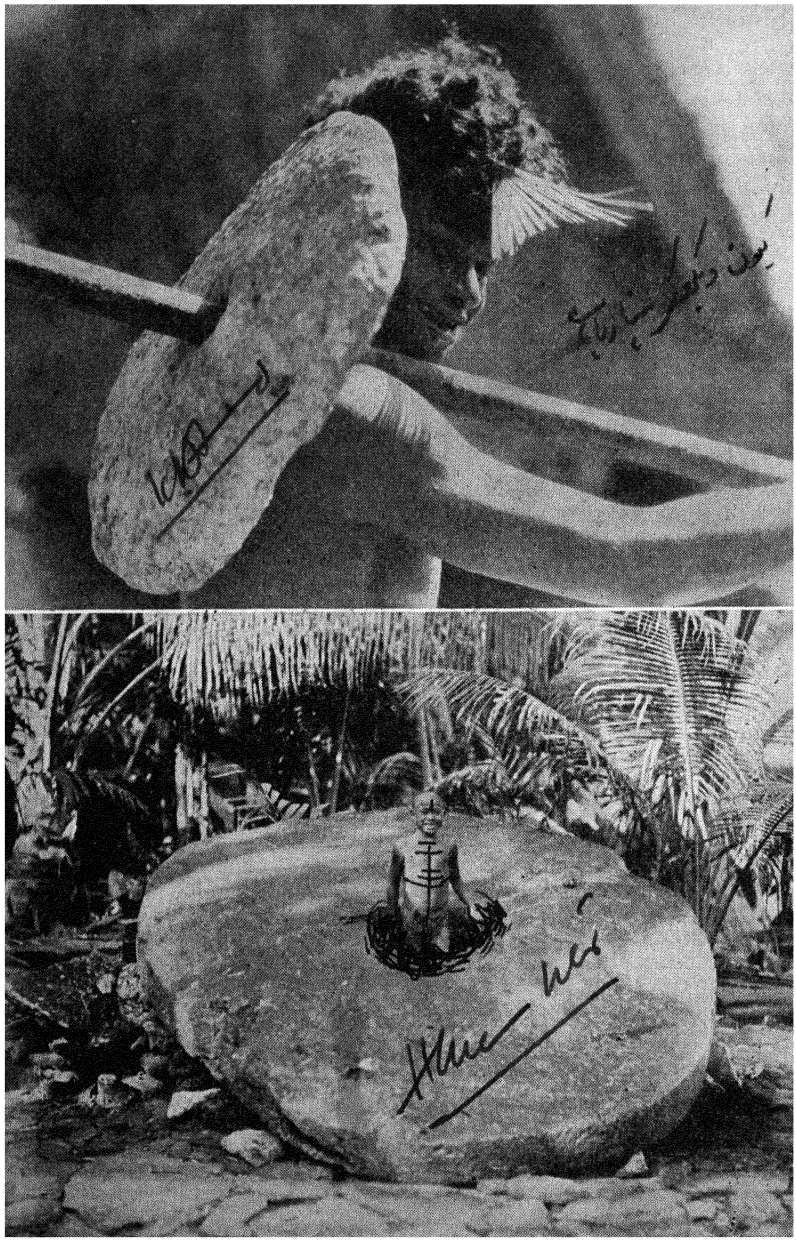


FIG. 18 • 1.—Millstone money used on the island of Yap. *Upper:* A native goes to market with a "small coin." *Lower:* One of the more valuable "coins" which the natives leave in front of their houses to advertise their affluence, secure in the knowledge that the money is too unwieldy to be stolen. (*Photographs by Willard Price, reproduced by permission.*)

Ocean, the natives have long used millstones as money (see Fig. 18 • 1). They are in the habit of leaving their money on display in the front yard as evidence of their economic position. The difficulties of carrying on any extensive commerce on the basis of such money are patent.

We would also naturally like to get a monetary material that was ***homogeneous,*** so that one unit had the same value as any other. If we were to use cattle as money, which many people have done, we would discover that all cows are not of equal value and, when we were paid our wages, it would make a decided difference to us which cattle we received. Today men make little or no distinction between different five-dollar bills, even though we shall see later that they have different histories and different backing. For practical purposes of trade, they are homogeneous.

It is convenient, in addition, to use as money something that is easily *divisible.* Debts may be large or small, and our money should be so divisible that we can pay even the smallest debt. If we use beaver skins as money and have a small debt to pay, we cannot cut off a tenth of a beaver skin without spoiling the rest. A tenth of a beaver skin does not have one-tenth the value of a whole one. On the other hand, gold or silver can be divided into smaller or smaller pieces, each of which retains its full proportionate share of the original value.

Durability is another characteristic for which we would look in choosing a money material. If we were to use grain, our money might mold and spoil. We would feel that we had to spend our money right away, before it lost its value. Even gold and silver were not durable enough to meet this requirement until men learned to alloy them by adding copper to make them harder. Not only do we want the material itself to last, but we would like it to have a *constant value.* We do not want to save up money for a "rainy day," or to send our children to college, or for our old age, and then find, when the time comes, that it has lost part or all of its value. Neither, if we owe a debt, do we want to find, when the debt comes due, that the money that we have to pay has increased in value and we have to pay a greater value than we had expected. Unfortunately no commodity is known that has a constant value. Some things fluctuate more than others, but nothing stays completely uniform. Gold and silver have been better in this regard than many other commodities, largely because of their durability. We still have in human possession a very large proportion of all the precious metals that were ever mined. Some has been lost and some has worn out, but, because it is durable and valuable, most of it has been carefully kept. The total amount mined in any one year is small compared with the entire stock on hand. Despite the fact that gold-mining output has recently been at an unusually high level, the combined monetary stocks of the nations are about twenty times the annual production. This means that even quite sizable variations in the annual rate

FIG. 18 · 2.—Privately produced gold coins issued and circulated in the United States.

1. Ten-dollar piece issued by Templeton Reid in Georgia.
2. Five-dollar piece issued by C. Bechtler in North Carolina.
3. Ten-dollar piece issued in 1849 by the Cincinnati Mining and Trading Company of San Francisco.
4. Fifty-dollar piece issued in 1851 by August Humbert, assayer, San Francisco.
5. Ten-dollar piece issued by J. S. Ormsby of Sacramento.
6. Ten-dollar piece issued by J. J. Conway, Denver.
7. Sixteen-dollar piece issued by Moffat & Co., San Francisco.
8. Twenty-dollar piece, issued by the Mormons at Salt Lake City in 1849.
9. Five-dollar piece issued by John Parson & Co., Tarryall Mines, Colo.

(*Catalog of Coins, Tokens, and Medals in the Numismatic Collection of the Mint of United States at Philadelphia, Pa., Washington, Government Printing Office,* 1914, *Plate* VIII.)

TEN
DOLLARS
GEORGIA
GOLD
1
2
3
4
5
6
7
8
9
J.J. IDDINGS
& CO.
BANKERS
MOFFAT & Co
20% CM AT
$16.00

of output bring but minor changes in the total stock on hand and, consequently, but minor variations in the value.

18 · 3. Coinage. If we do select as a monetary material something combining the properties that we have just listed, we still have the problem of recognizing and testing it in practice. Thus, if we were to select gold as the basis for our money, a man accepting it in payment of a debt would still have to satisfy himself that the metal offered was really gold, and he would have to test it physically and chemically to ascertain how much it weighed and how pure it was. Thus, it was common in mining camps to weigh gold dust that was presented in payment. This is a nuisance, and it is natural that people would prefer to have the weighing and testing done in advance. Such preliminary weighing, testing, and certification of the results comprise the essence of coinage.

The function of coinage may be carried on either by the government or by private individuals. Of course, individuals may not legally copy the designs of the coins put out by the government, but if they wish to make coins of their own they may do so, and if other people wish to accept them as mediums of exchange they will be money. Actually, we have many historical examples of such privately minted coins, including several in the United States, some of which are pictured in Fig. 18 · 2. These coins circulated freely for a considerable period. The private coiner certified to the weight and purity of the metal and, as long as other people had confidence in his honesty, they were willing to accept the coin. The coin gained no value through the process of coinage. This process merely guaranteed that a given weight of metal of given purity was present, and the metallic content gave the coin its value.

It is perhaps natural that more people trust a government than trust any individual, although in monetary matters such trust has all too often been misplaced. Although coinage can be carried on by anyone, in all commercial countries the process has in practice become a public monopoly, since people accept government coins and do not accept others. But the process of coinage as carried on by a government is essentially the same as that carried on by private parties: it consists in certification as to the weight and purity of the metal in the coin. Usually, however, there are laws which regulate coinage by the government, which limit the quantities that can be issued and specify the way, if any, in which the coins are to be redeemed or exchanged for other kinds of money. These laws usually become even more important than the metallic content of the coins in determining their value in exchange.

In technical terms, the purity of the precious metals is referred to as their "fineness." If we say that a coin is 9/10 fine, we mean that it is 9/10 precious metal and 1/10 alloying metal. The gold and silver coins of the United States have long been 9/10 fine, 90 per cent gold or silver and 10 per cent copper. Sterling silver is 0.92 fine, or slightly finer than the metal in our silver coins.

The gold used in the arts, on the contrary, is much less pure than our coined gold. The purity of gold in the arts is measured by the carat, with one carat being 1/24. The highest quality gold used for jewelry and things of that sort is usually 14 carats, or 14/24 gold and 10/24 alloying metal. This commercial gold would thus be about 0.58 fine, as compared with that in our gold coins which is 0.9 fine.

Gold coins are not at present circulating in the United States, but under a presidential proclamation of Jan. 31, 1934, the monetary unit of the United States is the gold dollar defined as 15 5/21 grains of gold 9/10 fine. This corresponds to a value of $35 per ounce of pure gold, and the gold reserves of the United States Treasury are evaluated at this rate. "Standard gold" or "mint-fine gold" is not pure gold at all, but gold that is 9/10 fine and 1/10 copper.

The silver coins at present issued are the silver dollar, the half dollar, the quarter dollar, and the dime. These are 9/10 fine silver and 1/10 copper. Table 18 · 1 shows the weights of these coins. The last column gives a figure by which

Table 18 · 1. Silver Coins Issued by the United States

Denomination	Weight in grains		Value factor*
	Total	Silver	
Dollar	412.5000	371.2500	0.7734
Half dollar	192.9045	173 6140	0.3617
Quarter dollar	96.4522	86.8070	0 1808
Dime	38.5809	34.7228	0 0723

* This factor, multiplied by the current market quotation for silver in cents per ounce, will give the value of the silver in the corresponding coin.

one can easily compute the value of the metal in each coin from the current market price of silver. It is merely necessary to multiply that price in cents per ounce by the factor in the last column of the table. Thus, for example, if the market price of silver is 75¼ cents per ounce, as quoted in the financial section of the *New York Times* on Mar. 27, 1947, the silver in a half dollar is worth 75¼ times 0.3617, or just over 27 cents.

The so-called "minor coins"—the nickel and the cent—have had their composition altered somewhat during the second World War, but prior to that time the nickel had weighed 77.16 grains of which 75 per cent was copper and 25 per cent was nickel, and the cent had weighed 48 grains of which 95 per cent was copper and 5 per cent was tin and zinc.

18 · 4. Standard Money. We have a number of kinds of money in the United States—gold, silver, nickel, and copper coins, and several types of paper money. Although most countries have had similar heterogeneous systems, many of them have had one kind of money which was peculiar in

certain respects, and in terms of which all the other money was valued. In the United States that special kind of money was gold, and we said that the United States was on a "gold standard." At one time most important commercial nations were on the gold standard, although some have had a silver standard. Almost any kind of money could be standard money, depending on its characteristics. *Standard money has three identifying characteristics: it enjoys free coinage, it has unlimited legal-tender power, and it may be freely imported and exported.* And money with all three of these characteristics is standard money; any money that lacks one or more of them is not standard money.

When we say that a given kind of money enjoys "free coinage," we mean that the mints are required to coin all of it that is brought, without any limit on the amount. We do not mean that they are necessarily required to coin it for nothing, as the word "free" might seem to imply. If no charge is made for the coining, we call it "gratuitous coinage," so that coining may be both gratuitous and free, as it was in the United States when we were on the gold standard. On the other hand, some countries have followed the practice of requiring the mint to coin all of the specified metal that was brought, but have made a charge that just covered the cost of doing so. Such a charge, if equal to the cost involved, is called "brassage" and, as long as the charge is kept at this level, we would still call the system one of free coinage, since there is no limit on the quantity coined. On the other hand, when a government makes subsidiary coins, it is common for the mint to buy the necessary metal on the market as it is needed, and to put into the coins a smaller value of metal than the face value of the coins. Thus, the United States cent contains far less than a cent's worth of copper, and the dime far less than 10 cents' worth of silver. The profit that the government makes in producing such lightweight coins is called "seigniorage." These latter lightweight coins, however, are not standard money, since the mint is not required to coin them when requested. We have free coinage only when the mints are required to coin all of the metal that is presented, either without charge or at a charge no larger than the actual cost of the work.

When we speak of "legal tender," we are talking of a legal offer to pay a debt, the word "tender" meaning "offer," as in the expression "to tender a resignation." If I owe a man $10, how can I pay him? Can I offer him $10 worth of tomato catchup and assert that I have paid the debt? I cannot, for the law would say that I had not made a legal tender—a legal offer to pay the debt. Catchup does not have legal-tender power under our laws. *Those things are legal tender which may be offered in payment of any debt incurred in monetary terms, and which must be accepted by the creditor if he is to accept anything.* The creditor may refuse my offer of tomato catchup and may insist on something else in payment, since tomato catchup is not legal

tender. If I offer him my check for $10, he may refuse that, too, and insist on some other sort of payment. The law does not require creditors to accept personal checks. If I offer him an ordinary ten-dollar bill and he turns it down, I need offer him nothing else. Of course, the law will not insist that he accept anything. He need not be paid if he prefers not to. But if he is to be paid, he will have to take the ten-dollar bill or nothing, since the ten-dollar bill has legal-tender power.

Until the early 1930's there were many varied sorts of legal-tender restrictions on the money of the United States. Some kinds had no legal-tender power at all. Creditors might accept them if they wished, but they were not required to do so. The fractional silver coins (half dollar, quarters, and dimes) were legal tender only up to amounts of $10, and the minor coins (nickels and cents) only to the extent of 25 cents. This situation was greatly simplified by the act of June 5, 1933, which stipulated in part that, "all money and currencies of the United States (including Federal Reserve notes and circulating notes of the Federal Reserve Banks and national banking associations) heretofore or hereafter coined or issued, shall be legal tender for all debts, public or private . . . ." Thus at present all of our money has unlimited legal-tender power, although it seems likely that we shall sooner or later return to the policy of limiting the legal-tender power of minor coins to thwart the practical joker who insists on paying for his new automobile in pennies.

When the United States was on the gold standard, some money had unlimited legal-tender power and some did not, but only gold enjoyed the privilege of free coinage. In addition, gold had unlimited legal-tender power. As a result, gold coins had a peculiarity that is shared by all standard money; namely, the face value of gold coins was exactly the same as the value of their contents. In those days the dollar was defined as 25.8 grains of gold 9/10 fine. It contained 23.22 grains of gold and 2.58 grains of copper. A ten-dollar gold piece contained 232.2 grains of gold, and that much gold was worth just exactly $10. Perhaps it would be more exact to say that much gold was $10, since a dollar was defined as 25.8 grains of 9/10 fine gold. It is impossible for the face value and the value of the contents to vary with standard money because, if the contents are worth more than $10, people melt down the coin; if the coin is worth $10 and the material worth, say, $9, it pays people to take the $9 worth of metal and have it coined. The mint is required to coin it, so that there is no problem on that score. When gold is standard money, it is possible to convert freely from bullion to coin or from coin to bullion, so the face value and the value of the contents must always be identical. This is a characteristic of all standard money.

From a purely domestic standpoint, within any given country, free coinage and unlimited legal-tender power are the only things necessary to estab-

lish standard money; when we come to the problems of international trade and finance,[1] we must add the requirement that to be standard money a commodity must enjoy freedom of import and export. A country might remain on the gold standard at home, with free coinage and unlimited legal-tender power, but forbid the exportation of gold. Under such circumstances, gold would not be available for the payment of debts owed abroad, and the country would have left the gold standard as far as its international transactions were concerned.

18 · 5. Bimetallism. At present, relatively few countries have any standard money. They have left the gold standard, for reasons that we can best discuss somewhat later. Before the first World War most of the commercial nations of the world were on the gold standard, and earlier than that there was a time when many countries had two or more kinds of standard money at once. *Bimetallism* is a monetary system in which there are two kinds of standard money circulating at a fixed ratio established by law. In practice, the two metals have ordinarily been gold and silver, and we shall illustrate the problems of bimetallism with them.

When we say that gold and silver are, under bimetallism, both standard money, we mean, of course, that they both enjoy free coinage and that a debtor may pay his debt in either metal he chooses, since each has unlimited legal-tender power. When we say that the two metals circulate at a fixed ratio established by law, we mean merely that the law must specify how much of each metal it takes to pay a given debt—how much gold there is in a gold dollar, and how much silver in a silver dollar. For example, when the United States was on a bimetallic basis in the middle of the nineteenth century, the law defined the gold dollar as 25.8 grains of 9/10 fine gold, containing 23.22 grains of pure gold. The silver dollar was defined as 412.5 grains of 9/10 fine silver, containing 371.25 grains of pure silver. Thus the silver dollar weighed approximately 15.988 times as much as the gold dollar, usually referred to as "16 to 1." This ratio of the weights of the two standard coins is called the "mint ratio," since it is the ratio of the weights at which the coins are minted, or the "legal ratio," since it is the ratio at which debts can legally be paid. One who owes money can pay either with a given amount of gold or with sixteen times as much silver. Under a bimetallic system, there must be some mint ratio. If the two metals were given equal debt-paying power, the ratio would be 1 : 1. In practice, gold has been considerably more valuable than silver; therefore the silver dollar has been heavier than the gold dollar.

18 · 6. Gresham's Law. No difficulty would arise from bimetallism if the mint ratio were actually equal to the relative values of the contents of the two coins. But the value of the two metals, gold and silver, is determined, like other values, in the immediate short run by supply and demand, and in the

[1] See Chaps. 23 and 24, especially the latter.

long run by costs of production. The ratio of the value of the metals on the market is called the "commercial ratio" or the "market ratio." Thus, if we say that the market ratio is 16 : 1, we mean that it takes 16 ounces of silver to purchase 1 ounce of gold when we buy the metals on the market. When the mint ratio and market ratio are both 16:1 or when they coincide at any other figure, bimetallism works as intended. However, even if the mint ratio is originally set in the law exactly equal to the market ratio, the latter will soon change as conditions of supply, demand, and cost change. Thus, in practice the mint ratio and the market ratio soon come to differ, and under those circumstances one of the metals is always driven out of circulation. The reason for this result is easy to see if we take an hypothetical example.

Let us suppose that the law has set the mint ratio at 16:1, with the silver dollar weighing sixteen times as much as the gold dollar; but that on the market it really takes but 15 ounces of silver to buy 1 ounce of gold. The mint ratio is 16:1 and the market ratio is 15:1. Let us suppose that a man owes a debt that can be paid with either 1 ounce of gold or 16 ounces of silver. It is clear that he will not pay the debt in silver, because if he has 16 ounces of silver with which to pay the debt, he can take 15 ounces of it to the market and buy 1 ounce of gold. Since there is free coinage, he can take the gold to the mint, have it coined, and pay his debt with it. This leaves him an ounce of silver. If he had paid the debt in silver, he would have had to pay his entire 16 ounces. If he had had an ounce of gold with which to pay the debt, there would have been no advantage in exchanging it first for silver on the market, since he could exchange it for but 15 of silver, which would not be enough to pay the debt. If he has 1 ounce of gold, he pays the debt in gold. If he has 16 ounces of silver, he exchanges 15 of them for 1 ounce of gold and again pays in gold. With these ratios, it always pays him to settle any debt with gold.

If, on the other hand, the mint ratio had been 16:1 and the market ratio had been 17:1, the situation would have been reversed. Under these circumstances, if the debtor had the 16 ounces of silver necessary to pay the debt, he would be foolish to exchange them for gold, since it takes 17 ounces of silver now to buy 1 ounce of gold, and his 16 ounces of silver will not buy enough gold to pay the debt. If he has the silver, he is better off to pay his debt in silver. If he has an ounce of gold with which to pay the debt, he will find it more advantageous to exchange it for 17 of silver on the market, use 16 of silver to pay the debt, and retain the other ounce of silver. This time we find that, if he has the silver, he pays his debts in it; if he has the gold, he first exchanges it for silver and then pays the debt in silver. With these ratios all debtors find it to their advantage to settle their debts in silver. Gold is not used as money in payment of debts but is withdrawn from circulation and used on the market to exchange for silver. The silver drives the gold out of circulation.

The difference between the two cases is easy to identify, so that we can tell for the general case what will happen, instead of merely for the two cases with which we have already illustrated. In the first case, the actual values of the two metals were 15 ounces of silver to 1 ounce of gold; but the mint had set the ratio at 16 ounces of silver to 1 of gold. The mint was saying in effect than 1 ounce of gold was worth 16 ounces of silver when it was really worth but 15. The mint was setting too high a value on gold—was overvaluing it—and as a result only gold was used as money, and silver was driven out of circulation. In the second case, the actual value of the two metals was 17 ounces of silver to 1 of gold; but the mint had set the ratio at 16 ounces of silver to 1 of gold. The mint was saying in effect that 16 ounces of silver were worth 1 of gold, while in fact it took 17 ounces of silver to be worth this much. The mint was setting too high a value on silver—was overvaluing it—and as a result only silver was used as money, and gold was driven out of circulation. We see here two illustrations of the general principle, which is usually named after one of its early formulators, Sir Thomas Gresham, and called "Gresham's law": *Under a bimetallic monetary system, when the mint ratio and the market ratio differ, the metal that is overvalued at the mint tends to circulate to the exclusion of the other.* When the mint overvalues a metal, people take that metal to the mint, since it is there that it has the greatest value. The other metal they sell or exchange on the market rather than at the mint, since it has a relatively higher value on the market.

The working of Gresham's law under conditions of bimetallism is easy to understand. Debtors are given two ways to pay a debt, since both kinds of money have unlimited legal-tender power. Since the creditor has to accept either kind of money that is offered, the debtor naturally offers the one that is to his own advantage, the cheaper one. If one of the metals depreciates in value, debtors always tend to use it. Thus, if we assume that the market ratio was originally 16:1 and that it changed to 15:1, we see that gold has depreciated in value. One ounce of gold used to be worth 16 of silver and is now worth but 15. This is the case in which we discovered that debts would be paid in gold. If, on the other hand, the market ratio changes from 16:1 to 17:1, we see that silver has depreciated. It once took but 16 ounces of silver to buy 1 of gold; it now takes 17 ounces of silver to buy 1 of gold. Silver is relatively less valuable. This is the case in which we discovered that debtors chose to pay all debts in silver. Since the debtor has his choice, he always pays in the cheapest way.

Gresham's law is really somewhat broader than the statement of it given in italics above. It is commonly stated in some such terms as these: *Gresham's law: Bad money tends to drive out good money.* We have seen that under bimetallism debtors always use the depreciated or "bad" money. But it is

not necessary to have bimetallism for Gresham's law to operate. Most people, on getting a half dollar that appears to them possibly counterfeit, spend it at once in preference to other half dollars that appear good to them. They spend the "bad" money and keep the "good." People living along the northern border of the United States, and finding a Canadian dime in their change, tend to spend it as soon as they can, getting rid of it quickly instead of spending other coins that they think are better. Again they spend the "bad" or questionable money and keep the "good." In the case of bimetallism both kinds of money, to be sure, are "good" in the legal sense; according to law they are equal. But the debtors themselves do not think of them as equal, and they spend the ones that are the cheaper. If the mint ratio and the market ratio are the same, then there is no advantage in trading one metal for another on the market. If the two ratios differ, debtors will find it always to their advantage to spend the depreciated or "bad" money and keep the "good" money for the purpose of exchange on the market.

18 • 7. Gresham's Law in Operation. The typical student is so accustomed to accepting his money as all equal that it is hard for him to realize that people ever actually do, in practice, distinguish between their different dollars, paying some in perference to others. Although the numerical illustrations of the preceding section were made up to suit the case, nevertheless they do illustrate the kind of thing that has happened over and over again in real life.

Proponents of bimetallism have often urged that the system would work because, if either of the metals began to depreciate, there would be an added demand for it for monetary uses, thus raising its value, while the metal that had appreciated in value would be driven out of monetary uses and would thus become plentiful on the market, thus reducing its value. The supposition was that this *compensatory action* would keep the market ratio from drifting away from the mint ratio. Reference to Table 18 • 2 will show, however, that the market ratio has never stayed fixed for very long. Moreover, the history of every country that has tried bimetallism demonstrates not only that the two ratios do drift apart, but that they drift far enough apart so that actually only one metal is used as money.

The first coinage system of the United States was set up by the coinage act of 1792. It established a bimetallic system at the ratio of 15:1, which was at the time almost exactly the market ratio (see Table 18 • 2). Soon, however, silver began to depreciate somewhat, and as a result gold was driven out of circulation in accordance with Gresham's law. The country continued for some years legally on a bimetallic basis but actually on a silver basis. In June, 1834, Congress altered the mint ratio, making it 16.002:1. In 1837, another slight change was made in order to standardize both coins at 9/10 fine, and

Table 18·2. Commercial Ratio of Silver to Gold, Annually from 1687-1944*

Year	Ratio	Year	Ratio	Year	Ratio	Year	Ratio	Year	Ratio	Year	Ratio
1687	14 94	1731	14.94	1775	14 72	1819	15.33	1863	15.37	1907	31.24
1688	14 94	1732	15.09	1776	14.55	1820	15.62	1864	15.37	1908	38.64
1689	15.02	1733	15.18	1777	14 54	1821	15.95	1865	15.44	1909	39.74
1690	15.02	1734	15.39	1778	14.68	1822	15.80	1866	15.43	1910	38.22
1691	14.98	1735	15.41	1779	14.80	1823	15.84	1867	15.57	1911	38.33
1692	14.92	1736	15.18	1780	14.72	1824	15.82	1868	15.59	1912	33.62
1693	14.83	1737	15.02	1781	14 78	1825	15.70	1869	15.60	1913	34.19
1694	14.87	1738	14.91	1782	14.42	1826	15.76	1870	15.57	1914	37.37
1695	15.02	1739	14 91	1783	14.48	1827	15.74	1871	15.57	1915	40.48
1696	15.00	1740	14 94	1784	14.70	1828	15.78	1872	15.63	1916	30.78
1697	15.20	1741	14.92	1785	14.92	1829	15.78	1873	15.92	1917	24.61
1698	15.07	1742	14 85	1786	14.96	1830	15.82	1874	16.17	1918	21.00
1699	14.94	1743	14 85	1787	14.92	1831	15.72	1875	16.59	1919	18.44
1700	14 81	1744	14 87	1788	14 65	1832	15 73	1876	17.88	1920	20.28
1701	15.07	1745	14.98	1789	14.75	1833	15.93	1877	17.22	1921	32.76
1702	15.52	1746	15.13	1790	15 04	1834	15.73	1878	17.94	1922	30.43
1703	15.17	1747	15.26	1791	15.05	1835	15.80	1879	18.40	1923	31.69
1704	15.22	1748	15.11	1792	15.17	1836	15.72	1880	18.05	1924	30.80
1705	15.11	1749	14 80	1793	15 00	1837	15.83	1881	18.16	1925	29.78
1706	15.27	1750	14 55	1794	15.37	1838	15.85	1882	18.19	1926	33.11
1707	15.44	1751	14 39	1795	15.55	1839	15.62	1883	18.64	1927	36.47
1708	15 41	1752	14 54	1796	15.65	1840	15.62	1884	18.57	1928	35.34
1709	15.31	1753	14.53	1797	15.41	1841	15.70	1885	19.41	1929	38.78
1710	15 22	1754	14 48	1798	15.59	1842	15.87	1886	20.78	1930	53.74
1711	15 29	1755	14 68	1799	15.74	1843	15.93	1887	21.13	1931	71.25
1712	15.31	1756	14 94	1800	15.68	1844	15.85	1888	21.99	1932	73.29
1713	15.24	1757	14 87	1801	15.46	1845	15 92	1889	22.10	1933	59.06
1714	15 13	1758	14.85	1802	15.26	1846	15 90	1890	19.76	1934	72.49
1715	15 11	1759	14.15	1803	15.41	1847	15 80	1891	20.92	1935	54.19
1716	15 09	1760	14.14	1804	15.41	1848	15.85	1892	23.72	1936	77.09
1717	15 13	1761	14.54	1805	15 79	1849	15.78	1893	26.49	1937	77.44
1718	15 11	1762	15.27	1806	15 52	1850	15.70	1894	32.56	1938	80.39
1719	15.09	1763	14.99	1807	15 43	1851	15.46	1895	31.60	1939	88.84
1720	15 04	1764	14.70	1808	16.08	1852	15.59	1896	30.66	1940	99.76
1721	15 05	1765	14.83	1809	15.96	1853	15.33	1897	34.20	1941	99.73
1722	15.17	1766	14.80	1810	15.77	1854	15.33	1898	35.03	1942	90.57
1723	15.20	1767	14.85	1811	15.53	1855	15.38	1899	34.36	1943	77.67
1724	15.11	1768	14.80	1812	16.11	1856	15.38	1900	33.33	1944	77.67
1725	15.11	1769	14.72	1813	16.25	1857	15.27	1901	34.68	1945	
1726	15.15	1770	14.62	1814	15.04	1858	15.38	1902	39.15		
1727	15.24	1771	14 66	1815	15.26	1859	15.19	1903	38.10		
1728	15 11	1772	14.52	1816	15.28	1860	15.29	1904	35.70		
1729	14 92	1773	14.62	1817	15.11	1861	15.50	1905	33.87		
1730	14.81	1774	14.62	1818	15.35	1862	15.35	1906	30.54		

* "Annual Report of the Director of the Mint," U. S. Treasury Department, fiscal year ended June 30, 1944, p. 91

the ratio then became 15.988:1. These last two ratios were almost identical,[1] and in practice men have referred to the ratio since 1834 as 16:1. Reference to Table 18 • 2 will show that this ratio went slightly too far in the other direction and overvalued gold. As a result, silver was driven from circulation. Although the country was still legally on a bimetallic basis, actually it was on a gold basis. Thus even a slight difference between the mint and the market ratios was sufficient to drive the more expensive metal out of monetary use.

In 1873, the coinage laws were again revised, and at that time the silver dollar, which had practically disappeared from circulation anyway, was dropped from the list of coins that could be minted. Thus, after 1873, the United States was really on a gold standard, although the law did not explicitly so state until 1900. Strangely enough at almost the time when the silver dollar was dropped, discoveries of new silver deposits in the West brought about a depreciation in the value of silver to such an extent that it would have been profitable again to coin it in preference to gold by 1874 or 1875 (see Table 18 • 2). Silver producers, discovering that this source of profit had been taken from them, asserted that the dropping of the silver dollar in 1873 had been engineered by Wall Street speculators through a conspiracy with Congressmen, and they called the act itself "the Crime of '73," a political slogan which became widely known in the free silver period that followed. The change in the market ratio, taken in conjunction with a long period of painful financial depression, led to much discontentment and to strong agitation for a return to bimetallism. This movement was supported in part by silver producers who had selfish interests in having silver monetized again, and in part by debtor classes who always stand to gain by depreciation of money.[2] In addition, times of depression are always times during which un-

[1] The actual figures for the three cases described were the following:

1. Act of 1792:
   Gold dollar contained 24.75 grains of pure gold and 2.25 grains of alloy. Total weight, 27 grains. Fineness, 0.91667. The silver dollar contained 371.25 grains of pure silver and 44.75 grains of alloy. Total weight, 416 grains. Fineness, 0.892. Mint ratio, 15:1.
2. Act of 1834:
   Gold dollar contained 23.2 grains of pure gold and 2.6 grains of alloy, making a total weight of 25.8 grains, 0.899225 fine. Silver dollar remained as in act of 1872. Mint ratio, 16.002:1.
3. Act of 1837:
   Gold dollar contained 23.22 grains of pure gold and 2.58 grains of alloy, making a total weight of 25.8 grains, 0.900 fine. The silver dollar contained 371.25 grains of pure silver and 41.25 grains of alloy, making a total weight of 412.5 grains, 0.900 fine. Mint ratio, 15.988:1.

[2] For an explanation of this fact, see Sec. 21 • 6.

employed and debt-ridden sufferers are "willing to try anything," without regard to its reasonableness or ultimate desirability. The movement to return to bimetallism was headed by William Jennings Bryan, and the presidential campaign of 1896 was fought mainly on the issue of bimetallism under the campaign slogan of "16 to 1." The bimetallists lost that campaign and, although there are still advocates of bimetallism in Congress, the issue has not been one of major importance politically since that year. The silver interests have found other ways of gaining their advantage at public expense, as we shall see, and those who wished to depreciate the currency have found other means more effective than bimetallism.

18·8. Classification of Money. We have described standard money and discussed some of the problems that arise in connection with it. Let us now turn to the other kinds of money usually making up the major part of the circulating medium even when a country has standard money. Although no classification is entirely satisfactory for all purposes, we shall find it useful to divide money into the following classes:

I. Government money:
  A. Coins.
    1. Standard money.
    2. Subsidiary coins.
  B. Paper money.
    1. Representative money.
    2. Credit money.
    3. Fiat money.
II. Bank notes:
  A. Issued by a central bank.
  B. Issued by other banks.
III. Private money.

18·9. Subsidiary Coins. Standard coins are sometimes called "full-bodied coins" because they contain material worth the full face value. Most countries, even if they have standard coins, have issued other coins, especially in small denominations, made of nonstandard metal and having a face value greater than the value of the metallic contents. Silver has commonly been used for such subsidiary coins in nations that were on the gold standard, although the coins of smallest denomination have been made of nickel, copper, bronze, and other materials. Since these coins do not contain enough metal to be worth their full face value, the question arises how their value is maintained. If, for example, a dime has a face value of 10 cents but the metal in it is worth only 4 cents, why are people willing to give 10 cents' worth of goods in exchange for it? Subsidiary coins obtain their value from two things: (1) from the fact that their issue is limited in amount and (2) from the fact that the government exchanges other kinds of money for them. Scarcity and con-

vertibility combine to give them their artificially high value. If the government offered to coin into dimes all the silver that was presented, as is done with standard money, it would be to the advantage of the public to buy silver for 4 cents on the market and have it made into dimes. But the government does not coin subsidiary coins on demand. When the authorities feel that more small change is needed, they buy the necessary metal on the market themselves and coin only the quantity that they think wise. We have already said that the profit made on this coinage is called "seigniorage."

If the country issuing the subsidiary coins is on the gold standard, it stands ready to exchange gold for its subsidiary coins; and, since holders of these coins can obtain gold for them, the coins have as much value as the equivalent amount of gold. Of course, if the government did not limit the amount of coinage of these subsidiary coins, they could not maintain convertibility. The limitation of the issue of subsidiary coins to the amount needed in trade is fundamental to the maintenance of their full value.

If the country issuing the subsidiary coins is not on a gold standard, it maintains convertibility of them for whatever kinds of money are in use, and the value of the subsidiary coins is thus kept on a par with the value of these other kinds of money.

Subsidiary coins are made lightweight perhaps in some small part so that the government may make a profit from their coinage, but the major reason is to prevent their being melted down for the metal in them. If a half dollar had today a full half-dollar's worth of silver in it, even a small change in the market value of silver might make the contents worth, say, 51 cents. Then people who wanted silver for nonmonetary uses would find that they could melt down half dollars and get 51 cents' worth of silver by spending 50 cents. Under such circumstances the coins would disappear.

Subsidiary coins usually have only limited legal-tender power, since payments of large quantities of small coins are a nuisance to the recipients. At present the subsidiary coins of the United States enjoy full legal-tender power, but it is reasonable to suppose that restrictions will again be placed on this power in the near future. The fiscal authorities of the United States divide the subsidiary coins into fractional silver (the half dollar, the quarter, and the dime) and minor coins (the nickel and the cent). The term "minor" is usually reserved for those coins of such small denomination that they have to be made out of one of the baser metals rather than out of gold or silver.

18 · 10. Paper Money. In Sec. 18 · 2, when we were discussing the qualities of a good money material, we said that it was desirable to use something which was relatively scarce and which had large value in a small bulk. It would seem as though paper would depart from these specifications about as far as anything could, yet paper money has been used for centuries. As was mentioned in Sec. 18 · 1, Marco Polo found the Chinese using paper

money in the thirteenth century, and it has been said that the first paper money was put into use more than a century before the birth of Christ.

Paper money has the advantage of convenience and portability. A ten-thousand-dollar bill is no heavier and no larger than a dollar bill. The use of paper makes it unnecessary to carry and store large quantities of coin, and saves the wear and tear and danger of loss of the metal. As a partial offset to these advantages, paper money has the disadvantage that there is no inherent limitation on the quantity issued as there is when the money material is scarce and costly. Since the value of the paper itself is negligible as compared with the face value, paper money circulates entirely on the basis of credit—on the faith and confidence of the people who use it. This confidence may rest on the belief that it will be redeemed in other sorts of money either on demand or at some later date, or it may rest merely on custom and habit—on the belief that the holder of the money will be able to exchange it for valuable commodities on the market. Experience shows that people do continue to use and accept paper money long after convertibility ceases, and in such cases its value depends on the limitation of the quantity issued.

The government has two ways in which it can encourage its citizens to accept and use paper money that is not convertible. In the first place, it can make it legal tender. Then creditors who are owed money must accept the paper money or nothing at all and, even though the paper money may be depreciated, it may well be better than nothing at all. Of course, the fact that paper money is legal tender does not force people to accept it. They do not need to incur their debts in monetary terms. Prior to the abandonment of the gold standard in 1933, citizens of the United States might specify the kind of money that they would accept, and many contracts were actually so drawn that the creditor had a right to demand payment in gold. He might, similarly, have insisted in his contract on payment in silver, in nickels, or in some particular kind of paper money. Since the act of 1933 was sustained by a questionable decision of the Supreme Court in 1935, the right to insist on payment of monetary debts in particular kinds of money is no longer enforceable. Still, one who wished to avoid the use of paper money that he distrusted could refuse to enter into contracts calling for payment of money. He might insist on payment in goods—on barter. However, the difficulties involved in any such procedure are evident. In practice, when the government gives legal-tender power to paper money, it does in effect force people to accept it.

The government may also encourage the acceptance and use of inconvertible paper money by accepting the money itself in payment of taxes or other dues. Yet paper money has circulated which neither had legal-tender power nor was acceptable by the government at par, as in the case of certain of the bank notes issued by private banks before the Civil War. When people

get in the habit of using paper money, they continue to use it even when it is inconvertible, and even when they are not forced to accept it by legal-tender provisions. In such cases, nothing can stop the depreciation of the money if it is issued in too large quantities.

It may be of some historical interest in passing (though hardly of current practical importance) to note that there was some question as to whether or not the United States had the constitutional right to issue paper money. A proposal that it be given such a right was made in the Constitutional Convention, and there voted down. For well over half a century the question was moot until the Supreme Court ruled in 1869 that the "greenbacks" issued by the Federal government during the Civil War were constitutional.

18 • 11. Representative Money. Sometimes paper money is issued as a warehouse receipt for gold or silver coins deposited in the Treasury and there held in trust for the redemption of the paper money. Our silver certificates and gold certificates are money of this kind. A silver certificate, for example, carries some such statement as the following: "This certifies that there is on deposit in the Treasury of the United States of America Five Dollars in silver payable to the bearer on demand." Such money is called "representative money" since it represents other pieces of money and circulates in their place. It is essentially a method for allowing gold and silver to serve as means of payment while retaining some of the advantages of paper money.

Gold certificates were first issued in 1863 but had no legal-tender power until 1919. They were originally intended to be issued only in large denominations, and until 1907 were limited to sizes of $20 and over. Since that time, they have been issued in all denominations of $10 and over. When the United States left the gold standard in 1933, these certificates were withdrawn from circulation and are at present used only by the Federal Reserve Banks. Almost the entire outstanding amount is now in $100,000 bills.[1] The gold certificates, formerly backed by gold coin, are now backed by a 100 per cent reserve in gold bullion.

Silver certificates have been issued since 1878, because the large size and weight of silver dollars made them inconvenient for day-to-day use. They have largely replaced silver dollars in circulation except in some parts of the West where people have preferred "hard money." Most of the small bills in circulation today are silver certificates, which are issued in denominations from $1 to $1000. They had no legal-tender power whatever until 1933, since when, as we have learned, all kinds of American money have had unlimited legal-tender power. Silver certificates, once backed entirely by coined silver dollars, are now backed by both coin and bullion of a monetary value equal to the face value of the certificates. Since the market value of the metal in a

[1] See Table 18 • 3.

silver dollar is considerably below the face value, these certificates really have but a partial backing and to that extent are credit money; gold certificates are backed by full-bodied money and, when really redeemable, are not credit money.

18·12. Credit Money. Credit paper money is also known as fiduciary, redeemable, or convertible paper money. It consists of promises to pay standard money on demand, but is only partly backed by reserve. Instead of being backed dollar for dollar, as are the gold and silver certificates which merely

Table 18·3. Denominations in Which United States Paper Money Is Outstanding, and Amounts Outstanding, May 31, 1946*

Denomination	Gold Certificates	Silver Certificates	Treasury Notes of 1890	United States Notes	Federal Reserve Notes	Federal Reserve Bank Notes	National Bank Notes	Total
$1	—	X	X	o	—	o	o	o
2	—	o	X	X	—	o	o	o
5	—	X	X	X	o	o	X	o
10	o	X	X	o	X	X	X	X
20	o	o	o	o	X	X	X	X
50	o	o	o	o	o	X	o	o
100	o	o	o	o	X	X	X	X
500	o	o	—	o	o	—	o	o
1,000	o	o	o	o	o	—	o	o
5,000	o	—	—	—	o	—	—	o
10,000	o	—	—	o	o	—	—	o
100,000	X	—	—	—	—	—	—	o
Total outstanding, billions.....	18.1	2.2	0.0	0.3	24.7	0.5	0.1	45.9

* The symbols in the body of the table have the following meaning
— means none outstanding.
o means that this denomination is outstanding, but the total amount is less than 10 per cent of the total of this kind of money outstanding.
X means that this denomination is outstanding, and that it makes up more than 10 per cent of the total of this kind of money outstanding.
Data are from Treasury Department's monthly statement and *Federal Reserve Bulletin.*

replace coins, it represents an actual net addition to the amount of money outstanding. The government promises to pay when demanded and keeps a certain amount of lawful money on hand to take care of those cases in which such payment is desired. As long as these notes are actually redeemed, they are credit money but, if the government fails to redeem them, they become fiat money. Thus a given bill may at one time be fiat, when the government refuses to convert it, and later be credit money when the government resumes payment. Coins may also be credit money. When a coin has a face value in excess of the value of its contents, it is clear that the holder of

the coin has not received full value but relies on the credit of the government to make good the difference.

The most important kind of paper credit money in the United States is the so-called United States note. These were first issued during the Civil War in 1862, and at that time were fiat money (see Sec. 18 • 13). By 1864 there were about $450 million of them in circulation, but since 1878 the number has been constant at just under $347 million. The law requires that they be reissued whenever they are redeemed. These notes, known as "greenbacks" at the time of their issue, were promises to pay, but actually the government did not redeem them until 1879. Thereafter the government did redeem them on demand as long as the country was on the gold standard; yet there was not a 100 per cent reserve behind them. Since 1900 the law has required the Treasury to maintain a reserve of $150 million behind this $347 million of notes, so that the reserve amounts to about 43 per cent. As long as the notes were actually redeemed, they were credit money. They derived their value from the fact that other types of money could be obtained in exchange for them, and, of course, their quantity was limited. We shall comment further on these notes in Sec. 18 • 13.

The government has in the past issued other sorts of credit paper money, and still carries on its circulation reports a small quantity of Treasury notes of 1890. These notes were issued to pay for silver purchased under the Sherman Silver Purchase Act (see Sec. 18 • 17), but since 1900 they have been canceled as fast as they come in to the Treasury and only about $100,000 of them are now outstanding on its books. The quantity outstanding has changed hardly at all for the past few decades. Presumably the notes listed as still outstanding are either lost or in the hands of collectors. Certainly these notes are no longer plentiful enough for the student to reckon with in our present circulation, and their importance is now almost entirely historical. They are reserved 100 per cent with silver dollars but, since the silver dollars are themselves not full value, we list these notes here as credit money, or at least would have at the time when they could have been redeemed in standard money at least indirectly.

18 • 13. Fiat Money. Fiat money is money that does not have full value in itself and cannot be redeemed for any full bodied money. It is not standard money and cannot be converted into standard money. Such money always in practice carries a promise to pay, so that one cannot determine its fiat character by inspecting it. Thus the United States notes, or "greenbacks," have always carried a statement that "The United States of America will pay to the bearer on demand Five Dollars," or whatever amount of money corresponded with the denomination of the bill. Unfortunately, the promises of governments on their money have seldom been trustworthy over any considerable period. Despite the promise to pay, the government did not redeem

these notes from the time of their original issue in 1862 until 1879. During these 17 years, the money was fiat. Then the government did actually start to redeem them, setting up a partial reserve for that purpose as we saw in the preceding section. Notes that had been fiat suddenly became credit money. There was no physical change in the notes themselves. To discover whether or not money is fiat we do not inspect it, we ascertain whether or not the government will in fact exchange it for some type of fullbodied or standard money. Mere exchangeability is not enough. If the government exchanges one kind of fiat money for another kind, it is still fiat. Unless one can obtain full-value money for fiat money, either directly or indirectly, it is still fiat money.

It is a strange commentary on morals in monetary matters that governments always give fiat money full legal-tender power. The very best money (standard money) and the very worst money (fiat money) always have this power. The reason for giving full legal-tender power to standard money is evident: if the creditor is to be forced to accept anything, it seems reasonable to force him to accept a kind of money that carries full value in itself. The reason for giving full legal-tender power to fiat money is very different. In this case, the government gives the money full legal-tender power to force people to take it who would prefer other kinds of money if given a choice.

Governments are costly to operate, and in times of great financial need, as during war or depression, they sometimes feel that the usual sources of funds (taxation, borrowing, etc.) are insufficient to meet their needs. Under such circumstances they have often resorted to the printing press, turning out fiat money which they force on their citizens by the legal-tender provision. It seems fairly well established that fiat money can circulate and retain a reasonable value despite the fact that it is not redeemable if the government limits the quantity of it sufficiently. Money is a great convenience—it has utility—and, if it also has scarcity, it will be valuable for the same reasons that other economic goods are valuable. Usually, however, the very exigencies that have caused the government to issue fiat money in the first place encourage it to issue more and more. As more of the money is issued, its value falls, and the government finds itself in the position of needing to issue ever more and more. Under such conditions, there is nothing that can maintain the value of the money, and it depreciates. Thus, during the Civil War the value of the greenbacks depreciated until they were worth roughly 35 cents on the dollar. Merchants quoted two prices on their goods, one in gold and the other in greenbacks. For a time during 1864, it took approximately $2.85 in greenbacks to buy what one could get with $1 in gold. In other countries and at other times depreciation has gone much further, as in the case of the German mark after the first World War. This paper money depreciated until it finally took 1 trillion paper marks to equal 1 gold mark. In July, 1946, the

Hungarian pengo fell to the point where it took 500 quintillion of them (that is, 5×10^{20}) to equal $1, as compared with a value of 29.12 cents to the pengo in June, 1939.

If we stick strictly to our definitions, we would have to say that all of the present money of the United States is fiat. None of it can now be redeemed for standard money. Except under certain restrictions, no one can get full value in gold or in any other kind of money in exchange for his monetary holdings. It is true that businessmen can exchange their money for gold for shipment abroad under certain restrictions, but even this is a matter of privilege rather than of right. One can also buy gold for bona fide use in industry or the arts, thus "converting" his money into gold.

Although the repudiation of government contracts, which was part and parcel of our departure from the gold standard seems to have involved the sort of barefaced dishonesty that has characterized the monetary policies of governments from the earliest times, nevertheless we discover again that society can operate reasonably well with fiat money as long as the quantity issued is carefully controlled. The major advantages, as well as the major dangers, of fiat money spring from this possibility of human control. The country operating a monetary system on the gold standard finds that the value of its money is irrevocably tied to the value of gold. It cannot be controlled for social advantage, but similarly it is not possible to manipulate it for party or class advantage. Differences of opinion on the question of the desirability of fiat money as compared with the desirability of the gold standard reflect largely the holder's faith in human motives and judgment or his lack of it.

18 · 14. Bank Notes. As we have said, governments are not the only issuers of money. Private individuals and business firms have made money from the earliest times, and actually the substantial monopolization of the issue of money by governments is a rather recent thing. Among the non-governmental agencies issuing money, the most important have been banks. These may be either private banks or those which are operated by the government. In the United States it was long true that the only paper money issued was that put out by private banks chartered by the states. It was clear that the states had no right to issue paper money themselves, and it was thought that the Federal government also was denied that right in the Constitution; but the states, which could not themselves issue money, chartered banks that could and did. These bank notes continued to circulate as one of the major sorts of money until they were taxed out of existence by the Federal government in 1866. State banks still have the right to issue bank notes, but they are taxed at such a high rate that their issuance is unprofitable.

When the national banking system was set up during the Civil War, it provided for the issuance of bank notes by the national banks. Any bank issuing such notes was required to deposit with the Treasurer of the United

States government bonds to the full amount of the notes. In addition, it was necessary for the bank to deposit a "redemption fund" of 5 per cent from which the Treasurer could pay off the notes on demand.[1] Thus they had as their backing 100 per cent in bonds and 5 per cent in cash. The notes did not enjoy any legal-tender power, but in practice they were accepted on a par with those kinds of money which did have such power. The notes were in the form of promises to pay by the bank: "The Farmers' National Bank of Gurleyville will pay to the bearer on demand, etc." An attempt was made to retire these notes when the Federal Reserve System was started in 1913, but it was unsuccessful. In 1935 the government redeemed all those bonds which could legally be used as reserve for the notes, and the notes rapidly passed out of circulation until now they are a negligible part of our monetary system.

The first attempt to retire the national-bank notes called for replacing them by similar notes to be issued by the newly constituted Federal Reserve Banks. These new banks were authorized to buy the bonds which the national banks had formerly been required to own and which served as backing for the national-bank notes, and to issue new Federal Reserve bank notes against the same bonds. The new notes were to replace the old ones and to have the same security. The plan did not work, however, and very few Federal Reserve bank notes were issued. On two later occasions, there were fairly large temporary issues of these notes, but when the bonds which could be used for backing were redeemed in 1935, as mentioned in the preceding paragraph, both the national-bank notes and the Federal Reserve bank notes were retired and have almost entirely disappeared from circulation.

18·15. Federal Reserve Notes. When the Federal Reserve System was set up in 1913, provision was also made for another type of bank notes which have come to be the most important part of our paper money. These were the Federal Reserve notes. On May 31, 1946, over 53 per cent of the entire volume of paper money outstanding was in Federal Reserve notes; they were more important than all the rest of the paper money put together. For that reason they merit somewhat more careful study on our part. To some extent we shall have to postpone discussion of the problem until we have taken up the description of the banking system, but we can at least make some start on the matter now when we are describing the monetary system in general.

All the kinds of money that the United States had issued prior to the use of the Federal Reserve notes were what we call "inelastic currency"; that is, the quantity outstanding was relatively fixed and unchangeable and did not adapt itself readily to the varying needs of business. An "elastic currency" is one that varies in volume according to current needs in such a way that, when business is brisk and many transactions are taking place, the volume of money expands and, when business is dull and few transactions

[1] National-bank notes are also redeemable over the counter of the issuing bank.

are occurring, it contracts. The quantity of gold or silver does not alter with the volume of business. Similarly, the amount of outstanding government bonds stays approximately constant or, at least, the variations are not correlated highly with the monetary needs of business. Our money was either gold or silver, or was backed by gold or silver or government bonds. Since the amounts of these were not elastic, the money itself was inelastic. The Federal Reserve notes were designed to provide an elastic currency, and the proposal was to get elasticity by adopting a new and surprising type of backing for the money.

The laws relating to Federal Reserve notes have been changed several times in the years since their original issue, and it is not important for us at this point to trace the changes historically. Let us say merely that, since June, 1945, each Federal Reserve Bank is required to maintain a reserve of at least 25 per cent in gold certificates behind its Federal Reserve notes. This is for the time being the main use of gold certificates. The remainder of the backing must be made up of "commercial paper" or government securities. The notes must have 100 per cent backing, which may be entirely of gold certificates; but whatever part is not covered by gold certificates must be made up of commercial paper or government securities up to 75 per cent. Thus the backing might be 75 per cent in commercial paper or securities and 25 per cent in gold certificates, or the gold backing might be any larger proportion up to 100 per cent.

We shall discuss commercial paper in much more detail later[1] but suffice it to say for the moment that it consists of drafts, promissory notes, and other such documents arising out of ordinary business transactions. If a businessman borrows money for a month or two to carry on his business and gives his promissory note to the lender, this note is commercial paper. It is, of course, the businessman's promise to pay at some later date. Federal Reserve notes are the government's promise to pay, and strangely enough they are backed in large part by the promises of businessmen to pay. In the following chapter we shall learn how this commercial paper comes into the hands of the banks, but for the moment let us point out merely that when business is brisk and businessmen are doing a lot of buying and selling, there is much commercial paper, and the amount of possible backing for money is large. On the other hand, when business is slack, the volume of commercial paper quickly diminishes[2] and it is necessary to diminish the volume of Federal Reserve notes as the available backing disappears. The Federal Reserve notes are an elastic currency because of the peculiarity of their backing, and they are such an important part of our present monetary system that they give elasticity to the entire system.

[1] See Secs. 19 • 3 to 19 • 5.

[2] This is especially true because, as we shall see in the following chapter, commercial paper represents only short-term credit transactions, which have at most 2 or 3 months to run.

Federal Reserve notes are issued to the Federal Reserve Banks at their request and upon their deposit of the required gold certificates and commercial paper. The Board of Governors of the Federal Reserve System is empowered to require additional security at any time if it seems desirable to protect those notes which are issued. Although the Federal Reserve notes are issued by the individual Federal Reserve Bank, they are legally obligations of the United States Government and a lien on all the assets of the issuing bank. Provision is made in the statutes for temporary suspension by the Board of Governors of the Federal Reserve System of the requirement that the bank have 25 per cent in gold certificates behind its notes, but the

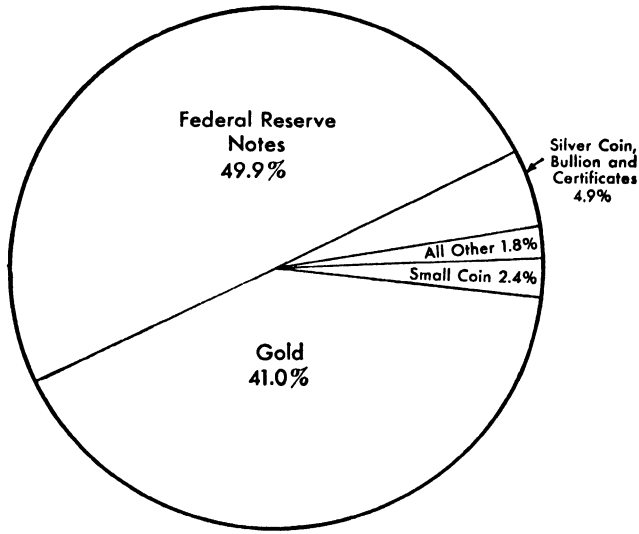


FIG. 18 • 3.—Kinds of money in the United States monetary stock, May 31, 1946.

issuance of notes under such temporary permit without the usual minimum backing in gold certificates is subject to a graduated tax.

Out of a total United States monetary supply of 49.5 billion dollars on May 31, 1946, 24.7 billion was in Federal Reserve notes and 20.2 billion in gold. Of the money outside the Treasury and actually in circulaton, 23.9 billion dollars was in Federal Reserve notes out of a total of 28.1 billion. The present outstanding importance of the Federal Reserve note in our monetary system is readily seen from these data or from Figs. 18 • 3 and 18 • 4.

Bank notes have not usually been given legal-tender power in the United States, evidently on the theory that creditors should not be forced to accept the promises of such private organizations as banks, even when they are great government-sponsored central banks such as the Federal Reserve Banks.

The national-bank notes, the Federal Reserve bank notes, and the Federal Reserve notes alike had no legal-tender power whatever until 1933, and we had the peculiar situation in the United States that the most important part of our actual circulating medium[1] had no legal-tender power. People accepted it, however, because they had confidence in it, because they felt secure in the belief that they would be able to exchange it for valuable things when they went into the market to buy.

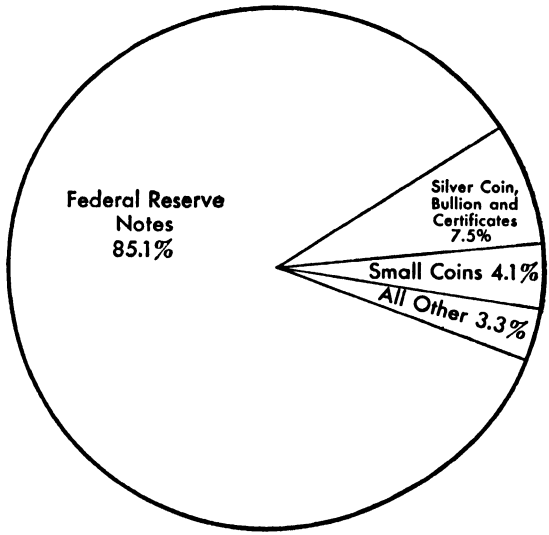


FIG. 18 • 4.—Relative amounts of different kinds of money actually in circulation in the United States on May 31, 1946.

18 • 16. Private Money. We become so accustomed to the use of money issued by the government or by governmentally sponsored agencies that we are apt to assume that there is some necessary connection—that money *must* come from some governmental agency. On the contrary, people may get their money from private sources and have often done so. The famous pine-tree shillings of the early Massachusetts Colony were privately made for private profit. From about 1839 to 1853, a considerable part of the circulating money in the Northwest and the Mississippi Valley was George Smith's money, issued, perhaps in contravention of law, by a Scottish immigrant, and accepted without question until Mr. Smith voluntarily ceased to issue it.[2] During the great depression of the 1930's in the United States, many

[1] Even when we were on the gold standard, very little gold actually circulated, and at that time the Federal Reserve notes were the most important part of the *circulating* money.

[2] For a short description of George Smith and his monetary adventures, see Horace White, "Money and Banking," pp. 465*ff*, revised by C. S. Tippetts and L. A. Frohman, Ginn and Company, Boston, 1935.

banks suspended payment, and people were often unable to get their ordinary supplies of money. Under these circumstances they did in many cases supply their own, sometimes in the form of scrip issued by cities to pay teachers, firemen, policemen, etc., and sometimes by other means. A small group of men in Tenino, Wash., for example, issued wooden money, printed on thin sheets of wood such as are used for making berry baskets, and this money circulated for some time in that vicinity. Leather money and oyster-shell money were issued in other places by men who combined a flair for publicity with their need for money. These things were actually accepted in their communities and passed freely from hand to hand. They were not used "instead of" money—they were money. Several of the individual states in this country have issued tokens with value less than a cent—five mill pieces, one mill pieces, etc.—which are used in payment of state sales taxes. In some cases these have been made of metal, sometimes of plastic. In at least one state they were for a time made from milk-bottle tops, and for the time being people actually could and did buy things in their stores and pay with large round cardboard milk-bottle tops! These, of course, were not really private money, since they were issued by the states, but they do serve to show that, when the ordinary sources of supply are not available or when they do not furnish money in all the denominations needed for trade, people are quick to furnish their own supplies and to accept them. The characteristic that identifies anything as money is its acceptability and use, not the provisions of the statutes. The provisions of the statutes may help to make something acceptable and to bring it into use, but the people themselves determine by their own behavior what is and what is not money.

18·17. Silver in the United States Monetary System. As we noted in Sec. 18·7, silver was standard money in the United States under the bimetallic system from 1792 to 1873. Thus, for over half our history silver was on a basis of equality with gold. The bimetallic system did not operate satisfactorily, since the mint ratio and the market ratio did not stay together and either gold or silver was driven out of circulation. It is possible that this would not occur if all countries adopted a uniform bimetallic system with the same mint ratio, since under such circumstances the compensatory action described in Sec. 18·7 might be strong enough to control the market ratio. Be that as it may, the system did not work well as it did exist, and silver lost its free coinage right in 1873. Since that time, however, right down to the present, there have been strong advocates of the use of silver money—people who want to "do something for silver." This pressure has been largely selfish, coming from silver-producing interests who want to sell their product at artificially high prices, and the problem has become far more a political than an economic one. Unfortunately, the political influence of the silver interests has been very considerable, and it has been coupled with

pressure from other groups of people with inflationary interests—people who wish to force prices to rise. Let us trace very briefly the history of silver in our monetary system subsequent to its demonitization in 1873.

Almost immediately after its demonitization in 1873, silver started to depreciate as compared with gold. This was reflected by a continued change in the commercial ratio of the two metals, as shown in Table 18 • 2. Silver depreciated almost at once to the point where it would have been profitable to coin it if free coinage had still been required, and the silver producers felt that they had been unjustly discriminated against, since, now that it would have paid to coin silver at a mint ratio of 16:1 for the first time in 60 years, the mint was not required to buy. Pressure was brought at once in Congress for a return to free coinage and, when that move failed by a narrow margin,[1] other legislation was introduced, which eventually passed as the Bland-Allison Act in 1878. Under this act the Treasury was required to buy between 2 and 4 million dollars' worth of silver each month and to coin it into silver dollars, which were again given full legal-tender power. This act also authorized the issuance of silver certificates redeemable in such silver dollars, and the silver certificates that now make up most of our small bills first appeared at this time. This act, was, it will be noted, a compromise. It provided for limited rather than unlimited coinage of silver, but it did reestablish a silver coin which had a face value only slightly in excess of the value of the contents and which had unlimited legal-tender power. Silver had become almost but not quite standard money. Some people referred to it as a "limping standard"—one in which silver, as they said, was not quite able to get along unaided, but limped along borne up by its stronger neighbor, gold. Fortunately the revenues of the government were sufficient to enable it to buy this large quantity of silver each month, and the silver hoard in the Treasury slowly grew larger. But pressure for further aid to silver continued, and in 1890 the chance came. The Republicans were anxious to pass the McKinley Tariff bill, and the silver interests followed up their advantage. In a political trade they threw their support to the tariff bill, and in return they received sufficient support to pass the Sherman Silver Purchase Act of 1890.[2] This act required the Treasury to buy 4.5 million ounces of silver per month at the market price, and to pay for it with legal-tender notes—the Treasury notes of 1890 to which we

[1] The bill came up for passage under suspension of the rules, a parliamentary situation which required a two-thirds favorable vote. The bill actually received more than a majority but not the necessary two-thirds, the vote being 119 to 68. This gives a good indication of the feeling of Congress on the matter at that time.

[2] There are two important pieces of legislation each of which is often called "The Sherman Act" or "The Sherman Act of 1890." To distinguish them we shall call them the Sherman Silver Purchase Act, to which we are now referring, and the Sherman Antitrust Act, which we shall describe in Chap. 25. The student should realize that they are entirely separate pieces of legislation.

referred in Sec. 18 • 12. Since silver at that time was selling at approximately $1 an ounce, it is easy to see that the drain on the Treasury was enormous, and over 150 million dollars of these notes were issued. The McKinley Tariff bill reduced government revenues, and the Sherman Silver Purchase Act increased the expenditures as the Treasury went into the market to buy some 50 million dollars' worth of silver a year. Additional expenditures for Civil War pensions likewise drew heavily on the Treasury, and this combination of circumstances brought headaches to a harassed Treasury. The Treasury notes of 1890, issued for the purchase of silver under the Sherman Act, were redeemable in either gold or silver. Under the circumstances people did in fact demand their redemption in gold. With tremendous numbers of the Treasury notes being issued each month in accordance with the provisions of the law, and with people redeeming them at once in gold, the gold reserves of the Treasury were depleted very rapidly until it began to appear that they would be completely exhausted and that the government would have to suspend payment. The situation became so alarming that a severe financial panic was precipitated in 1893. Finally, President Cleveland was forced to call a special session of Congress for the purpose of repealing the Sherman Silver Purchase Act. This was done and the act repealed in 1893.

This, however, did not end the pressure from the silver groups. Economic conditions within the country were such as to bring about widespread suffering and discontent,[1] and many people mistakenly believed that the issuance of more money would automatically remove the difficulties. The presidential campaign of 1896 was fought almost exclusively on the issue of bimetallism, and only after the defeat of the "free silverites" in that campaign (and perhaps even more important, after the long downward trend of the price level had been reversed at about the same time) was the silver question settled for the time being. In 1900, Congress passed the Gold Standard Act, which gave formal and explicit recognition to the single gold standard, which had in fact existed implicitly since the demonitization of silver in 1873. Under the Bland-Allison Act and the Sherman Silver Purchase Act together the Treasury had spent over 450 million dollars for silver and had almost wrecked the monetary system in the process.

For several years the United States continued to carry this very large amount of silver bullion. The silver dollars were too heavy for convenient use and, except in some parts of the West, they never passed into circulation to any appreciable extent. With the advent of the first World War there was a greatly increased demand for silver, and the value increased until finally the government could have melted down its silver dollars, sold the

[1] This was the result of a long-continued fall in the price level, which we shall discuss in Sec. 21 • 5.

bullion, and realized a profit. The pressure of the silver interests was still too strong to allow this to take place. In 1918, Congress did pass the Pittman Act, which authorized the Treasury to call in silver certificates and melt down the silver dollars that were held for their redemption. Under this law over 250 million dollars of silver were melted down and exported. But, in order to secure the agreement for the passage of the act, the Administration had to agree that when the emergency was over the full amount would be bought again at a price of at least $1 an ounce, and that silver dollars would again be coined and silver certificates issued against them. These latter provisions were included in the original Pittman Act itself, so that there was, therefore, only a very temporary reduction in the silver hoard of the Treasury.

The most recent chapters in our silver experience date from the early 1930's. In the welter of monetary legislation that accompanied the start of the New Deal were several measures having to do with silver. As early as 1933, President Roosevelt authorized by proclamation the purchase of domestically mined silver at the rate of 64½ cents per ounce despite the fact that silver could at the time be purchased on the open market at about 40 cents per ounce. By this means the silver producers were in effect given a subsidy of over 60 per cent. Half the silver so purchased was coined at a coinage value of $1.29 per ounce, or just twice the purchase price. By such a seigniorage arrangement the government could coin half its silver into enough money to pay for all of it, leaving the other half as clear gain. But this subsidy was not sufficient to quiet the silver interests.[1] After trying unsuccessfully to get approval for several other measures, they finally scored a major victory with the passage of the Silver Purchase Act of 1934. This act declares that it is the policy of the United States to increase the silver in the country's monetary stocks until ultimately one-fourth of the monetary value of the entire stock shall be silver. The Secretary of the Treasury is directed to issue silver certificates against silver purchased for this purpose, and to put them in circulation. As a result, almost all the small bills of the country have come to be silver certificates.[2] The President was empowered to "require the delivery to the United States of all silver by whomever owned or possessed," and under this authorization the silver stocks of the country were nationalized in August, 1934. On May 31, 1946, the total gold coin

[1] It is not meant to insinuate here that the silver interests are any worse than any of many other selfish organized groups who seek to further their own advantage at public expense. Perhaps those who accede to such demands are more to be censured than are those who make them. As we shall see later, the railroads, other public utilities, manufacturing interests protected by tariffs, organized labor groups, and many others have acted in the same selfish manner as that which has been displayed for many years by the silver bloc in Congress and their backers in the industry.

[2] On June 30, 1946, there were 1,179 million dollars outstanding in one-dollar bills, of which 1,172 million dollars were silver certificates.

and bullion in our monetary stock amounted to 20.2 billion dollars, while the silver dollars and bullion amounted to 2.4 billion. Thus, at that date silver made up about 10 per cent of the monetary stock, and the purchases of silver continue. It seems fair to say that these steps were taken more for political than for economic reasons, and they left our monetary system more muddled and less stable than it had been before. The artificially high prices paid for silver by the Treasury under these acts has resulted in a large increase in the silver stocks, to the great embarrassment of countries that were on the silver standard, such as China and Mexico. And all to what end? One competent monetary specialist[1] has summarized the situation in the following words:

> Viewed from the standpoint of domestic results it must be concluded that the silver program has accomplished precisely nothing. There is no conclusive evidence that it has had any effect upon the general level of commodity prices. The issue of certificates against the silver acquired has increased somewhat the proportion of that form of currency in circulation by driving out a corresponding amount of other kinds of money, and has contributed some $350 million to the total of about $3 billion of recent excess bank reserves. The purchases of domestic silver at prices away above even the artificially high world market has given a handsome subsidy or bonus to the silver mines. The purchase of foreign silver has stimulated the drooping silver-mining industry of Mexico and Peru and has given Indian silver hoards an artificially marked-up value. It has thrown China, Hong Kong, and other silver standard countries off their old and satisfactory standards, has confused their monetary systems, dislocated their normal trade relations, and caused a marked shrinkage in our trade with those countries. The silver policy has saddled the United States Treasury with a program of costly buying of a commodity for which we have no possible use. For years before the Silver Agreement plans had been considered for getting rid of our huge "silver hoard." The operations of the silver program if carried to conclusion will only add further millions of ounces of useless silver to that hoard, and, when the loss has to be taken, further burdens of taxation on all the American people. What, then, shall we do with this enormous pile of useless silver? The answer to this question no living man knows. "We are left to carry this bag through eternity."

18·18. Present Monetary System of the United States. As has been intimated earlier in this chapter, the United States is, at least for the time being, not on a gold standard nor a silver standard nor a bimetallic standard. In fact, our monetary situation is not easily classified on any basis, and most people look on it as temporary without having any way of knowing what will replace it. Both gold and silver have been "nationalized," and gold and silver bullion are held by the Treasury for some purpose that has

[1] HOLDSWORTH, JOHN T., "Money and Banking," p. 105, D. Appleton-Century Company, Inc., New York, 1937.

never been explained. The Federal Reserve Banks hold gold certificates for reserve purposes which they cannot legally pay out, nor can they legally be redeemed for the gold which they purport to represent. For purposes of foreign trade, businessmen may obtain gold for export under special license, but at home the United States is on what we call an "irredeemable paper standard." This is a fiat money system in which the paper money enjoys full legal-tender power but is not convertible for standard money. Such a standard has usually been adopted historically only under compulsion, in time of financial distress caused by war or by business depression; but it seems to have been adopted voluntarily by the United States in 1934. The changes of the past decade and a half have, however, given us a much simpler monetary system in several respects—one in which national-bank notes, Federal Reserve Bank notes, and Treasury notes of 1890 have all but disappeared, and both gold coin and gold certificates have passed out of circulation. There is no good reason for continuing the use of United States notes, silver dollars, or silver certificates; yet the inflationary influence in Congress is strong enough to make it unlikely that they will soon be withdrawn. Certainly the purchase of more silver should be stopped at once. Even a return of the gold standard would probably not bring a return to the coinage of gold, since it is quite as convenient to circulate gold certificates redeemable in gold bullion. The monetary system has also been simplified by the changes that gave full legal-tender power to all kinds of money—a change that seems likely to be retained with the exception of small coins, where some limitation on legal-tender power is desirable.

At present, almost every country in the world is on some form of inconvertible paper standard. Until recently, these standards have been thought of as temporary devices, to be abandoned as soon as possible, and the countries that use them still express their monetary units in terms of the former standard coins. Yet there is a definite tendency on the part of some people to urge an indefinite continuation of the use of inconvertible paper. This tendency has been enhanced by the fact that the United States, because of a combination of circumstances, finds itself in possession of a very large proportion of the entire gold stock of the world, and other nations may be unwilling to base their currencies on gold unless and until there is a more equal distribution of it. But, entirely aside from the question of the distribution of gold, there are many advocates of inconvertible paper. This can best be understood if we examine briefly the advantages and disadvantages of a gold standard as compared with the use of inconvertible paper.

18 · 19. Advantages of the Gold Standard. Without going into great detail, we may say that the gold standard had the advantage that it was not open to human manipulation and caprice. The value of money in a country on the gold standard is fixed by the supply of and the demand for gold—

forces that cannot be arbitrarily determined or altered. Exchange rates between countries on the gold standard were stable and fixed within narrow limits,[1] and the use of this standard sets a definite upper limit to the possible expansion of the currency. Countries on this standard can issue other sorts of money only up to the limit that they can cover with gold when redemption was demanded. Thus, there is much less likelihood of uncontrolled inflation with its serious disadvantages. When countries are on the gold standard, we also have the advantage of an international standard by means of which exchange can easily be carried on between countries that have monetary units of diverse weights and names, but commensurable in the fact that they are all expressed in gold units.

18 • 20. Disadvantages of the Gold Standard. To offset the advantages mentioned in the preceding section, we must confess that there were ways in which the gold standard worked very poorly indeed. While international exchange rates were kept relatively stable, price levels within countries fluctuated rather violently,[2] and it is felt by many people that the gold standard sacrificed the more important domestic trade advantage in order to secure advantages in foreign commerce. Gold, like other commodities, fluctuates in value and, if we are to base our money on gold, we are likely to have money that fluctuates in value. Many people think that it is better to have a *managed currency* than one that is bound to fluctuate with every change in the gold-mining industry. They would prefer a currency the value of which is subject to manipulation, thinking that in this way the value of the currency can be manipulated to human advantage rather than being left to itself. This is probably the nub of the difference of opinion over any proposed return to the gold standard. Those who trust human agencies and administrators to pursue wise policies reflecting general social advantage generally advocate a managed currency and oppose a return to gold. They realize that human beings err, but they feel that the danger of loss from that source is less than the danger of throwing up our hands in despair and letting nature take its course. Those, on the other hand, who favor a return to the gold standard, seem to feel for the most part that the use of a managed currency concentrates a dangerously large power in the hands of politicians, which can (and, they think, probably would) be used for personal, party, or class ends. They would point to the long historical record of government dishonesty in monetary matters which runs back almost as far as recorded history, and would urge that it is better to tie to something that human beings cannot change, even if it is imperfect, than to trust such important matters to fallible human beings. They would point out that our other standards, such as those of weights and measures, are not ordinarily subject to change

[1] This matter is discussed in detail in Chap. 24

[2] This matter is discussed in Sec. 21 • 5 and presented graphically in Fig. 21 • 1.

at the whim of the government, and they would argue that the standard of values should be similarly treated.

It is quite impossible to make any categorical statement on such matters which will meet approval from all concerned. When this book goes to press, there is no indication of the course that will ultimately be pursued, either by the United States or by other countries. Certain it is that the events of the past decade or two have done much to destroy the blind worship of the gold standard that led many of its advocates to assume that any other kind of monetary standard was almost immoral and surely unworkable. The problem today has come to be one largely one of expediency, with opinions varying according to one's judgment of human nature and his experience with the operation of government as a system of practical politics. Perhaps it is well for us to remember that the gold standard is itself a very recent invention, and it may be reasonable to suppose that the systems in use in another generation will be better than any so far proposed.

18·21. Functions of Money. Before we can take up the other problems of money that need treatment in such an elementary book, it will be necessary for us to turn in the following chapter to the problems of banking. But, before we leave temporarily our consideration of money and monetary problems, let us summarize some of the things that we have been saying and other things that have been implicit in our treatment, by noting that the money of any society performs a number of different functions. Some of them it performs well, and some of them poorly. First of all, we have already said that money serves as a *medium of exchange*. We have noted the inconvenience of barter and the advantage of using money to effectuate exchanges. Our money has served this purpose very well. It has been generally acceptable, and a very large proportion of actual exchanges have taken place through money as a medium. We must remember, of course, that people are free to use other things as mediums of exchange if they wish. Use as a medium of exchange is a necessary but not a sufficient condition for money. The element of general acceptability must also be added. In addition, money serves as a *measure or standard of value*. It is quite possible for something to be a medium of exchange without being at the same time a standard of value. Thus, I might swap my potatoes for corn meal with the intention of swapping the corn meal in turn for clothing. Corn meal would have served as a medium of exchange, but not as a standard of value. When different men think of the values of commodities, they naturally tend to think of them in different terms. The farmer values a pair of shoes as being the equivalent of so many dozens of eggs or so many baskets of apples; the tailor values a pair of shoes as being worth so many suits of clothing; the butcher values a pair of shoes as being the equivalent of so many pork chops; etc. Each man thinks of values in terms of what he personally has to

give up. But, in a monetary society, these men all get a common language—they can all talk in terms of a common unit of account, such as the dollar. This unit may not actually be coined and used as money. The gold dollar was not actually coined during much of the period when we were on a gold standard. Yet values were reckoned in gold dollars as units, and it was this gold dollar that served as a measure or standard of value. Today we speak loosely of the dollar as containing 15-5/21 grains of gold, yet no such dollar has ever actually been coined or used as a medium of exchange. The medium of exchange and the standard of value may be different; but a good monetary system supplies both. The monetary system of the United States has worked rather well in furnishing a measure of value which was commonly used and which furnished a workable method of comparing the diverse value ideas of different people. The unit of account changes much less often than the money in which it is embodied. Thus, since 1792, people in the United States have kept their accounts in units of dollars, but we have already seen that the dollars referred to have changed over and over again during the period.

The other two common functions of money both involve lapse of time and, because of that fact, they have been much less satisfactorily performed by our monetary system than have those already discussed. Money is expected to serve as a *standard of deferred payments.* If a man enters into a contract today for the purchase or sale of goods a year from today, it is necessary to specify somehow what value he is to give up after the lapse of the year. A man buying a house or an automobile on the installment plan promises to pay over a longer or shorter future period, and it is necessary to evaluate these deferred payments. Thus, we may say that the buyer will pay $40 a month for 3 years. In this case, we are measuring future values with the dollar rather than present ones. Now, although the dollar serves reasonably well in comparing the values of present things, it serves rather poorly as a measure of future values, because the dollar itself varies in value from time to time.[1] This is evidenced by the fact that a dollar will sometimes buy more and sometimes less—when prices are high, it buys little; when prices are low, the same dollar buys much. The dollar does not embody some definite and permanent quantity of value, but itself changes in value. It would be very inconvenient to have a measure of length that varied—a yardstick that sometimes stretched around the block and sometimes shrank to vestpocket length. Under such conditions a man might today be 10 yards tall (when the yardstick was small) and tomorrow be but a fraction of an inch tall (when the yardstick had been stretched out). It would be difficult to compare our ideas of length from day to day, and the man who contracted to buy 1000 yards of wire fencing a year from today would have to

[1] This matter we shall take up for separate discussion in Sec. 21 • 4.

run the risk that the yardstick might then be very short, and his 1000 yards might not reach around the chicken yard. On the other hand, at the time when the wire was delivered, the yardstick might be very long, and he might get far more wire than he had anticipated, and far more than the seller could afford to give him at the stipulated price. Changes in the value of money, as reflected in changes in the general level of prices, make money a poor standard of deferred payments. To be sure, gold and silver vary less in value than most commodities do, for reasons mentioned in Sec. 18 • 2, but the variations are nonetheless large and their effect is serious, as we shall see in Sec. 21 • 6.

Closely allied to this function of money is another; namely, that money serves as a *store of value*. The man who has raised a crop of sweet corn and finds it matured and ready for market may not yet be ready to use all its value—to convert all its purchasing power into other goods. He may want to use some of its value now to buy groceries and chewing tobacco, but he may wish to save up some of the value of the sweet corn to use later to put his youngsters through college or to live on in his old age. The sweet corn cannot in itself be kept until these times, but its value can, in a sense, be stored up and kept for future use. The farmer can sell his corn today and keep the money for 10 or 50 years, spending it at the later time for things that he then wants. He will in a very real sense be utilizing the value of the sweet corn long after the corn itself has passed out of existence. He will have used the money as a method of storing up the value. If money kept a constant value, he would be able to store the value without gain or loss. As it is, he keeps, let us say, $100 out of the money obtained for the corn and saves it for 20 years. He then spends it for commodities of various sorts. If prices have gone up during the two intervening decades, he cannot buy so much with the $100 as he had expected. He put a given amount of purchasing power in storage, and some of it disappeared because of the fluctuations in the value of the money. If, on the other hand, prices have fallen during the 20-year period, he finds that he can buy more than he anticipated. He has made a fortuitous gain. Although our monetary system seems to have been very good in furnishing us with a medium of exchange and with a standard of values, the fluctuations in the values of money have made it a rather poor standard of deferred payments and storehouse of value. We shall return to these problems again in Chap. 21.

Questions

1. Some people have a superstitious dislike for $2 bills. If this antipathy became general and people refused to take $2 bills, but if such bills were still issued by the government in their present form, would you say that they were or were not money?

2. Do you think that it would be possible to adopt for money something that was too scarce? What, for example, would be the disadvantage of using radium as money?

3. Why was it, when we were on the gold standard, that small coins such as 10-cent pieces and 1-cent pieces were not made of gold?

4. The stone money of Yap is criticized in Sect. 18•2 for its lack of portability. Yet when the United States was on a gold standard people did not in practice really carry gold with them at all, but pieces of paper money representing gold. Do you think it would be possible to use representative money in Yap, redeemable in millstones?

5. Explain the relationship between durability and relative constancy in value.

6. Would the tokens commonly used by street-car companies and bus companies for fares be considered money? Why or why not?

7. During the second World War housewives who bought meat or sugar or shoes were required to give ration points for them. In some cases these were small disks of colored plastic. Would you consider that these were money? Why or why not?

8. In a country on the gold standard, would a gold coin 0.900 fine be worth more or less than another gold coin of the same weight and 0.950 fine? Why?

9. When silver is selling on the market at 40 cents per ounce, what is the value of silver in a silver half dollar?

10. What price of silver would make the silver in a half dollar worth a half dollar?

11. If a copper cent weighs 48 grains of which 95 per cent is copper, what must be the price of copper in order that the cent contain a cent's worth of copper?

12. How does standard money differ from other money?

13. Standard money always has value of contents exactly equal to the face value, while other money does not. Could this other money have value of contents in excess of the face value, or must it always be smaller than the face value? Why?

14. Distinguish brassage and seigniorage.

15. Define legal tender.

16. Explain why it is that bimetallism did not work in practice.

17. State, explain, and illustrate Gresham's law.

18. A given country has bimetallism. Its two standard coins are a gold dollar weighing 20 grains and a silver dollar weighing 60 grains. On the market, 4 pounds of silver actually are worth the same as 1 pound of gold.

    a. What is the mint ratio?

    b. What is the market ratio?

c. Which metal will be used in payment of debts?

d. Exactly how will debtors gain by using this metal in preference to the others when they pay their debts?

19. What is compensatory action?

20. Using the figures of Table 18 • 2 and recalling that the present United States dollar contains 15 5/21 grains of gold, what weight would the silver dollar have had in 1941 to make the mint ratio and the market ratio equal?

21. Since the silver half dollar today contains less than 30 cents' worth of silver, why do debtors not pay all their debts in silver half dollars, driving other kinds of money out of circulation? Would one expect this from Gresham's Law?

22. Why is fiat money made legal tender?

23. Why was legal-tender power given to all United States money soon after we left the gold standard?

24. What were the principal differences between United States notes and Treasury notes of 1890? In what ways were they alike?

25. What is an elastic currency?

26. How was elasticity obtained in the case of Federal Reserve notes?

27. When we establish an elastic currency, is it our idea merely to get a currency that will fluctuate greatly in amount, or is there something more involved? Comment.

28. The statement that Federal Reserve notes are backed 25 per cent in gold and 75 per cent in commercial paper is not correct without qualification. What qualifications are needed to make the statement accurate?

29. What objections, if any, would there be to having all of our money be private money?

30. What were the major differences between the Bland-Allison Act and the Sherman Silver Purchase Act?

31. Explain how the Sherman Silver Purchase Act almost forced the United States government to suspend the gold standard.

32. Can anything be a standard of value without being a medium of exchange? A medium of exchange without being a standard of value?

19. THE COMMERCIAL BANK

19 • 1. Kinds of Credit. Modern business is operated to a surprising extent on credit rather than on a cash basis. Manufacturers commonly buy their raw materials on credit, manufacture the finished products, and sell them to raise funds to pay off their debt. Merchants commonly buy the goods that appear on their shelves on credit, sell them, and collect the funds to repay their debt. No one knows just what proportion of business transactions are credit transactions, but it would certainly run to a very great majority and estimates usually run to over 90 per cent.[1]

We find it convenient to divide credit first into two major classes: consumer credit and commercial credit. As the name implies, *consumer credit* is that which is extended to assist consumers in obtaining and enjoying consumption goods. When a family buys a new automobile, a new radio, a new refrigerator, or a new house, it is common for them to obtain some sort of consumer credit. *Producer credit* or *commercial credit,* on the other hand, is extended to assist in business operations—in the production of goods rather than in their consumption.[2] Businessmen and bankers often describe these two kinds of credit in other terms. They speak of "unproductive" credit, and of "productive" or "self-liquidating" credit. If we lend to a manufacturer to enable him to buy raw material, we have made it possible for him to set productive processes in motion which will, in the ordinary course of events, bring forth finished products that can be sold to repay the loan. Credit is

[1] It might be better to call them "guesses" rather than "estimates," since there is no way of getting more than very rough accuracy. This is especially true because, as we shall see, people often use credit without knowing it.

[2] It will be recalled that we have defined production in such a way as to include all creation of utility (see Sec. 10 • 1). Thus the railroad company that transports goods to market, the storage company that holds them until wanted, or the retail merchant who brings them to the consumer's doorstep and holds them until they are desired, are all included as producers, and all use commercial credit. For those who are not fully accustomed to the technical use of terms in economics, we might as well have said above that commercial credit is extended to assist in the *production and distribution* of goods.

self-liquidating or *productive* if it is used for the purpose of producing things that can be sold to repay the credit. Similarly, when a merchant borrows money to buy a stock of goods, the credit has made possible productive processes that enable the borrower to repay the loan. The credit has enabled the borrower to buy goods that will be resold, and from the proceeds the loan may be repaid. If, however, a consumer borrows money to buy a new suit of clothes or living room furniture, no processes are set in motion which in themselves make it possible to repay the loan. The credit has been *unproductive*. The student should be sure to understand that the use of the terms "productive" and "unproductive" is not intended to convey approval or disapproval. We do not think of them as distinguishing good from bad or desirable from undesirable kinds of credit.

Having distinguished credit according to its use, we also often divide it into two or more categories according to the length of time for which it is extended. Sometimes the borrower obligates himself to pay whenever the lender demands. Such credit is known as "call loans" or "demand loans." More commonly the credit extends for a definite period. The borrower is under no obligation to repay during this period, but at the end of it he is required to repay the full sum borrowed. Of course, if the intervening interval is long enough, the borrower may make payments of interest meanwhile, and in some cases arrangements are made to pay off the principal in small regular installments rather than waiting to pay it all in one lump sum at the end. When a debt is repaid in such installments, we say that it is "amortized." Amortization is the gradual extinguishment of a debt, "killing it off" by degrees, as the word implies. Naturally, amortization is used only if the credit period is a rather long one. Businessmen have fallen into the habit of classifying credit periods rather arbitrarily into short-term and long-term credit. "Short-term credit" is that which runs for periods of 3 months or less—usually for 30, 60, or 90 days. Credit that runs for more than 3 months is usually called "long-term credit," although sometimes, especially in the case of agricultural credit, loans of between 3 months and 3 years in length are called "intermediate credit."

19·2. Commercial Banks. Credit is furnished by many different kinds of institutions: pawnbrokers, small-loan companies, building and loan associations, insurance companies, investment houses, etc. Some of these agencies specialize in consumer credit, others in producer credit. Some are primarily in the long-term and some primarily in the short-term field. The type of agency to which one would go for credit would depend largely on the purpose for which it was to be extended and on the length of time for which it was wanted. The commercial bank, to which we shall direct our major attention in this chapter, is mainly an institution that extends short-term commercial credit. Such banks often perform other services in addition, such

as renting safe deposit boxes to people who want safekeeping for their valuables, or acting as trustee, administrator, receiver, and the like. In practice they are often called "trust companies,"[1] "national banks," etc., and often it is impossible to be certain from the name of the bank just what sort of business it does do. We must, however, distinguish commercial banking from savings banking. The savings bank, as its name implies, is primarily an institution established to safeguard the small savings of individuals. These savings are ordinarily invested by the savings bank in long-term investments, such as mortgages or bonds. The commercial bank, on the other hand, exists mainly for borrowers rather than for savers, and its business is one of short-term rather than one of long-term credit. To be sure, many modern commercial banks have savings departments, and conduct a savings banking business along with their business of commercial banking; but this is somewhat the same thing that happens when an undertaker conducts a furniture business along with his other work. A commercial bank with a savings department has added another very different sort of business which does, however, have some connection with its main business and which is expected to be a profitable side line.

Commercial banks are not unusual things, beyond the experience of the ordinary student. The *Federal Reserve Bulletin*[2] lists 14,011 commercial banks in the United States on Dec. 31, 1945, and a considerable majority of all the institutions that call themselves banks are doing a commercial banking business. These banks are for the most part private corporations, owned by their stockholders and, like any other corporation, run in the hope of making a profit. Some few of them are partnerships or individual enterprises. They are usually subject to all the laws regulating other corporations and, in addition, to specialized banking laws, which govern the more technical and specialized parts of their businesses. We saw in Chap. 9 that most business corporations are brought into existence by the issuance of a charter by the state. In the case of commercial banks about two-thirds are chartered by individual states and about one-third by the Federal government.[3] The former are called "state banks," the latter "national banks." State banks are subject to state law and regulated by state authorities (except as they may voluntarily submit to Federal regulation in order to obtain special benefits, about which we shall speak later), while national banks are subject to Federal law and are regulated by Federal authorities. The banking laws of the individual states differ; consequently, the state banks can be described only in a

[1] Whatever distinction there may have been originally between trust companies and banks seems largely to have disappeared. Most commercial banks do perform trust functions, and most trust companies do perform commercial banking functions.

[2] Vol. 32, no. 7, July, 1946, p. 772.

[3] The figures for Dec. 31, 1945, were 5017 national banks and 8994 state banks to make up the total of 14,011 commercial banks. See *Federal Reserve Bulletin, op cit.*

general way. The national banks are uniform throughout the country, and we shall be able to describe them in somewhat more detail.

19·3. Credit Instruments. Some loans are made on the basis of oral agreements without documentary evidence, as is usually true, for example, when one borrows a quarter from his roommate. Most business loans are, however, evidenced by some sort of written document, which shows how much is owed and when and how it is to be repaid. These documents or credit instruments may be classified into two broad groups: promises to pay and orders to pay. The former are made out by the borrower or debtor, promising to repay the sum borrowed. The latter are made out by the lender or creditor, ordering the borrower to repay the sum of money involved. In either case, it will be necessary to include the appropriate facts with regard to the sum involved, the interest to be paid, etc. The commonest form of promise to pay is the *promissory note,* sometimes called an "I.O.U." The commonest form of order to pay is the *draft* or *bill of exchange.*

In either of these cases, the person to whom the funds are to be paid may be restricted or, on the other hand, the instrument may be so drawn that the borrower has no way of knowing, when he borrows the money, to whom he will ultimately have to pay it. A promissory note might be made out in the form, "I promise to pay to Harrison Carter," or the corresponding draft might be made out in the form, "Pay to Harrison Carter." In either of these cases, Mr. Carter is the only person who may collect when the instrument falls due. In business practice, credit instruments are always made payable "to the order" of someone. Thus a promissory note would be in the form, "I promise to pay to the order of Paul Taylor," or a draft would be in the form, "Pay to the order of Paul Taylor." If the instrument is in this form, Mr. Taylor may collect the funds himself when they fall due, or, if he prefers, he may order that they be paid to someone else. This he would do by an *endorsement* on the instrument. A draft or promissory note is endorsed by writing one's signature on the back. If the signature of the payee alone is written, we say that the instrument has been *endorsed in blank,* or that it bears a *general endorsement.* Often, however, the payee specifies in his endorsement to whom or to whose order the instrument is to be paid. Thus if Charles Sedgewick borrows from Wendell Kinsey and gives Kinsey his promissory note payable to the order of Wendell Kinsey, Kinsey can sign his name on the back and turn the note over to someone else. In such a case, when the note bears the general endorsement, it will be payable to the bearer. But Kinsey might endorse it by writing, "Pay to the order of Sumner Dole," in which case only Dole or someone designated by him could collect. This is called "endorsement in full," as distinguished from endorsement in blank or general endorsement. When a credit instrument is written, dated, and contains an unconditional promise or order to pay to another person or to his order or to the bearer a definite sum of money either on demand or at a

specified future date, the lawyers call it a "negotiable instrument." The ordinary bank check is a negotiable instrument, since it obviously meets all these requirements. In fact, it is a form of draft, since it is an order to pay rather than a promise to pay. In the typical bank check the depositor is ordering his bank to pay a specified sum of money to some named payee or to his order.

19·4. The Promissory Note in Use. Let us imagine that the Keystone Sack Company wants to buy $1000 worth of heavy twine from the Foul Line Company for use in the manufacture of its sacks, but the sack company is not in a position to pay cash. If it can get the twine now, it can manufacture and sell the sacks within, say, 2 months. Under such circumstances, the promissory note might be used in either one of two ways. In the first place, the Keystone Sack Company might borrow the money directly from its bank, which, we will suppose, is the Merchant's National Bank of Gurleyville. In such a case the sack company, through its treasurer or other designated agent, would make out a promissory note and take it to the bank, where it would collect the desired funds. The note would be in some such form as the following:

$1000.00 Gurleyville, Conn., Sept. 28, 1947.

Sixty days after date we promise to pay to the order of the Merchant's National Bank of Gurleyville One thousand dollars. Value received.

(Signed)
Keystone Sack Company
by J. H. Arjona, Treas.

Having borrowed the funds from the bank, the sack company is now able to pay cash to the Foul Line Company for the twine, and the seller will not have any way of knowing that the transaction was a credit transaction. On the other hand, the seller might himself extend the credit. In this case, the buyer would make out the note as before, but this time would make it out payable to the order of the seller, and the document would have something like the following form:[1]

$1000.00 Gurleyville, Conn., Sept. 28, 1947

Sixty days after date we promise to pay to the order of the Foul Line Company of Amherst, Mass., One thousand dollars. Value received.

(Signed)
Keystone Sack Company
by J. H. Arjona, Treas.

[1] In practice, both promissory notes and drafts are printed forms, like bank checks, on which the person drawing up the documents merely fills in the appropriate blanks to show the date, the amount of money involved, etc.

In this latter case, however, the document is still likely to fall into the hands of a bank and to be collected by them. Most business firms are not in the credit business. The seller in this case is in the business of making twine and will want to keep its funds invested in that business rather than keeping them tied up in credit. So, in this case, the chances are that the Foul Line Company will take or send this promissory note to its own bank and turn it over to them, letting them collect it when it comes due. The bank will turn the money over to the Foul Line Company at once[1] and the bank will then collect the funds at the end of the 60-day period. It is obvious that, in this case as in the former one, the credit has been furnished by a bank, even though this time the promissory note was originally made payable to someone else. The Foul Line Company will endorse the note before turning it over to its bank, and we shall see the effect of such endorsement later.

19 · 5. The Draft in Use. Suppose that the Gurleyville Furniture Company wants to stock up with a new line of linoleum, and a salesman for the Eagleville Linoleum Company drops in. They go over the line and make out an order for $1500 worth of linoleum, but the furniture company is either unable or unwilling to pay cash. If the two firms have done business before, the salesman will know whether or not he is allowed to extend credit to this store and to what extent. If it is a new customer, the salesman may have to get in touch with his credit department. At any rate, we shall assume that it is decided to extend the credit. Under such circumstances, the order is made out and sent to the linoleum factory, and the buyer tells the seller that "you may draw on me for $1500." This means that they may draw a draft against him—draw up an order requiring him to pay the money. The buyer himself does not draw up the document in this case.

The order for the goods goes to the factory in Eagleville and, when it has been assembled, it is put on the car at the railroad station. The railroad company furnished the shipper with a document known as a "bill of lading." This is a receipt for the goods, and the original copy carries title to the goods. It shows who has the right to take them from the car at the destination. The shipper (linoleum company) now draws up a draft ordering the buyer (furniture company) to pay the sum that is due and takes this draft, along with the bill of lading, to its own bank. The draft will have a form similar to the following:

$1500.00 Eagleville, Conn., Sept. 28, 1947.

To Gurleyville Furniture Company
Gurleyville, Conn.

Sixty days after date, pay to the order of the Eagleville Bank and Trust Company One thousand five hundred dollars.

(Signed)
Eagleville Linoleum Company
C. J. Turner, Treas.

[1] As we shall soon see, the bank does not give the full face value of the note, $1000, to the payee, but discounts it somewhat. This matter, however, will be explained in Sec. 19 • 6.

At the bank, the seller of the goods would turn over both the draft and the bill of lading, and in return would receive the money at once.[1] The bank in turn would send the draft and bill of lading through the mail to some bank in the buyer's neighborhood. This bank, possibly the buyer's own bank or possibly some other, would notify the Gurleyville Furniture Company that it held a bill of lading for a shipment of linoleum and a draft for $1500. The treasurer or some other authorized official of the furniture company would then go to the bank and "accept" the draft. This he does by writing the word "Accepted" and his name on the face of the instrument. A draft that has been accepted in this way by the person on whom it is drawn is called a "trade acceptance," or merely an "acceptance." The furniture company is now legally obligated to pay the draft when it comes due. As soon as the draft is accepted, the bank turns over the bill of lading to the furniture company, which enables them to get delivery of the linoleum from the railroad. If the furniture company had refused to accept the draft, two things would have happened: (1) they would have been unable to get the linoleum and (2) unless they had some valid reason, they would lose their credit standing with their bank. Thus the shipper of the goods is protected by the device of using the draft and bill of lading. In 60 days, when the draft falls due, the bank in the buyer's neighborhood collects the amount due and remits it to the bank that originally cashed it for the seller. The operation is then complete. The buyer got the goods right away, and the seller got the money right away. The money was put up by the bank.

19·6. Discount and Interest. The bank, of course, does not furnish the credit in these cases merely out of the goodness of its heart. When business firms extend credit, they make a charge for the service. This charge takes either one or the other of two forms: interest or discount. These differ primarily in the time when the charge is made: whether at the time of the original extension of credit or at the time when the final payment falls due. *Interest* is usually added to the amount borrowed and paid at the end of the transaction. Thus, if a merchant borrows $100 from his bank for a year and if interest is charged at 6 per cent, the merchant will receive $100 now and after the lapse of the year he will pay $106 of which $100 will repay the principal and $6 will be interest on the transaction. Much more common in business circles, however, is discount. When the mathematician computes the discount on $100 for 1 year at 6 per cent, he finds the sum of money which, if put at interest for 1 year at 6 per cent, would amount to $100. This is called "pure discount" and is found by dividing $100 by 1.06. The discounted amount would be $100 ÷ 1.06 = $94.34. The commercial banker uses a simpler method, largely because it takes less time but also because he makes

[1] Again, as in the case of the promissory note, the draft would be discounted and the seller would not receive the full amount. For explanation, see the following section.

more money that way. If we are to compute *bank discount on* $100 for 1 year at 6 per cent, we merely deduct 6 per cent of $100 or $6, and get $94 as the discounted amount. In discounted transactions, the charge is deducted at the time the credit is extended. Instead of getting the full amount of the note, the borrower is given the discounted amount and, when the note falls due, he pays its face value. Thus, if a man gives a commercial bank his note for $100 due in 1 year and if the discount rate is 6 per cent, the bank will give him $94 now (the $100 less $6 discount) and, when the note falls due, he will repay the $100 face value of the note. We can see the difference between these different charges by noticing what would happen in the case of an extension of $100 of credit for a year at 6 per cent under each of the three systems. The amounts received by the borrower at the start and repaid by him after the year are shown in the accompanying table.

	Amount received	Amount repaid
Interest.... .. .	$100	$106
Pure discount. . .	94.34	100
Bank discount... . .	94	100

Evidently, pure discount is somewhat more expensive than interest and bank discount is somewhat more expensive than pure discount, if they are all charged at the same nominal rate. On an interest transaction, the borrower paid $6 to have the use of $100 for a year; on the bank discount transaction, he paid $6 to have the use of but $94 for a year. Commercial banks almost invariably extend credit on the basis of bank discount, and the student should assume in what follows that any charges are computed on that basis unless the contrary is specified. Also, we shall use the single word "discount" to represent "bank discount" unless the term "pure discount" is used in full.

Of course, commercial banks do not ordinarily lend for periods as long as a year; consequently, the illustrations just used are not typical. A large proportion of commercial bank loans are made for one of three periods: 30, 60, or 90 days.[1] Interest rates and discount rates, however, are always stated on an annual basis. A discount rate of 6 per cent means 6 per centum per annum—6 per cent for a full year. If a businessman borrows from his bank for 60 days, he has borrowed for one-sixth of a year and the discount will be but one-sixth the annual amount. Perhaps we should illustrate this specifically. Suppose that a man borrows from his bank, giving his promissory note for 60 days. The bank's discount rate is 5 per cent, and the note is for

[1] The banker usually simplifies his calendar by assuming that there are 30 days to the month and 360 days to the year; so the terms 30, 60, and 90 days are equivalent to 1, 2, and 3 months or one-twelfth, one-sixth, and one-quarter of a year.

$4500. How much will the man get, and how much will he repay? A commercial bank will not pay him the full $4500 now but will discount it. If he had borrowed for a full year (which of course he could not have done at a commercial bank), the amount deducted would have been 5 per cent of $4500 or $225. Since he is to have the funds for but a sixth of a year, the charge will be one-sixth of $225, or $37.50. He will therefore receive $37.50 less than the amount of his note, or $4462.50 and, when the note comes due in 2 months, he will repay the full amount of the note, or $4500. Every discount transaction may be figured in this same general way. Knowing the discount rate, we compute what the charge would be on the sum borrowed for a full year. We then note the fraction of a year for which the sum is borrowed, and deduct a charge that is the same fraction of the full year's discount. If a man borrows x dollars for y days at z per cent discount, we first take z per cent of x dollars to find the annual charge, and we then take $y \div 360$ of this annual charge (since the year is thought of as containing 360 days) as the actual discount on the particular transaction. If the student wishes to check his understanding of the process by another example, he may compute that the discount on a 30-day note for $2000, when the discount rate is 3 per cent, will amount to $5.

If, then, we go back to the transaction of Sec. 19 • 4, in which the Keystone Sack Company took its promissory note for $1000 to the bank, the bank would in this case have given $990 for the note if the discount rate was 6 per cent. If the company needed more than $990, it would have to make out a note for more than $1000. Or if the company gave its note to the Foul Line Company, the latter would endorse it and take it to its bank in Amherst, Mass. The Amherst bank would discount it, giving $990 to the Foul Line Company. At first sight it looks as though the seller of the goods, in this case the Foul Line Company, loses by the transaction. They have sold goods priced at $1000 and taken a promissory note for $1000 on which they actually realize but $990. But the loss is only apparent. Most business firms make a practice of giving discounts for cash payment. Thus, although the Foul Line Company quoted its twine at $1000, they would have given the Keystone Sack Company a discount if they had been willing to pay cash. Thus they would not have received the full $1000 if the transaction had been a cash one, and they do not receive the full $1000 when the transaction is a credit one. The price was quoted in the first place on the assumption that the buyer would want credit and that the seller would have to draw a draft that would be discounted. If the buyer is willing to pay cash, the seller can afford to give a discount from the quoted price.

In the transaction of Sec. 19 • 5, the Eagleville Linoleum Company draws a draft for $1500, which it presents at its bank, the Eagleville Bank and Trust Company. The draft has 60 days to run and, if the discount rate is, say, 5

per cent, the bank will charge $12.50 for the credit and will give the linoleum Company $1487.50. At the expiration of 60 days, the bank in Gurleyville will collect the full $1500 from the furniture company and remit it to the bank that extended the credit in the first place.

19·7. Deposits. We have assumed so far that when people borrow, either with a draft or with a promissory note, they receive the proceeds in cash. Thus we have assumed that the linoleum company of Sec. 19·5 took a draft to its bank and received $1487.50 for it. In practice the company probably would prefer not to have such a large sum lying around in cash. It would be dangerous and unnecessary. Instead of receiving the cash, the company will almost certainly prefer to leave the cash at the bank and use it as it wishes. In order to do this, the bank gives the company a deposit. A deposit is not really money in the bank, as we can see when we recall that the company did not take any money to the bank. Most deposits represent borrowing. Instead of giving the borrower cash, the bank gives him a right to demand cash. *A deposit is a right to demand money from a bank.* Sometimes people get this right by actually taking money to the bank and leaving it there for safekeeping, but most deposits in commercial banks arise from credit transactions such as those which we have been describing. A man borrows from a bank on his promissory note and takes a deposit instead of cash; or a bank discounts a draft for the man who drew it and gives him a deposit in place of cash. The bank would be willing to give the cash, but the depositor usually prefers the deposit. He can draw the funds when he wishes, using checks instead of cash, and this will give him a number of advantages, which we shall note later.

Deposits are of two kinds: time deposits and demand deposits. In the case of a *time deposit* the person who has a right to demand money must give notice in advance, if requested. The length of the notice differs, but usually a bank can require 30 days' notice or more on withdrawal of time deposits if it wishes. Of course, the bank can pay such deposits without notice, but it has the right to demand notice if it wishes. Deposits in savings banks (and in the savings departments of commercial banks) are ordinarily time deposits; when a savings bank depositor goes to the bank to withdraw funds, the bank can put him off for 30 days if such postponement is to its advantage. *Demand deposits,* on the contrary, must be paid at any time when the depositor asks for payment during banking hours. Deposits in commercial banks are ordinarily demand deposits, and the demand for their payment is usually made by means of a check. A bank check is merely a form on which the depositor demands payment of part or all of his deposit. The deposit is a right to demand money, and the check is the mechanism by which the demand is made. A check is one form of draft, since it is one person's order directed to a second person requiring him to pay money to

a third person. (To be sure, one may make out a check or other draft payable to himself; in that case, the first person and the third person are the same.) The depositor thinks of the deposit as an asset, of course, since it is something owed to him by the bank. From the standpoint of the bank, however, the deposit is a liability, since it owes the money to its depositor and must be ready to pay it when demanded.

19·8. Operation of a Bank. The actual operation of a bank can best be understood by taking a hypothetical example. Let us suppose that a group of men form a corporation and go into the banking business. As in other corporations,[1] they raise their initial capital by the sale of stock. We shall suppose that our new corporation sells $500,000 worth of stock and in return receives from its stockholders (the owners of the bank) $500,000 in cash. The bank's balance sheet would at this time be as follows:

Assets		Liabilities	
Cash on hand	$500,000	Capital stock	$500,000

If the bank now buys a building and equips it with safe-deposit boxes, safe, computing machines, tellers' cages, etc., at an expense of $300,000, its balance sheet may look as follows:

Assets		Liabilities	
Cash on hand	$200,000	Capital stock	$500,000
Land and buildings	200,000		
Furniture and fixtures	100,000		
	$500,000		$500,000

The bank is now ready to do business. The directors decide to set their discount rate, let us say, at 6 per cent. We must remember that this means a rate of 6 per centum per annum but that the bank will lend for only short periods. If, now, the treasurer of the Eagleville Linoleum Company comes to the bank with the bill of lading and the 60-day draft for $1500 described in Sec. 19·5, the bank discounts the draft and charges $15 for the transaction (this being the charge for $1500 for 60 days at 6 per cent). They give the linoleum company a deposit for $1485 and hold the draft for $1500 as an asset. Their balance sheet will now be altered, and perhaps it will be best for us first to show merely the changes that would be made in the assets and the liabilities. These would be the following:

Assets		Liabilities	
Loans and discounts	+ $1500	Deposits	+ $1485
		Undivided profits	+ 15
Net changes	+ $1500		+ $1500

[1] See Chap. 9.

The draft is listed under the heading of "loans and discounts." Sometimes a bank uses merely the word "loans" or merely the word "discounts" to represent drafts and promissory notes held by it. Of course, these represent amounts owed to the bank by people who have borrowed from it, and are therefore assets of the bank. The deposit, as explained in the preceding section, is the bank's liability. But the deposit amounts to only $1485 while the draft amounts to $1500. The bank has made a profit of $15 on the transaction (or expects to when it collects the draft 60 days later), and this it lists as a liability for reasons explained in Chap. 9. If the balance sheet balanced before, it will balance now, since we have added $1500 to each side. In fact, the new balance sheet will appear as follows:

Assets		Liabilities	
Cash on hand	$200,000	Capital stock	$500,000
Land and buildings	200,000	Undivided profits	15
Furniture and fixtures	100,000	Deposits	1,485
Loans and discounts	1,500		$501,500
	$501,500		

On any particular transaction, we can show either the entire balance sheet before and after or, more simply, the changes that took place in the balance sheet as a result of the transaction. In the latter case, it is necessary to mark each change plus or minus to show whether the amount of the asset or liability involved was increased or decreased.

Suppose that a man comes to our bank now and brings $100,000 in cash, which he leaves on deposit. The changes in our balance sheet will be these.

Assets		Liabilities	
Cash	+ $100,000	Deposits	+ $100,000

The transaction has been a big one, but the bank has made nothing at all by it. It is important for depositors to understand that *a commercial bank makes its profits by lending money.* One does not do a favor to a bank when he takes money there and leaves it on deposit. The bank makes no profit from it but does incur expenses for bookkeeping, sending out statements, etc. For this reason most commercial banks make a small service charge on small deposits. This is to cover, at least in part, the expense that the bank incurs from carrying the deposit. It is when one borrows from the bank that the bank makes profits.

If, at the end of 60 days, our draft comes due and is paid off (we shall illustrate again with the same $1500 draft mentioned two paragraphs back), our balance sheet will be changed again. The particular changes that are made will depend on the way in which the draft is repaid. If the doctor

brings in $1500 in cash and pays it to the bank, the changes in the balance sheet will be as follows:

Assets		Liabilities	
Cash	+ $1500	No change	
Loan and discounts	− 1500		$0000
	$0000		

The bank receives the cash and gives back the draft. The businessman no longer owes the money to the bank, so the loans and discounts item is reduced by $1500; but the bank has $1500 in cash which it did not have before. There are no changes whatever on the liability side, but there are two changes on the asset side which cancel each other out.

But the merchant might, of course, have saved up money from day to day during the 60-day period, putting it on deposit with his bank each night. In this case, he would be likely at the end of the 60 days merely to write a check for the $1500 instead of bringing in the cash. In this case, the changes in the balance sheet would be different. The bank would receive no cash, but the depositor would give up $1500 of deposits. The changes would be the following:

Assets		Liabilities	
Loans and discounts	−$1500	Deposits	−$1500
	−$1500		−$1500

This is a good place to point out to the student that the total assets and total liabilities are not of any great significance. Here we have had the same draft paid off in two different ways: by cash and by surrender of deposits. It makes little difference to the bank which method is used; yet, in one case the bank's assets and liabilities are decreased and in the other case they are not.

A bank is not likely to keep all its cash in its own vaults but is apt to deposit some of it in other banks. If the bank is a member of the Federal Reserve System, it will be required by law to keep a deposit at its Federal Reserve Bank. This is a matter that we shall describe later.[1] In such a case, it may lump all its cash together in its balance sheet, or it may publish it in more detail, showing which cash is actually in its vault and which it holds in reserve at the Federal Reserve Bank. In the latter case, the bank's balance sheet would show an item called "Reserve with Federal Reserve Bank" among the assets as well as the item of cash on hand. Also the bank may invest some of its funds in government bonds or in other securities,[2] in

[1] See Sec. 20 • 10.

[2] Banks are not usually allowed to invest in all sorts of securities, but are limited by law to those securities which are thought to be safest and least speculative.

which case there may be listings of government bonds and of other securities among the assets.

Suppose that one of our depositors comes into the bank this morning and deposits his salary check, which is drawn on a bank in a distant city. Perhaps the check is for $200. We give him a deposit for $200 right away, yet we do not have the cash to show for it. The check is a means by which we may demand cash from his bank in the distant city, but so far as we have the check and not cash. How can we show this transaction on our books? We shall have increased the deposits on the liability side of our balance sheet by $200, and we must have some offsetting item. We realize that other distant bank owes our bank the money and will have to pay it when we demand it; so we merely list on our asset side $200 as due from other banks. Thus the changes made in our balance sheet by the deposit of the check would be as follows:

Assets		Liabilities	
Due from other banks	+ $200	Deposits	+ $200

Let us illustrate the meaning of one other asset of a bank by another example. Suppose that one of our depositors makes an arithmetical mistake on the stubs of his checkbook, and as a result his record shows that he has $150 on deposit when he really has but $15. This depositor now buys a new suit of clothes for $50, and gives his check to the clothier for that amount in good faith, thinking that he has sufficient funds to cover the check. When the check gets to the bank, it is discovered that the depositor has a right to demand but $15, yet he has drawn a check for $50. The bank may, of course, refuse the check, stamp it "No Funds," and return it to the clothier who deposited it. This would be a perfectly reasonable thing for them to do, yet, of course, it will be embarrassing to the depositor when he is told that his check "bounced." Had he done it intentionally, it would have been a serious affair, called "passing bad checks," but we are assuming that it was based solely on an error. In such a case, if the amount is not too large and if the bank has confidence in the depositor, they may do him the favor of accepting the check and giving the clothier the full $50 without saying anything to him about it. They will, of course, notify the depositor immediately that his account is overdrawn by $35 and, if he is an understanding person, he will immediately come to the bank and make it good, either by depositing $35 in cash or by giving his note and borrowing the funds. He will realize that the bank has done him a favor to save him from embarrassment, and he will rectify his error at once. But in the meantime, how is the bank to show the facts? They have paid out $50 in cash and have canceled the $15 deposit which the depositor had left. But how are they to show the facts concerning the other $35? They do it by listing $35

of overdrafts among the assets. Overdrafts are assets from the standpoint of the bank, because they are owed to the bank by the depositor. The changes that would take place in the bank's balance sheet as a result of this affair would be as follows:

Assets		Liabilities	
Cash	−$50	Deposits	−$15
Overdrafts	+ 35		
	−$15		−$15

Although banks list a very large number of different assets and liabilities, we have now mentioned and explained the more common and more important of the assets. A few additional liabilities should also be mentioned. Just as a manufacturing or mercantile corporation often has a surplus in addition to its capital stock and its undivided profit,[1] so banks often carry surpluses. In fact, financial institutions such as banks and insurance companies usually carry much larger surplus items than do other kinds of corporations. It is not uncommon for financial corporations to have surpluses far greater than the amounts of the capital stock. Likewise, just as a bank carries an asset item called "due from other banks," if often carries a liability item called "due to other banks," and both items arise from the same type of transaction. One bank owes money to a second bank. The money is due to one of them and due from the other. National banks still occasionally have bank notes outstanding—the national-bank notes which are now being retired and which we described at some length in Chap. 18. These bank notes are money, and the man who holds them thinks of them as assets. But they are the bank's promises to pay money, and from the standpoint of the bank they are liabilities. Thus the balance sheet of a commercial bank might contain the following items, in addition to some others that are either seldom found or of minor importance:

Assets	Liabilities
Cash on hand	Capital stock
Reserve with Federal Reserve Bank	Undivided profits
Land and buildings	Surplus
Furniture and fixtures	Deposits:
Loans and discounts	Demand deposits
Due from other banks	Time deposits
Overdrafts	Due to other banks
Government bonds	Bank notes outstanding
Other securities	

The student should make sure that he understands the meaning of each of these items and knows how they arise in practice.

[1] For explanation see Chap. 9.

19 · 9. Bank Clearings. We have already seen that most bank deposits arise without the depositor's having taken any money to the bank. When a deposit does arise from the depositor's actually leaving money at the bank for safekeeping, it is called a "primary deposit," as distinct from "derivative deposits," which arise from lending transactions. Now we should note that not only are most deposits derivative, arising without the taking of any money to the bank, but also that in most cases no money will ever be withdrawn from the bank against the deposit. Deposits not only are not money in themselves, but they usually arise without the use of money and disappear without the use of money.

This would be easiest to see if we were to consider first an artificially simple case. Let us imagine that there is but one bank in a city, and everyone there does business with it. Let us suppose, moreover, that every business transaction is settled by check rather than by money. Under such circumstances, a moment's reflection will show that the bank would not have to keep any money at all. When Mr. Waggoner, the butcher, buys shoes from Mr. McPeek, the cobbler, he pays by check, and Mr. McPeek deposits the check in his bank—the only bank that there is. The bank deducts the amount from Mr. Waggoner's account and adds it to Mr. McPeek's. When Mr. McPeek consults Dr. Gilman about his lumbago, he pays Dr. Gilman by check, and Dr. Gilman deposits it in the bank. The bank deducts the amount from Mr. McPeek's account and adds it to Dr. Gilman's. Finally, when Dr. Gilman buys his Thanksgiving turkey from the butcher, Mr. Waggoner, the transaction is again handled by check, which Mr. Waggoner deposits at the bank. The bank now adds to Waggoner's account and deducts from Gilman's. No money has changed hands on these transactions. The only changes have been on the books of the bank. It is quite possible, of course, that the shoes, the advice about lumbago, and the turkey were all items of exactly the same size. If, for example, each of the transactions involved $5, then the three men have exactly the same balances at the bank after the transactions as before. Each has been credited with $5 once and debited with $5 once. Fifteen dollars' worth of transactions have taken place, and not a penny of actual cash has changed hands; yet all the debts have been settled. It is evident that, under these hypothetical circumstances, the use of the bank made it possible to get along without money. The bank in effect canceled off one debt against another.

If, now, we can become somewhat more realistic, let us imagine that there are two banks in the town and that some people do business with one and some with the other. The two banks are the Gurleyville Trust Company and the First National Bank of Gurleyville. Under these circumstances, when the butcher, Mr. Waggoner, buys his shoes from Mr. McPeek and pays by check, it is quite possible that the two men will be patrons of

different banks. Mr. Waggoner gives Mr. McPeek a check, let us say, on the Gurleyville Trust Company, and Mr. McPeek deposits the check in the First National Bank. The First National Bank will, of course, add the amount to Mr. McPeek's deposit, but they cannot subtract it from Mr. Waggoner's deposit, since he has none at their bank. The amount, as we have already seen, will be listed as "due from other banks," (sometimes it is called "cash items in the process of collection," which means the same thing.) How is the National Bank, which has given Mr. McPeek a deposit, to be reimbursed? The answer is that the National Bank could send a messenger across the street to the Trust Company, have the check cashed, and bring the check back. Each time either bank received a check drawn against the other, they could send a messenger across to bring back the cash. This, however, they would not really do. Instead, each bank would set aside the checks received during the day, and when the bank closed these checks would be sorted. Many of them would probably be checks drawn against the other local bank. Let us suppose that, when the day is done, the Trust Company finds that it holds checks drawn on the National Bank to a total amount of $3500; and that the National Bank holds checks drawn on the Trust Company to a total amount of $3600. Each bank could presumably demand payment on all the checks. The National Bank could collect $3600 from the Trust Company, and the Trust Company could collect $3500 from the National Bank. This would obviously be foolish, and in practice only the difference would be paid. The Trust Company in this case would pay $100 to the National Bank, the two banks would exchange the checks, and the whole thing would be settled. A total of $7100 of checking transactions would have been settled with the use of but $100 in cash. All the other transactions would have canceled each other out. Thus we see that for most of these transactions no money is paid out on the deposits, just as no money was put in to originate the deposit. The deposits are money in name only. Actually, they represent credit rather than cash.

Let us now consider a city larger than Gurleyville—one in which there are several banks. At the end of the day each bank might send a messenger to each of the other banks in the way that we have described, and each bank would make one single settlement to cover all its transactions with each of the other banks. If there were 20 banks in the city and if each pair of banks settled separately, there would be 190 separate settlements to make.[1] It is possible, however, to accomplish the same results much more easily, through what is called a "clearing house." In our city of 20 banks, each bank would sort out its checks at the end of the day and send a messenger to some

[1] This is the number of combinations of 20 things taken 2 at a time. In general, if there are n banks in a city and if each pair settles the differences, the number of settlements will be

$$(n)(n\text{-}1)/2 = (n^2\text{-}n)/2$$

common meeting place, the clearing house. This may be merely the director's room in one of the banks, or it may be in a separate building. It really is but a meeting place for the messengers of the banks. Each messenger arrives at the clearing house with the checks on the other local banks, all sorted out and with the amounts totaled. It is easy then for a clerk at the clearing house to figure out how much is owed to each bank by all the other banks in town. This will be the sum of the checks brought in. But it is also easy to compute how much each bank owes to all the other banks, since this is the sum of the checks brought in by messengers of the other banks and drawn against this bank. It would take up too much room here to illustrate with 20 banks but the principle involved will be clear if we illustrate with four banks, which we shall call bank A, bank B, bank C, and bank D. Let us suppose that the clerks at bank A discover, when the day is done, that they have checks drawn against Bank B to the amount of

Table 19·1. Hypothetical Clearing House Account

Amount owed to each bank	Amount owed by each bank				Total
	Bank A	Bank B	Bank C	Bank D	
Bank A		$25,000	$ 8,000	$19,000	$ 52,000
Bank B	$30,000		5,000	21,000	56,000
Bank C	10,000	12,000		6,000	28,000
Bank D	40,000	30,000	12,000		82,000
Total	$80,000	$67,000	$25,000	$46,000	$218,000

$25,000. They have checks drawn against bank C to the amount of $8,000, and checks against bank D to the amount of $19,000. These they sort out into three piles and fasten them together with rubber bands, including in each pile a slip showing the total amount of the checks enclosed. These checks are then sent off to the clearing house, where messengers are also coming in from the other banks with similar piles of checks. Bank B sends in, let us say, $30,000 of checks against A, $5000 against B, and $21,000 against D. Bank C sends in $10,000 against A, $12,000 against B, and $6,000 against D. Finally bank D sends in $40,000 against A, $30,000 against B, and $12,000 against C. The piles of checks are handed to a clerk, who takes off the totals to form a table, such as Table 19·1. In this table we see the figures just given for the individual banks, arranged now in columns and rows. Each row shows the amounts owed to the bank named at the left by each bank listed at the top. Each column shows the amounts owed by the bank named at the top of each of the banks listed at the left. The sums of the rows show us the total amounts owed to each bank; the sums of the columns show

the total amounts owed by each bank. Thus, bank A is owed $52,000 by the other banks and owes $80,000 to them. Bank A will therefore pay the difference of $28,000 to the clearing house. Bank C is owed $28,000 by the other banks and owes them $25,000, so it receives the difference or $3000 from the clearing house. Inspection of the totals in the table shows that bank A pays $28,000 and bank B pays $11,000 to the clearing house, a total of $39,000 paid in. On the other hand, bank C receives $3000 from the clearing house and bank D receives $36,000—a total of $39,000 paid out by it. From the nature of the transactions, it is evident that the amount that some banks pay to the clearing house will be exactly equal to the amounts that other banks receive from it. In fact, the clearing house is merely a method for settling the debts and, since each debt is owed to someone and also owed by someone, the total amounts "owed to" and the total amounts "owed by" must exactly balance.

In our example $39,000 in cash changed hands. Even this might well have been unnecessary, since bank D's gain of today might be offset by a loss tomorrow, and it is quite possible to carry the amounts on the books with only an occasional settlement. Even if the entire $39,000 balance were settled in cash today, it should be noticed that we would have settled $218,000 worth of checks with a payment of but $39,000 in cash. Most of the checks would have been canceled off and never paid at all. When a man buys a new lamp for his wife's birthday and pays for it by check,, he is apt to say that he paid cash for it, yet we now see that it is probable that no cash was actually involved in the transaction. The check will, in all likelihood, be canceled against someone else's check. This process of canceling checks rather than paying them is called "clearing" checks.

Later, when we have studied the present American banking system, we shall find that checks on out-of-town banks are also usually cleared and not paid in cash; when we study foreign trade, we shall see that international indebtedness also is usually cleared and canceled off without the payment of cash.[1]

19·10. Bank Reserves. We have described the various items that make up a bank's balance sheet almost as though they were separate and disconnected, although our description of bank operations in Sec. 19·8 shows that they are really tied together more or less closely. For example, we have seen that the item of loans and discounts is connected with the terms of deposits and of undivided profits, since when the bank discounts a note it increases the loans and discounts, the deposits, and the undivided profits. The major business of a commercial bank in this business of deposit and loan, and the more such business it can do the more profit it will make. Yet there are limits, as can easily be seen. We noted in Sec. 19·9 that a

[1] See Secs. 20·7 and 24·3.

bank could get along without any cash at all if everyone were willing to do business by check and if interbank items exactly balanced on clearing. But, in practice, banks do need small amounts of cash occasionally to settle clearings balances, and also people do come to the bank to cash checks. The bank has to be ready to pay checks in cash on demand. Every depositor has a right to demand cash, even though in the ordinary course of business he will seldom do so. Therefore the bank does have to keep some cash on hand for these exigencies. The amount necessary for the banker to keep depends on many things. It depends on the time of year, since usually more cash is required around Christmas time or, in agricultural regions, at the time of planting or harvest. It varies according to the nationality of the bank's clients and their familiarity with the use of checks. It varies with the size of the city in which the bank is located, and many other things. The banker has to learn by experience how much cash he has to keep on hand to be safe, because, although a bank makes more profit if it lends more, it is safer if it lends less. This is clear when we remember that every loan increases the bank's deposits—the amount of cash which people have a right to demand—yet it does not increase the cash available for payment. The bank's officers are constantly torn between the desire for profit and the desire for safety. The profit motive urges them to lend more and more and the safety motive urges them to lend less and less. Somewhere they have to compromise.

In most countries the decision on this matter is left to the bankers themselves. It is assumed that their own experience will tell them how far they can safely go. In the United States, however, this matter is regulated by law. For a state bank it is regulated by the laws of the state that chartered it; for a national bank by the laws of the United States. The state laws are too many and too diverse for us to take up here,[1] but the Federal laws cover all national banks and also those state banks which are members of the Federal Reserve System. Thus almost half the commercial banks of the United States are governed by the Federal laws.[2] These laws divide the banks into three classes, primarily according to the size of the city in which the bank is located. New York and Chicago are called "central reserve cities," and the member banks in these cities are called "central reserve city banks." Then comes another group of large cities, at present 58 in number,[3] called "reserve cities," and the member banks in these cities are called "reserve city banks." All other member banks are called "country banks." For many

[1] The Federal Reserve Board has issued a compilation of the state laws relating to bank reserves as they existed on Dec. 31, 1944, copies of which can be had on request to the Board.

[2] On June 29, 1946, there were 14,026 commercial banks in the United States, of which 6887 were members of the Federal Reserve System.

[3] See Sec. 20 • 4.

years the law required member banks in the central reserve cities to maintain a reserve equal to at least 13 per cent of their demand deposits. This reserve must be kept at the Federal Reserve Bank. Reserve city banks were required to keep a reserve with their Federal Reserve Bank of 10 per cent of their demand deposits; and country banks a reserve of 7 per cent of their demand deposits. The required reserve on time deposits was 3 per cent regardless of the city in which the bank was located. Nothing can be counted as reserve save funds deposited with the Federal Reserve Bank. Since 1935, under the provisions of the Banking Act of that year, the board of governors of the Federal Reserve System has the power to set the reserve requirements for these different classes of banks anywhere between the minima of 7, 10, and 13 per cent then in force, and double those amounts. Thus the limits within which the minimum reserves can now be set are:

	Demand deposits, per cent	Time deposits, per cent
Central reserve city banks	13–26	3–6
Reserve city banks...	10–20	3–6
Country banks ...	7–14	3–6

Under the provisions of this act the minimum reserve requirements have actually been altered several times, as shown in Table 19·2. At the time of this writing (April, 1947), the requirements are that member banks

Table 19·2. Minimum Reserve Requirements for Banks That Are Members of the Federal Reserve System (Per cent of deposits)

Period in effect	Net demand deposits			Time deposits (all member banks)
	Central reserve city banks	Reserve city banks	Country banks	
6/21/17–8/15/36	13	10	7	3
8/16/36–2/28/37	19½	15	10½	4½
3/1/37–4/30/37	22¾	17½	12¼	5¼
5/1/37–4/15/38	26	20	14	6
4/16/38–10/31/41	22¾	17½	12	5
11/1/41–8/19/42	26	20	14	6
8/20/42–9/13/42	24	20	14	6
9/14/42–10/2/42	22	20	14	6
10/3/42 and since	20	20	14	6

in central reserve cities and in reserve cities keep at least 20 per cent of the amount of their demand deposits, and country banks at least 14 per cent of their demand deposits, and that all members banks keep at least 6 per

cent of their time deposits, on reserve with the Federal Reserve Bank of their district. We can easily discover from the bank's balance sheet the amount of the time and demand deposits, and from these we can discover the minimum amount of reserve that the bank must have with its Federal Reserve Bank.[1] The total reserves actually held for all country banks on June 21, 1946, amounted to 4,432 million dollars, but the minimum amount which they could legally have held on that date was 3,968 million dollars. Thus, these country banks held 664 million dollars of excess reserves on that date, and all member banks together held 912 million of excess reserves.[2]

A bank may hold excess reserves because it does not feel that it is safe to lend as much as the law allows, or because it cannot find what it considers good business risks to which it can lend. If it holds excess reserves without good reason, it is sacrificing profits which it could otherwise make by lending. Thus, banks tend to reduce their excess reserves to as low a point as seems to them safe. Every time the bank lends money and gives a deposit, it increases its total deposits without increasing its reserve, and thereby decreases its *reserve ratio*—the ratio of reserves to deposits. The reserve ratio is the percentage which the reserve is of the deposits, so that a bank with deposits of $500,000 and maintaining a reserve of $250,000 with its Federal Reserve Bank has a reserve ratio of 50 per cent.

19·11. Methods of Controlling the Reserve Ratio. Since a bank is required by law to maintain some minimum reserve ratio, it is important to know how that ratio may be controlled. If it gets too low, how can it be increased? If it gets too large, how can it be reduced? There are in fact a number of procedures available by which the officers of a bank can alter the reserve ratio. The effects of these procedures can be most easily visualized if we start with a very simple bank statement and see how it would be altered by each sort of practice. Let us assume, therefore, that a bank statement at the moment appears as follows:

Assets		Liabilities	
Loans and discounts	$1,137,000	Capital stock	$ 300,000
Government securities	1,502,000	Surplus	200,000
Other securities	366,000	Undivided profit	32,000
Real estate	50,000	Demand deposits	3,250,000
Furniture, etc.	20,000		$3,782,000
Reserve with Federal Reserve Bank	650,000		
Cash on hand	57,000		
	$3,782,000		

[1] The process is slightly more complicated than has been described here, since various adjustments are made in the figures before computing the reserve requirements; but since we are trying here to give a general idea of the problem and not to train bank examiners, the description here will serve for our purposes.

[2] Data from *Federal Reserve Bulletin*, vol. 32, no. 7, July, 1946, p. 767.

Suppose that this bank's directors meet today, and that the bank is in a reserve city where the minimum reserve requirements are 20 per cent. This bank has deposits of $3,250,000 and reserves of $650,000, so that its reserve ratio is $650/$3250, or exactly 20 per cent. Evidently the bank is just on the ragged edge of the legal requirement, and the directors may well decide that they ought to do something to increase the reserve ratio at once.

One thing to do very evidently is to sell some of their securities. If they sell these to the Federal Reserve Bank, they will immediately increase their deposit there; if they sell them to the general public and collect the funds, they can send the funds for deposit to their Federal Reserve Bank. If we take the simpler case, let us assume that this bank sells $65,000 worth of government securities to its Federal Reserve Bank. The changes in the balance sheet will be as follows:

Assets		Liabilities	
Government securities	−$65,000	No change	
Reserve with Federal Reserve Bank	+ 65,000		$000
	$ 000		

Although no change has taken place in the total assets or total liabilities, nevertheless a change has taken place in the reserve ratio. The deposits remain unchanged at 3,250,000, but the reserve has been increased by $65,000 to $175,000. The reserve ratio is now $715/$3250, or 22 per cent. If this is not safe enough, the bank can sell more securities or use some of the other methods that are available. Buying securities tends to reduce the reserve ratio; selling securities tends to increase it.

We have noted that some loans are made for definite periods, while others are made payable "on call" or on demand. If this bank has some call loans outstanding, as most commercial banks do, it can also increase its reserve ratio by calling some of them. This means that it demands immediate repayment of the loans. If our bank calls $100,000 of such loans and receives cash for them, the changes in its balance sheet will be the following:

Assets		Liabilities	
Loans and discounts	−$100,000	No change	
Cash on hand	+ 100,000		
	$ 000		$000

The new $100,000 of cash could now be sent to the Federal Reserve Bank, thus increasing our deposit there by $100,000. If the appropriate changes were made in the original balance sheet, we would find the reserve increased by $100,000 from $650,000 to $750,000, while the deposits remain at $3,250,000.

The reserve ratio is then \$750/3250, or 23 per cent, instead of the original 20 per cent.[1] Calling call loans tends to increase the reserve ratio, and this is one of the reasons why commercial banks like to have at least a small amount of money lent out "on call."

A third thing that the bank can do, and one of the most common in practice, is to *rediscount* some of its commercial paper. *Commercial paper* consists of drafts and promissory notes which will mature in 90 days or less and which have arisen out of commercial transactions. In other words, they are documents representing short-term production credit. Businessmen sometimes refer to these documents as "short-term self-liquidating paper," but the phrase "commercial paper" is their most common designation. We have already learned how the commercial bank comes into possession of commercial paper by discounting it. In effect, the bank buys the paper at a discount from the man who holds it. Thus, they may buy a \$1500 draft by paying (giving a deposit for) \$1487.50, if we use again the example of Sec. 19 • 6. But if the draft is negotiable—that is, if it is payable "to the order of" someone—the bank can sell it again, turning it over to someone else who will collect it when it falls due. This would be the process of rediscount—discounting over again; and the process is exactly the same in every essential detail as the original process of discount. The bank may sell the draft to anyone, but the common thing is for them to sell it to their Federal Reserve Bank.[2] Let us assume that the bank that bought the draft of Sec. 19 • 6 has held it for a month and now wishes to rediscount it. The note had originally 60 days to run; it has but 30 days left before maturity. Let us suppose that the rediscount rate at the Federal Reserve Bank is 2 per cent. This, we assume, is the rate at which the reserve bank discounts commercial paper that has already been discounted by one of its member banks. Our bank sends the \$1500 note with 30 days to run to its reserve bank, which discounts it at the rate of 2 per centum per annum. This means that the reserve bank deducts \$2.50 and gives our bank a deposit for the remaining \$1497.50. Our bank had expected to make a total of \$15 profit when it discounted the draft in the first place, and it added \$15 to the undivided profit at that time.[3] But now it pays out \$2.50 of this as a rediscount charge and has to reduce the undivided profit accordingly. The changes in the balance

[1] It is here assumed that the bank called these loans *instead* of selling securities, not that it did both. In other words, the illustration of call loans has now been substituted for the illustration of bond sales, not added to it. If the bank had carried out both expedients, the reserve would have risen to \$815,000, and the reserve ratio would be 25 per cent.

[2] This, by the way, is the way in which the Federal Reserve Bank gets the commercial paper that it uses as partial reserve behind Federal Reserve notes—the commercial paper that gives the elasticity to our currency (see Sec. 18 • 15).

[3] This part of the transaction was described in Sec. 19 • 8.

sheet of the member bank as a result of rediscounting this note are the following:

Assets		Liabilities	
Loan and discounts	— $1500.00	Undivided profits .	—$2.50
Reserve with Federal Reserve Bank	+ 1497.50		
	—$ 2.50		—$2.50

If these changes were made in the original balance sheet given at the beginning of this section, the reserve with the Federal Reserve Bank would be increased by $1497.50 to $651,497.50, while the deposits would remain unchanged at $3,250,000. The reserve ratio would become 651,497.50/3,250,000, or 20.046 per cent. The reserve ratio has risen only very slightly, because the transaction was relatively small. If we had rediscounted several thousand dollars' worth of commercial paper, the effect would have been greater. Even this small transaction sufficies, however, to show the general principle: rediscounting commercial paper increases the reserve ratio.

Our bank directors, worried by the low reserve ratio, might also raise the discount rate of our own bank. The bank has been discounting commercial paper for its customers at, say, 6 per cent. If we raise our discount rate to 10 per cent, men who had intended to borrow will be discouraged from doing so. The cost will be too high. They will get their credit somewhere else or do without it entirely. As far as we are concerned, this will be an advantage. Our loans will come due and will be repaid, and we will not be making new ones; or, at least, we shall be making fewer new ones than would otherwise have been the case. Thus, we shall increase our cash or decrease our deposits as the old loans are paid off, depending on whether they are paid off by cash or by check. In either case, we shall improve our reserve position. If the loans are repaid to us in cash, we can deposit it with the Federal Reserve Bank and increase our deposit there. If the man repaying a loan pays by check and thus surrenders his deposit, we shall reduce the amount of our deposits. Either one will increase our reserve ratio.

Finally, we might borrow from some other bank on our own promissory note. This would increase our item "due to other banks" and would increase our cash, which we could send to the reserve bank for deposit. Since our reserve is increased and our deposits are not, our reserve ratio is increased. Thus the five major ways in which we might seek to raise our reserve ratio would be (1) by selling securities, (2) by calling call loans, (3) by rediscounting commercial paper, (4) by raising our discount rate, and (5) by borrowing from other banks.

19 · 12. Advantages of Loan and Deposit as Compared with Cash. We have seen that the use of deposits in commercial banks, to a considerable

extent, take the place, of cash. There are several reasons why people prefer to use bank deposits and checks rather than money for some kinds of transactions, even though there are many cases where checks cannot be used at all. We do not class checks as money, even though they are used as mediums of exchange, because they are not generally acceptable. Anyone who has tried to pass a check in a strange town has discovered this fact. Currency, except for the minor risk of counterfeiting, is good at its face value, whereas a check is good only if drawn in proper form by someone who has sufficient funds in the bank and if presented by the proper payee.

The advantages of checks are numerous for those kinds of transactions in which they can be used. To begin with, their use minimizes the danger of loss or theft. Stolen money usually cannot be traced and recovered. If destroyed by fire or accident, it is usually beyond recovery. The loss of a check is, however, a minor matter. Another check may be drawn to take its place, the precaution being taken of "stopping payment" on the original check.[1] If one loses his billfold full of money, the loss is a real one; if one loses his checkbook, the loss is negligible.

The use of checks, moreover, gives at least a rough account of expenditures—a record of where the money went. Both the stubs in the checkbook, and the checks themselves when they are returned by the bank with the monthly statement, show how the money was spent. The returned and canceled checks are also useful as receipts for payment. If a bill is paid in cash, it is difficult to prove payment unless a written receipt was obtained at the time, or unless witnesses to the payment can be found. Proof is particularly difficult if the person who received the payment has died. But, if the payment was by check, the written endorsement on it serves as acknowledgement that the payment was received, and shows who received it. When the bank pays the check, it stamps or perforates the word "Paid" on it and returns it to the person who drew it. The check may thereafter be retained as a receipt for the payment. In fact, the endorsements on the back of a check make it possible to trace the course taken by it from the time it left the hands of the man who originally drew it until it came back to his bank for payment.

Not only are checks safer to use than cash, but they are also much easier to handle for many transactions. For example, if one wishes to send the sum of $319.78 through the mail and uses cash, he must send at the least three one-hundred-dollar bills, one ten-dollar bill, a five-dollar bill, two two-dollar bills, a half dollar, a quarter dollar, and three cents—twelve separate pieces.

[1] One "stops payment" on a check which is lost or which was issued in error, by notifying the bank, before the check is presented for payment, that the check is not to be paid. The check is then without value to anyone who has stolen it or found it, or to anyone who has received it in error. This ability to stop payment is another advantage of checks over cash.

But checks have the advantage that they can be made out in odd amounts, and a single check can be made to cover the amount of $319.78 or any other amount. Even sums of millions of dollars are paid by check, without the difficulties of carriage or of protection that would be involved if cash were used. To be sure, there is no likelihood that checks will replace cash for small purchases or for purchases in which goods are to be delivered immediately to strangers, as in everyday retail trade. But their great advantages have already made them the major means of payment for many sorts of commercial transactions.

The endorsement of a check, as we have said, serves as a receipt for payment; but it is much more than a receipt. When a man endorses any negotiable instrument, such as a check, a promissory note, or a draft, he makes himself liable for payment of the instrument in case the original drawer does not pay. Anyone who takes title to the instrument after the endorser's name is on it can collect from him if the instrument is not paid in the regular course. Thus, let us suppose in our example of Sec. 19·9 that the butcher, Waggoner, had bought a pair of shoes for $5 from the cobbler, McPeek, and had paid with a $5 check. McPeek then takes his troubles to Gilman and, being charged $5, endorses the check and gives it to Gilman. Gilman buys a new hat for $5 from his haberdasher, McReynolds, endorses the check in his turn, and gives it to McReynolds. They may go on and on, but we shall assume that McReynolds endorses the check himself and sends it, bearing the original signature of Waggoner and the endorsements of McPeek and Gilman in addition to his own, to his bank for deposit. If, now, it is discovered that Waggoner really had no funds at the bank and that the check is worthless, the bank will refuse to give McReynolds a deposit. McReynolds, in turn, may collect from anyone whose name appeared on it before his; that is, from Gilman, McPeek, or Waggoner. If Waggoner passed a bad check to begin with, it is unlikely that it will do much good to look to him for payment, so McReynolds is likely to call on Gilman or McPeek for payment. They, in turn, can demand payment from those whose endorsements preceded theirs. Thus the check "backs up" over its old route, perhaps skipping some steps now and then, until someone is left holding a check with no predecessor who can or will pay. To be sure, when a bad check is passed, someone loses. This is the reason that checks are not generally acceptable and that strangers have trouble passing them. As they gather more and more endorsements, they are generally better and better risks. It pays to remember, when one endorses a check, that he has made himself liable for payment on it.

This matter of liability is characteristic of endorsements on all sorts of negotiable paper, not merely on checks. And it is particularly important in the case of drafts and promissory notes. When one borrows on a promissory

note, it is not uncommon for the lender to require the borrower to get an endorser, someone who will endorse the note for him. The endorser now shares the responsibility. If the original borrower does not pay, the endorser must. Many college and university loan funds insist that students who borrow get the endorsement of their fathers or of some responsible businessman. When a note or other negotiable instrument has been endorsed, it is called "two-name paper." It bears the name of the original borrower and also that of an endorser. Thus the security behind it is better than it would be on "one-name" paper.

A trade acceptance is always two-name paper, as will be seen when we recall how it arises. In Sec. 19 • 6 the Eagleville Linoleum Company drew a draft for $1500, which it took to its bank for discount. The bank sent the draft, with the bill of lading, to a correspondent bank[1] in the neighborhood of the buyer, and this bank called the Furniture Company and had them accept the draft. The signature of some responsible officer of the Linoleum Company was put on the draft when it was first drawn. The signature of some responsible officer of the Furniture Company is added when the draft is accepted and becomes an acceptance. The acceptance is, therefore, two-name paper. If the Furniture Company does not pay its debt when it falls due in 60 days, the Linoleum Company will have to make good. It may seem strange at first that the bank in Eagleville was willing to give a deposit in return for a draft drawn against some firm in a distant city—some firm with which it had probably never had dealings. We now see that it is not necessary for the bank to know the buyer of the goods. The bank is receiving the draft from the seller of the goods, who is one of its regular customers. His name appears on the draft and, if he is reliable and responsible, the draft will be good. This is another reason for preferring the draft and bill of lading as a method of financing commercial transactions. Before the draft is accepted, the bank is protected in two ways: (1) it can collect from the man who drew the draft, since his name appears on it, or (2) it can sell the goods (since it has the bill of lading) and collect the money from their sale to pay the debt. After the draft has been accepted, the bank gives up the bill of lading; but it now holds the acceptance, which is two-name paper on which both the buyer and the seller of the goods are responsible. The seller of the goods got his money (minus discount) right away, which is what he wanted. The buyer of the goods got the goods right away and credit for 60 days,

[1] When two banks agree to help each other out by presenting and collecting drafts for each other, they are called "correspondent banks." In the example of Sec. 19 • 6, the student may have wondered how the bank that presented the draft to the Furniture Company and later collected it when it came due, profited by the transaction. It did not. On other transactions, the business will go in the other direction, and this bank will be sending drafts to be collected for it gratis by the Eagleville Bank and Trust Company.

which is what he wanted. And the bank got $12.50 of profit on the transaction (see Sec. 19·6) which is what it wanted!

The system of loan and deposit thus works out to the satisfaction of all parties concerned, giving credit where credit is needed. If the seller is to be able to get cash at once and the buyer is to be able to wait, someone else must put up the cash in the meantime—someone else must finance the transaction. This is exactly what the commercial bank has done by its system of deposit and loan; and, in the meantime, those involved have been protected by the use of credit instruments of the various kinds that we have described. Without the bank, either the buyer would have had to postpone his purchases until he could pay cash or else the seller would have had to finance him by waiting for his money. Specialization of labor is advantageous here as anywhere else. The bank is a specialist in granting credit, and can do it more cheaply than the ordinary businessman who does not understand the credit business nor have the funds to invest. The use of the commercial bank makes it possible to carry on business transactions smoothly with a maximum of security and a minimum of red tape.

19·13. Bases of Credit. Credit is the ability to get valuable things in the present in return for a promise to pay a future equivalent. Almost everyone has some amount of credit, and the more he uses the less he has left. People differ greatly in their credit. Some have real difficulty in borrowing a quarter of a dollar, while others are able rather easily to borrow tens of thousands of dollars. Many factors influence the credit that any man or business firm can command, but someone has summarized them fairly well with four words beginning with the letter C, which are sometimes called "the four C's of credit." These are *character, capacity, chance,* and *collateral.* The words are not exact nor complete descriptions of the bases of credit, but they are close enough to suggest the important facts, and they possess obvious mnemonic advantages. One's reputation is, of course, more important in determining his credit than is his true character. If a man wishes to borrow money or to obtain goods on credit, the prospective creditor is interested in his honesty and integrity—in his willingness and determination to repay the debt. But even the best of intentions are not enough. Unless the borrower has capacity—the ability to earn money with which to repay the loan when due—no one cares to extend credit to him. An honest half-wit is not a good credit risk.

Given character and capacity, we are still likely to inquire how the borrower wishes to use the money. Does he have a business opportunity that seems reasonably likely to be profitable? Commercial or production credit is extended, we recall, for self-liquidating purposes only. When a businessman asks his bank for credit, the bank is likely to require a balance sheet and an income statement, so that it may satisfy itself as to the opportunity

for putting the borrowed funds to profitable us. This is what is meant by the word "chance" as a basis for credit.

Finally, if a man wishes to borrow money and can put up $1000 worth of government bonds or a modern office building on a busy street corner as security, he will have little trouble in borrowing. To be sure, he may have to establish his ownership of the security or collateral but, if he can pledge valuable property as security behind the loan, the lender will feel safe in lending a fair percentage of the value of the security or collateral. Thus we might substitute for the four C's the words "reputation, ability, opportunity, and security," as being slightly more descriptive, even though less easily remembered.

19 · 14. Conversion of Personal to Bank Credit. An individual's credit, then, depends on his character, his capacity, his chance, and his collateral, or on the lender's appraisal of these things. If these things are all thought to be good, the borrower's credit will be large; but, if they are thought to be bad or unknown, his credit will be small or entirely lacking.

Business firms, like individuals, have credit that depends on these same characteristics. The reputation, skill, etc., of a bank are apt to be known to more people, over a wider area, than the same characteristics of a single individual. Therefore, the bank is likely to have more credit extending over a wider area, than would be true for the ordinary individual or business firm. Because the bank is in this way thought of as "better" than the ordinary individual, the individual finds it advantageous to exchange his own personal credit for that of the bank. He goes to the bank with credit of his own, depending on his own characteristics. He exchanges some or all of his own credit for bank credit and, since the bank credit is better, he accepts some discount. He may give up, say, $1500 of his own credit in return for $1487.50 of the bank's credit. His own credit is evidenced by a draft; the bank's credit is evidenced by a deposit subject to check. It may at first seem unfair for the bank to give but $1487.50 of bank credit in return for $1500 of personal credit, but there are two reasons for the difference. One is, as we have just explained, that the bank's credit is superior to the individual's. If two farmers swap potatoes, it might well be that the farmer with high-quality potatoes would give but 1487 bushels of his product for 1500 bushels of potatoes of a lower grade; and the bank gives $1487.50 of somewhat higher grade for $1500 of credit of a somewhat lower grade. We must also remember that the bank is promising to pay right now on demand, while the borrower is promising to pay only after the lapse of some period, such as 60 days. Our recollection of time preference[1] will help us to understand that the right to demand money now is more valuable than the right to demand money at

[1] See Sec. 13 • 2.

some time in the future. For this reason also, the bank discounts the note—gives less for it than its face value.

When people talk, as they often do loosely, about the "creation" of credit by banks, they are overlooking the fact that banks do not manufacture credit out of thin air. The businessman has credit of his own and would have it even if there were no banks. It is quite possible to carry on a credit business without banks, just as it is quite possible to grow and consume wheat without railroads and middlemen. But the bank makes the job easier. It assembles spare funds from individuals who do not need them for the time being, and lends them out to people who want them. This is the bank's major business and its main contribution to society. The manager of the electrical supply house who needs funds for 90 days is not likely to know that the dentist has, for the time being, spare funds available. But the dentist deposits them in the bank, which makes them available to the electrician. Thus the bank assembles the credit of a community, and exchanges its own credit for that of individual businessmen.

19 · 15. Banks as Investors. Throughout this chapter we have been emphasizing the role of the bank as lender—lending money and giving deposits. This is, to be sure, its typical and peculiar business. We have mentioned that banks often perform other functions as well, such as furnishing safe deposit boxes for patrons who want places of safekeeping for their valuables, or giving advice to customers on financial matters. But we have overlooked, or at least under-emphasized, the role of the bank as an investor. On Dec. 31, 1945, all the member banks of the Federal Reserve System had 22.8 billion dollars of loans and discounts, as compared with 84.4 billion of investments. If we consider all the commercial banks in the United States on the same date, we find that they had 26.2 billion dollars of loans and discounts as compared with 97.3 billion of investments. In both cases, the investments were very largely government securities, including large amounts of war bonds. Yet the situation was by no means a war phenomenon. The commercial banks of the country, which had always before invested primarily in commercial paper (in items listed under loans and discounts), reduced these loans sharply during the early thirties as a result of the depression. In 1934, the investments for the first time exceeded the loans and discounts, and this situation has continued unchanged ever since. As long as this condition continues, the banks are emphasizing functions which are not truly banking in character, but which are more characteristic of investment houses. Even with the large amount of funds invested in securities and even with the reserve ratios set at present at a high figure, the commercial banks have huge excess reserves available for credit expansion if and when such expansion seems warranted.

19 · 16. Functions of the Commercial Bank. Let us now summarize the functions and activities of commercial banks. (1) They perform the service of assembling the available credit of a community and distributing it among those who need it. In this respect they are like other middlemen who assemble commodities and distribute them. (2) Although hardly a real banking activity, most commercial banks do provide places of safekeeping where their customers may leave their valuables—the business of safety deposit. (3) The bank provides loanable funds for short periods to businessmen who need them for productive purposes. Although there are many other agencies providing other kinds of credit—long-term credit, consumption credit, etc. —the commercial bank is preeminent in the field of short-term commercial credit. (4) The commercial bank provides for the community a considerable part of the medium of exchange. This it does in two ways. We have put most of our stress in this chapter on the use of deposits subject to check. We have seen that, although these deposits are not money, they do to a considerable extent replace money. We have given less attention to another activity that has long been considered one of the major functions of a bank—the furnishing of actual money in the form of bank notes. This function of "issue," as it is called, is an old and well-recognized function of banks, and at one time or another almost every class of commercial bank in the United States has practiced it. The old state banks, as was pointed out in Chap. 18, gave us our only paper money for a long period in our history. The national-bank notes are only now disappearing. And the most important part of our circulating money—the Federal Reserve note—is issued by the Federal Reserve Banks. As we take up the development of the American banking system in Chap. 20, it will be important to keep in mind the fact that historically the function of issue has been quite as important as any other function; in the eyes of many people, it has been the most important and characteristic function of a bank.

Questions

1. What do we mean by self-liquidating credit? How does it liquidate itself?
2. When we call certain credit productive, do we mean that the credit is itself productive, or that the activities financed with it are productive? Is credit ever productive as the word was defined in Sect. 10 · 1?
3. Would you expect the interest rate on call loans to be higher or lower than that on time loans? Discuss.
4. What are some kinds of debts that are commonly amortized? Some kinds that are not?
5. What is a national bank?

6. What are the major differences between drafts and promissory notes? The major points of similarity?

7. Assume that John D. Buyer purchases $4000 worth of goods from William A. Seller. Ninety days' credit are given on the transaction, and it is financed by a draft. Draw up a facsimile of the draft, showing how it would look and what it would say. Mr. Buyer's bank is the City Trust Company of Toledo. Mr. Seller's bank is the Corn Exchange Bank of Buffalo.

8. Explain in detail the financing of the transaction described in the preceding question. Tell who would do what, what documents would be used, etc.

9. What requirements must a document meet to be a negotiable instrument?

10. Distinguish endorsement in blank from endorsement in full.

11. What is a bill of lading? What part does it play in the financing of commercial transactions?

12. What are acceptances? How do they arise?

13. What is the difference between interest and discount?

14. A man borrows money from a bank for one year, and his note is discounted at 6 per cent. What is the equivalent rate of interest? In other words, if the discount rate is 6 per cent, what rate of interest is he paying on his money?

15. A bank discounts a 30-day note for $5000 when the discount rate is 5 per cent. What changes will be made in the bank's balance sheet? Mark the changes plus or minus to show whether the items would be increased or decreased.

16. A bank makes $27 profit by discounting a 90-day note at 3 per cent. What was the amount of the note?

17. When a businessman borrows from a bank, why does he ordinarily take a deposit instead of the cash?

18. What changes will occur in the balance sheet of a bank if

    a. A depositor withdraws $100 in cash?

    b. The bank sells for $525 government bonds, which it has been listing on its balance sheet as worth $500?

    c. A man who borrowed from the bank 60 days ago on his 60-day note, now comes to the bank and renews the note instead of repaying? Assume that he renews for 60 days more, that the note is for $1000, and that the discount rate is 4 per cent.

    d. A man deposits a $200 check drawn on another bank?

    e. A merchant borrows $1400 for 30 days on his endorsed promissory note? The discount rate is 4 per cent.

f. The bank rediscounts notes to the amount of $10,000. The notes now have 60 days to run, and the rediscount rate is 2 per cent.

19. What are overdrafts? How do they arise? Why does the bank list them as assets?

20. At the end of a day, bank A finds that it holds checks for $20,000 against bank B and for $25,000 against bank C. Bank B holds checks of $25,000 against A and of $30,000 against C. Bank C holds checks of $20,000 against A and of $35,000 against B. Assume these three are the only banks in the city.

    a. If each pair of banks settles separately, bank A settling with bank B and separately with bank C, and bank B settling with bank C, how much cash will change hands to settle the balances?

    b. If the three banks settle through a clearing house, how much cash will change hands to settle balances?

    c. If each check were settled separately, how much cash would change hands?

    d. Explain just how it is that the clearings process reduces the use of cash.

21. A country bank has $153,000 of demand deposits and $97,000 of time deposits. What is its minimum legal reserve with the Federal Reserve Bank under present conditions?

22. Explain the use of call loans in controlling the reserve ratio.

23. Explain the effect of endorsing a piece of negotiable paper.

24. Why is two-name paper ordinarily a better credit risk than one-name paper?

25. Would every piece of two-name paper be a better risk than any piece of one-name paper? Explain.

26. Which of the four C's of credit do you think is the least important? That is, if a man were to have three of them without the fourth, which one could be dispensed with most easily in your opinion? Discuss.

20. THE BANKING SYSTEM OF THE UNITED STATES

20 • 1. Historical Background. There were no real commercial banks in the early American colonies. Capital was badly needed on every hand and was furnished in part by the savings of the settlers themselves and in part by borrowing from abroad. Anyone in the colonies who had an excess of capital to lend found it easy to discover borrowers, and no intermediaries were necessary. Such borrowing was usually secured by personal promissory notes, which did to some small extent pass from hand to hand by endorsement as mediums of exchange. The financial exigencies of the Revolutionary War were such as to demonstrate the need for some institution to mobilize the credit resources of the country and, after several attempts in other directions, the Bank of North America was finally established in 1781 under charter from Congress. This, the first modern bank in the country, was organized by Robert Morris with the aid of a substantial subscription of government funds; yet it was privately owned and operated. The bank was of notable aid to the government during the remainder of the war, after which its relationship with the government ceased and it continued as a private bank, later getting a state charter from Pennsylvania.

In 1790, Alexander Hamilton recommended that the government set up a United States Bank, arguing that it would increase the availability of capital through the issue of notes, that it would enable the government to obtain loans more easily, and that it would make it easier for citizens to pay their taxes by increasing the volume of currency. Congress debated the proposal at length and with much feeling, and serious doubt was thrown on the constitutionality of the measure. The bill finally passed with the representatives of the Southern states almost unanimously in opposition and the representatives of the Northern states almost solidly in favor. Only after considerable hesitance and investigation did President Washington sign the bill. The United States Bank was given a 20-year charter, from 1791 to 1811. The act provided for 10 million dollars of capital stock, 2 million of which was bought by the government and the rest of private parties. The bank turned

out to be of very great advantage to the government and also of great assistance to private industry, although in the latter field the rapid growth of state banks was making the need less and less pressing. Its note issues circulated freely, and were redeemed promptly on demand. But the bank was unpopular with Congress, partly because many members feared the development of the "money power"—of monopolistic elements in the financial field—and partly because of criticism from the rapidly growing state banks which disliked government competition. Also a majority of the stock of the bank had passed into the hands of foreigners, and opponents stressed the likelihood of foreign domination, either not understanding or neglecting to state that foreign stockholders could not vote their stock. The constitutionality of the bank was also again questioned. The major criticisms were what might be characterized as political rather than economic, although one should not get the idea that politics and economics are mutually exclusive or antagonistic fields. At any rate, the bill to renew the charter came to a tie vote in the Senate and was lost on the negative vote of the vice-president, George Clinton.

For 5 years the government turned to local banks for borrowing and deposit of funds, this period including the War of 1812 with its troublesome financial problems. The government experienced considerable difficulty in financing this war, and many people became convinced that the failure to renew the charter of the United States Bank had been an error. Despite continued opposition from those who felt that such activities were unconstitutional, President Madison and his Secretary of the Treasury, Alexander Dallas, threw strong support behind a move to establish a second United States Bank. For 2 years there were proposals and counterproposals in Congress, but finally a bill was passed and the new bank established in 1816. Like its predecessor, its charter ran for 20 years. In an effort to force the state banks to resume specie payments on their issues of irredeemable money, the new bank was authorized to issue bank notes redeemable in specie, and in 1817 Congress passed a measure providing that all dues to the government must be paid either in the notes of this bank, or in cash, or in the notes of banks which were payable and paid on demand in United States currency. The capital of the new bank was 35 million dollars, of which 7 million were subscribed by the government. One-fifth of the directors were appointed by the President. Modern students of the subject seem to agree that the bank was badly managed during its first few years and, after the Baltimore branch failed in 1818, it was reorganized in 1819. Thereafter the bank was conservatively managed, and of considerable influence in the South and West, although relatively inactive in New England, where state banks were strong. It was successful in establishing a sound and nation-wide circulating medium through the issue of its own bank notes, and indirectly it was of benefit in improving the circulating currency of the state banks, which were forced

by competition to become somewhat more conservative in their own practices. The bank also was of decided financial assistance to the government and it was the major influence in the field of foreign exchange. The nation's specie reserve soon came almost entirely into its hands, and the bank was of value in many other ways. But there was strong opposition to it, as there had been to its predecessor. The issue of states' rights was being discussed heatedly in other connections, and there were still many people who were not satisfied with the Supreme Court's repeated declaration that the bank was constitutional. The state banks, growing rapidly in number and influence, were allied in opposition. The bank's insistence on a sound currency, which was perhaps one of its most valuable contributions, made it many enemies. President Andrew Jackson, newly elected in a spirit of opposition to monopoly, raised questions as to both the constitutionality and the desirability of the bank in his first inaugural address. Despite strong opposition, Congress did pass an act in 1832 renewing the charter with some modifications, but President Jackson vetoed the measure, and the bank died when its charter expired in 1836.

20·2. Period of State Banks. It was now clear that the reestablishment of a government central bank was politically impossible, at least for the time being. For a few years the government turned to state banks as depositories of public funds; but these banks, released from the influence of the United States Bank, began an orgy of speculation and in many cases threw off all restraint until it became evident that it was not safe to entrust government funds to their care. In 1840, Congress passed a measure, repealed almost at once but reenacted in 1846, establishing the Independent Treasury. The funds of the government were withdrawn from circulation and locked up in the Treasury for safekeeping. The government required payment of public dues in specie and was criticized by some people for withdrawing specie from circulation and hoarding it. In general, the period was one of prosperity, and even when business was bad, as during the panic of 1857, the government had no difficulty in weathering the storm under the Independent Treasury system. During this period, the government in effect divorced itself from banking and from the private banking institutions. The public was left to rely on the state banks.

The number of state banks continued to grow rapidly, doubling every 15 years or so. These banks were subject to state laws, which differed greatly from state to state. Some of the institutions were well managed, while others were, as one historian has put it, "mere batches of paper money." Many men started banks, not because they wanted to do a banking business, but because they wanted credit for their own enterprises and could not get it from existing banks. Especially in the Western states, legal restrictions were few. Almost every state bank issued bank notes, some of which were actually

redeemed in specie, but many of which circulated at a decided discount. One bank note might be worth 100 cents on the dollar, another worth 60 cents, and a third be entirely worthless. Banks were established far in the wilderness on the frontier—the so-called "wildcat banks"—to make it difficult for holders of the bank's notes to present them for redemption. Banks were considered speculative institutions by their managers, and charitable institutions by their borrowers. The ordinary citizen had to keep track as best he could of thousands of kinds of bank notes which differed greatly in value. One cannot properly speak of the system of state banks, since its outstanding characteristic was a lack of system. Many of the wildcat banks did not engage in the business of discount and loan at all, but made their profits through the issue of bank notes. The laws of the Northeastern states were, in general, somewhat more stringent than those in other parts of the country, and as a result some real strong commercial banks did spring up there. But, in general, chaos prevailed in the banking world from the closing of the second United States Bank in 1836 until the establishment of the National Banking System during the Civil War.

20 · 3. National Banks. The Civil War raised many new and difficult problems of finance. In December, 1861, the banks throughout the country suspended specie payments. In 1862, the government started the issuance of "greenbacks"—the United States notes with which we became familiar in Chap. 19. These notes being, like those of the state banks, inconvertible, soon depreciated in value. The problem of financing the war became more and more difficult. The Secretary of the Treasury, Salmon P. Chase, had consented to the issue of the greenbacks only with the greatest reluctance, and felt that something must be done to strengthen the credit of the government and to assist with the financing of the war. In 1861, he recommended the establishment of a national banking system, and in 1862 he elaborated the plan further. It was argued that there should be banks whose charters rested on national rather than on state authority, and that a system of national banks, uniform from state to state, would free the country from the evils of state bank notes. On Jan. 1, 1862, there were 1496 banks issuing circulating notes, chartered and controlled by 29 different states.[1] In 1863 and 1864, Congress passed laws which established a real *national banking system,* in part to improve the banking system, and in part to finance the war. It is unfortunate that the purposes had to be mixed, since certain provisions were included solely for war purposes which continued to plague the system for the following half century.

The National Banking Act provided for commercial banks chartered by the Federal government and uniform throughout the nation. A year or two

[1] DEWEY, DAVIS R., "Financial History of the United States," p. 321, Longmans, Green and Company, New York, 1902.

later a tax of 10 per cent per annum was laid on outstanding notes of state banks. Since this tax was prohibitive (the tax was larger than the profit that could be earned on the notes), these notes which had caused so much difficulty quickly disappeared. In their place, the national banks were authorized to issue national-bank notes. In an effort to finance the war, the act provided that every national bank must subscribe to government bonds, the amount depending on its capitalization. Banks whose capital exceeded $150,000 must buy and deposit with the Treasurer of the United States bonds to the amount of $50,000; and banks with smaller capitalization must deposit bonds to the amount of one-fourth their capital. The banks might then issue national-bank notes, using the bonds as security. The amount of national-bank notes might not exceed the face value of the bonds nor the capital of the bank. The government was required to accept these notes for all dues except import duties, and they were payable by the government for all purposes except to pay off its own indebtedness or the interest on its bonds. Aside from this, the national-bank notes were without any legal-tender power whatever save in transactions between national banks themselves. The Independent Treasury Act was modified to allow the deposit of public monies in selected private banks under certain restrictions.

The national-bank notes were redeemable at the bank that issued them, or by the U. S. Treasury. For the latter purpose, the bank was required to keep a deposit with the Treasury of 5 per cent of the outstanding bank notes. Being backed 100 per cent in government bonds and 5 per cent in cash, the notes were safe as long as the government was safe; but, since the amount of bonds available for backing was fixed at any time, the issue of national-bank notes was relatively inelastic. Also, since the bonds that could legally be used as backing for bank notes usually sold at a premium, and since there was a small tax on the issuance of the notes themselves, the profit that a bank could realize from issuing bank notes was very small—usually less than 1 per cent. Despite this fact, many banks felt that there was an advertising value in issuing bank notes, and it was only when all the bonds with the circulation privilege were called for redemption in 1935 that these notes began to disappear.

The national banking system was very successful in improving the situation with respect to circulating bank notes, but in other regards it left much still to be desired in the banking system. As has already been noted, the currency was still inelastic. The banking system, moreover, was still a system of many small banks. In most states banks were forbidden by law to establish branches, and the individualism that characterized all parts of the American economy was most marked in the field of banking. As the advantages of large-scale operation came to be better understood, the tendency in most countries was for banks to amalgamate into a few large systems, some-

what on the order of the chain store in the American grocery field. In England, most of the commercial banking has fallen into the hands of five great banks, known as "the Big Five"[1] with over 8000 branches. Canada has some 10 banks, but most of the business is in the hands of four banks[2] with some 2000 branches. The financial situation in Germany between the two World Wars was unsettled and rapidly changing, but most of the commercial banking was done by the branches of some five or six large banks. Yet the National Banking Act prohibited, with a minor exception, the establishment of branch banks, and most of the state banks were under similar legal restraint. The United States was, and still is, a nation of small banks, in which the fear of the money power has led legislative bodies to restrict the amalgamation of banking institutions and to insist on decentralization of banking.

There are a number of effects of this decentralization, but under the National Banking System one of the effects was a decentralization of bank reserves. The national banks were required to keep reserves behind their deposits, but only in the largest cities[3] were they required to keep their reserves in their own vaults. In cities of medium size, national banks were required to maintain 25 per cent reserves, but half of the reserve could be deposited in other banks. In the smaller cities, only 15 per cent reserves were required, and up to 60 per cent of the reserve might be deposited in another bank. Thus, we can imagine a bank in a small town with $320,000 of deposits. Behind these deposits it keeps reserves of $48,000. Forty per cent, or $19,200, it keeps in its own vaults, but the remaining $28,800 it deposits in a bank in some city of medium size. The bank that receives the $28,800 deposit does not have to keep it all. It is required to keep a reserve of but 25 per cent, or $7,200, and only half of this need be kept in its own vaults. The other $3,600 it sends on, perhaps, for deposit in a bank in New York. The New York bank keeps a 25 per cent reserve, or $900. Now if we try to trace the reserves of the original small-town bank we find the following:

Amount of deposits	$320,000
Cash in own vault .	19,200
Cash in second bank	3,600
Cash in third bank	900

The actual cash held behind the deposits amounts to $23,700, or 7.4 per cent, scattered through three different banks. In practice this pyramiding of

[1] Barclay's Bank, Ltd.; Lloyd's Bank, Ltd.; Midland Bank, Ltd.; National Provincial Bank, Ltd.; and Westminster Bank, Ltd.

[2] Royal Bank of Canada; Bank of Nova Scotia; Bank of Montreal; and Canadian Bank of Commerce.

[3] New York, Chicago, and St. Louis.

bank reserves often went much further, with the reserves scattered through many banks.

This decentralization of bank reserves was one of the serious weaknesses of the National Banking System. If a bank suddenly needed cash, it had to call on other banks. It took time to get the cash back when it was needed, and the other banks were embarrassed by withdrawals of funds at times when they, too, were apt to need cash. A considerable proportion of the bank reserves came to be deposited in the banks of the large cities, especially in New York banks, where they were lent out on call, largely for purposes of stock market speculation. In a time of money stringency, each bank tried to improve its position by cashing in its deposits with other banks and building up its local reserve; and in this way the other banks were put into an even more serious situation by withdrawals of cash which they needed.

In addition to the two weaknesses just described—the inelasticity of the currency and the decentralization of reserves—the National Banking System suffered from the lack of any ready and organized market for the rediscount of commercial paper. Banks needing more funds for any purpose could, of course, "sell" commercial paper if they could find a buyer. They could try to sell it to their own customers or to other banks. But at a time when one bank was in special need of funds, it was likely that other banks would be meeting the same problem, and rediscount was difficult and fortuitous.

These difficulties were recognized for some time, and were called vividly to the attention of interested parties by the Panic of 1907. Although this panic was short-lived, it was extremely serious while it lasted, and both businessmen and bankers insisted that our banking system was out of date and in urgent need of overhauling. As a result, Congress created a Monetary Commission, usually called the "Aldrich Commission" in honor of its chairman. This commission made a thorough study of the situation and brought in recommendations for a major revision of the banking system. Just at this time, there was a change of political administrations, as President Wilson and the Democrats came into office in 1913; and some people feared that the recommendations would be discarded on political grounds. On the contrary, the incoming administration took over the results of the careful study that had been made, added some improvements, and adopted the Federal Reserve Act as one of its major measures.

20 · 4. Federal Reserve System. The Federal Reserve Act was passed and signed by President Wilson on Dec. 23, 1913. It was entitled, "An Act to provide for the establishment of Federal Reserve Banks, to furnish an elastic currency, to afford means of rediscounting commercial paper, to establish a more effective supervision of banking in the United States, and for other purposes." This act has been amended over and over again, especially by the Banking Act of 1935, which made major revisions. It is not important for us in an elementary treatment to trace all the changes that have taken

place, but we shall try to describe the general outlines of the system as it operates today.

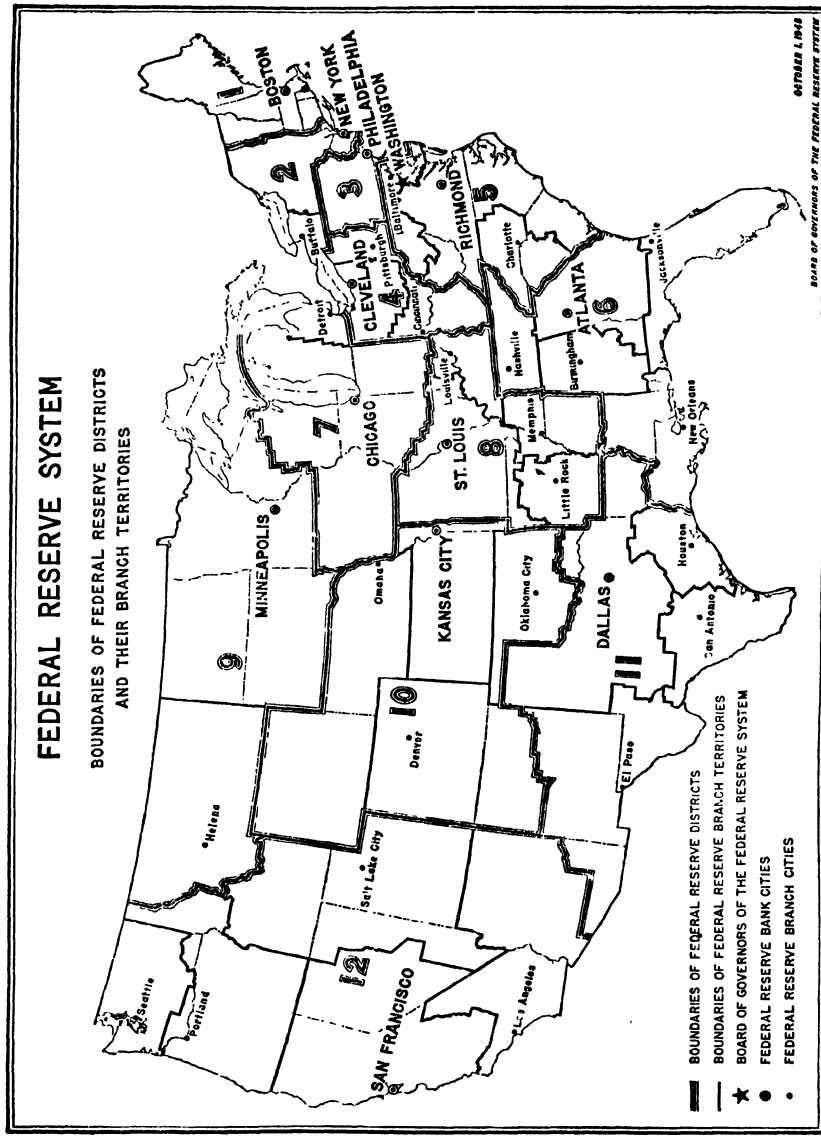


FIG. 20 • 1.—Federal Reserve System.

Under the Federal Reserve System, the territory of the United States is divided into 12 Federal Reserve Districts, each of which contains a Federal Reserve Bank. Thus, instead of having one single central bank as has been

common in other countries, the American fear of monopoly and centralization led to the establishment of a dozen separate and largely independent institutions. The present districts and locations of Federal Reserve Banks and their branches are shown on the map of Fig. 20 · 1. It will be noted at once that the districts differ greatly in size. The Federal Reserve cities and the boundaries of the districts were chosen as far as possible on the basis of the interbank relationships and the flow of funds that had already grown up. The present (May, 1947) locations of Federal Reserve Banks and of their branches (the latter being shown by indentation under the parent bank) are as follows:

1. Boston
2. New York
   - Buffalo
3. Philadelphia
4. Cleveland
   - Cincinnati
   - Pittsburgh
5. Richmond
   - Baltimore
   - Charlotte
6. Atlanta
   - Birmingham
   - Jacksonville
   - New Orleans
   - Nashville
7. Chicago
   - Detroit
8. St. Louis
   - Little Rock
   - Louisville
   - Memphis
9. Minneapolis
   - Helena
10. Kansas City
    - Denver
    - Oklahoma City
    - Omaha
11. Dallas
    - El Paso
    - Houston
    - San Antonio
12. San Francisco
    - Los Angeles
    - Portland
    - Salt Lake City
    - Seattle

The 12 Federal Reserve Banks, like other corporations, belong to their stockholders, which in this case, are commercial banks, called "member banks." Each member bank is required to buy stock in its Federal Reserve Bank in an amount equal to 6 per cent of the member bank's capital and surplus.[1] Every national bank is required to be a member, and other banks may become members if they wish although, in case they do join the system, they must submit to the regulations laid down for member banks. On Dec. 31, 1945, there were 14,011 commercial banks in the United States. Of these, 5,017 were national banks, all of which were compelled to be members. Of the 8,994 state commercial banks, 1,867 were and 7,127 were not members (see Fig. 20 · 2.) This hardly gives a fair picture of the situation, however, since on the average the member banks are much larger than the nonmember

[1] Only half this amount need be paid in, the rest being subject to call by the Board of Governors. To date, the remaining half has never been called for.

banks. Although just over half the banks are nonmembers, these nonmembers hold less than 15 per cent of the total deposits. If we can judge the situation

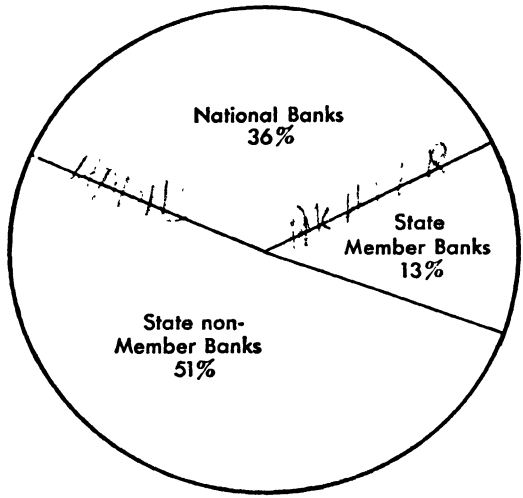


FIG. 20 • 2.—Relationship of commercial banks to the Federal Reserve System, Dec. 31, 1945.

somewhat more accurately on the basis of deposits, Fig. 20 • 3 will show the relative importance of member and nonmember banks.[1]

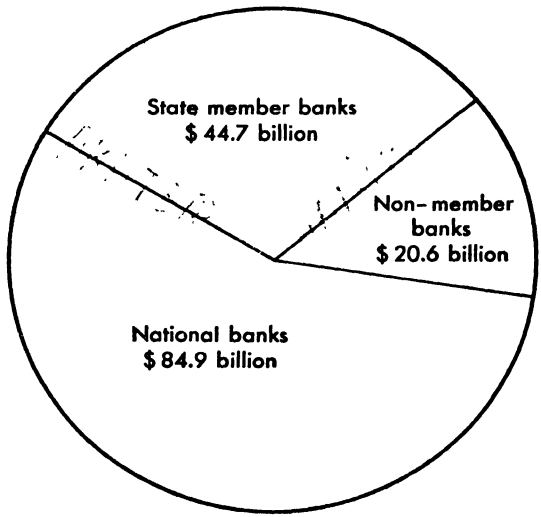


FIG. 20 • 3.—Relative importance of member and nonmember banks, measured by deposits, Dec. 31, 1945.

[1] It will be noted that the American banking system grows by accretion rather than by substitution. The adoption of the National Banking System did not do away with state banks, nor did the Federal Reserve System do away with national banks. Each added something to what went before, and today we have them all operating side by side.

All member banks, whether state or national, are required to maintain the reserve ratios described in Sec. 19 • 10. The amount of these minimum reserves depends on the city in which the member bank is located. New York and Chicago are today the only *central reserve cities,* and the 49 member banks in these two cities are required to maintain reserves of somewhere between 13 and 26 per cent. The following cities are today[1] *reserve cities,* and the 360 member banks in these 58 cities are required to keep reserves of somewhere between 10 and 20 per cent.

Atlanta	El Paso	Milwaukee	St. Paul
Baltimore	Fort Worth	Minneapolis	Salt Lake City
Birmingham	Galveston	Nashville	San Antonio
Boston	Grand Rapids	New Orleans	San Francisco
Buffalo	Helena	Ogden	Seattle
Cedar Rapids	Houston	Oklahoma City	Sioux City
Charlotte	Indianapolis	Omaha	Spokane
Cincinnati	Jacksonville	Philadelphia	Toledo
Cleveland	Kansas City, Kans.	Peoria	Topeka
Columbus, Ohio	Kansas City, Mo.	Pittsburgh	Tulsa
Dallas	Lincoln	Portland, Ore.	Waco
Denver	Little Rock	Pueblo	Washington
Detroit	Los Angeles	Richmond	Wichita
Des Moines	Louisville	St. Joseph	
Dubuque	Memphis	St. Louis	

Banks outside the 60 cities listed in these first two groups are called "country banks." On June 30, 1945, there were 6431 country banks. They are required to maintain reserves of between 7 and 14 per cent. The relative importance of the banks in the different groups cannot, of course, be determined from their number. Although less than 1 per cent of the member banks are in central reserve cities, these member banks hold approximately 35 per cent of all the demand deposits. The banks in reserve cities make up some 5 per cent of the member banks of the country but hold approximately 37 per cent of the deposits. The country banks, accounting for 94 per cent of the number, hold but 28 per cent of the deposits.

State banks wishing to become members of the Federal Reserve System must maintain the same reserves as those required of other member banks, and, in addition, they must submit to examination by Federal authorities.

Each of the 12 Federal Reserve Banks is a separate corporation, with its own board of 9 directors. Six of these directors are elected by the member

[1] The data in this paragraph listing cities in different classes and the number of member banks in each group are correct as of May, 1947. Only two changes in the designation of cities have taken place since 1930, one in 1940 and one in 1945. Thus the student will see that the cities in the various classes are fairly well fixed.

banks in the district and the other three (including the chairman of the board, who is known as the Federal Reserve Agent) are designated by the Board of Governors of the Federal Reserve System. These nine directors select the officers of the bank. Of the six directors selected by member banks, three are bankers representing the member banks, and the other three must be persons actively engaged within the district in commerce, agriculture, or some other industrial pursuit.

20 · 5. Board of Governors. From the foregoing description we see that the 12 Federal Reserve Banks are separate, privately owned institutions. Yet they are subject to certain external coordinating controls, the strength of which was greatly increased in 1935 as part of the new deal program of centralization and authoritarianism. Most important is the body known as the Board of Governors of the Federal Reserve System. This is a board of seven men serving overlapping 14-year terms at annual salaries of $15,000. They are appointed by the President of the United States, by and with the advice and consent of the Senate. The purpose of having overlapping terms, with one member appointed every other year, is to arrange it so that no administration can take immediate political control of the board. It would take 8 years in the ordinary course of events for any President to appoint a majority of the board, and the slow rotation of members is intended to divorce the board as far as possible from political interference. That such methods are actually ineffective is well known to any practicing politician, as demonstrated when the administration changed the name of the board in 1935 and on Feb. 1, 1936, displaced the existing board with a new and entirely inexperienced one with one exception. The board is empowered to make examinations of the Federal Reserve Banks, to require statements and reports from them, to require them to establish or discontinue branches, and to supervise the issue or retirement of Federal Reserve notes. As we have said, it selects one-third of the directors of each Federal Reserve Bank. In connection with member banks, the board has the power of examination, of passing on admission of state banks to membership in the system, or regulating monopolistic practices under the Clayton Act,[1] of limiting the amount of interest paid by member banks on savings accounts, etc. The board also either determines or influences the rediscount rates set by the Federal Reserve Banks, the reserve requirements of member banks, and the carrying-on of open-market operations, all of which are so important as to require separate mention.

20 · 6. Functions of the Federal Reserve System. The Federal Reserve System has come to be a very complicated organization, with many duties and responsibilities. Among the most important of its functions are

1. To act as fiscal agent of the government.

[1] See Sec. 25 · 16.

2. To issue bank notes and provide an elastic currency.
3. To rediscount paper for its members.
4. To reduce the evils of the business cycle.
5. To hold the reserves of member banks.
6. To act as a clearing system.

Some of these functions require little in the way of explanation or comment. Since the institution of the Federal Reserve System, the government no longer operates on the basis of the independent treasury, but uses the Federal Reserve Banks just as a businessman might use his local commercial bank. The 12 Federal Reserve Banks do their business either with the Federal government or with the member banks. They have sometimes been called "bankers' banks" because they act for bankers much as bankers act for businessmen. We have already seen that member banks can count as reserve only funds on deposit with their Federal Reserve Bank, so that the new system succeeded at the outset in centralizing the reserves of its members and thus in escaping some of the difficulties described in Sec. 20 • 3.

Just as member banks are required to keep reserves behind their deposits, so the Federal Reserve Banks must keep reserves behind theirs. Since June 12, 1945, they have been required to maintain a reserve of 25 per cent in gold certificates behind deposits, just as behind the Federal Reserve notes. Thus, if a Federal Reserve Bank has $7000 in gold certificates, it can build up deposits to the amount of $28,000. These deposits are used by the member banks as their reserves, and these $28,000 of reserves might then be used by member banks to support, in the extreme case, $400,000 of deposits.[1] Thus the deposits in the member banks might be backed by but 1¾ per cent in cash.

We have already spoken of the Federal Reserve notes, and their outstanding importance. We have also mentioned their elasticity, which depends on their backing of commercial paper. We now turn to some of the functions of the Federal Reserve System to which we have not given attention before.

20 • 7. Federal Reserve Clearings System. In Sec. 19 • 9, we learned how banks reduce the need for money by "canceling" the larger part of the checks that are issued, settling only small balances in cash. We described how this occurs within a given bank or among banks within a city. Let us now turn to the method of clearing used when the banks are in different cities.

Before the establishment of the Federal Reserve System, checks on out-of-town banks were handled through a system of *correspondent banks*. Each bank entered into reciprocal arrangements with a number of other banks, agreeing to handle local business for the other bank in return for their

[1] The reserve requirements for country banks may be set as low as 7 per cent, and it is this figure on which the computation above is based. The student can easily compute corresponding figures for any other member bank reserve requirements.

promise to do likewise. Thus a bank in Rochester, N. Y., would settle transactions with other Rochester banks through the local clearing house or by messenger. But it would also establish correspondent relationships with banks in all the surrounding towns, and in the more important larger cities of the state and surrounding states. If a Rochester man bought goods in St. Louis and sent his check on a Rochester bank in payment, the St. Louis merchant would deposit the check in his local bank. The St. Louis bank would not be likely to have a correspondent in Rochester, so it might send the check to a correspondent in Chicago. The Chicago bank might then send it to a correspondent in Cleveland, which would send it to a correspondent in Utica, which would send it back to a correspondent in Rochester. The bank in Rochester would now send it by messenger or through the clearing house to the bank on which it was originally drawn. Thus each check that was sent to a distant town followed a devious and uncertain route back to the bank on which it was drawn. This correspondent system took a good deal of time and expense and kept large amounts of funds tied up in checks that were in the process of collection.

Checks may still be cleared through correspondent banks, but with the establishment of the Federal Reserve System a new and faster system was set up. This works in two ways. In the first place, each Federal Reserve Bank acts as a clearing house for the member banks of its own district.[1] Each bank gathers together its checks on other banks and sorts them out at the end of the day. We have already seen that the checks on other banks in the same town are sent by messenger or through the local clearing house back to the bank on which they were drawn, most of the indebtedness being canceled in the process. But the out-of-town checks are sent to the Federal Reserve Bank of the district,[2] where the Federal Reserve Bank goes through a process of clearings just like that which the local clearing house carries out. It is necessary for the Federal Reserve Bank to compute the amount owed to each bank in the district (this is, of course, the sum of the checks sent in) and the amount owed by each bank in the district. Most of the indebtedness is then canceled off, and the bank is given credit or is debited for the uncanceled difference. Since each bank is required to carry a deposit with its Federal Reserve Bank, it is not necessary to send cash even to settle the differences. If our local bank has a small credit balance today, the amount is merely added to our deposit at the Federal Reserve Bank or, if we

[1] Nonmember banks are also allowed to avail themselves of the advantages of clearing through the Federal Reserve Banks. Such *nonmember clearing banks* are required to carry an account with the Federal Reserve Bank for the purposes of clearing and have relationships with the Federal Reserve Bank for purposes of clearing and collection of checks only.

[2] Federal Reserve Banks also give "direct sending privileges" to member or nonmember banks in other Federal Reserve districts if they have a sufficient daily volume of checks on the Federal Reserve Bank's district to warrant it.

have a small deficit, it is deducted from our deposit. Of course, if the balances stay against us day after day, our deposit will slowly be drawn down until finally we are forced to send funds to increase our deposit; if the balances continue for some time in our favor, we shall build up a large balance at the bank and decide to bring some of it back for use. Only in these cases are the checks occasionally settled by transfer of cash.

If the check is sent outside the Federal Reserve district, the process is slightly more complicated. Each Federal Reserve Bank is receiving each day checks from member banks all through its district. Some call for payment of funds by banks in other districts. For the purposes of settling these checks, the 12 Federal Reserve Banks maintain in Washington the *interdistrict settlement fund.* Each day each Federal Reserve Bank telegraphs to the Board of Governors in Washington, reporting the amount due it from each of the other Federal Reserve districts. The Treasury Department, which maintains the interdistrict settlement fund, then computes the total amount owed by and the total amount owed to each district, and credits or debits the difference in the settlement fund. Thus, it might turn out on a given day that the Federal Reserve Bank in Atlanta had 50 million dollars of checks drawn on other districts. It would report the amount owed to it by each of the other 11 districts. But at the same time, each of the other 11 districts would be reporting. The Treasury officials would note that the other districts owed the sixth or Atlanta district 50 million dollars but, when they looked over the reports of the other districts, they might find that the Atlanta district owed other districts 48 million dollars. In such a case, the Atlanta bank's account in the settlement fund would be increased by 2 million dollars. No actual cash would change hands. Some Federal Reserve Banks would find their equities in the settlement fund increased, and some would find them decreased. If trade continued for some reason to flow for a considerable time in a given direction, so that the Atlanta bank continued to build up a larger and larger credit in the settlement fund, it might finally withdraw funds. Or a bank might find that its equity in the settlement fund was growing so small that it was desirable to send funds to Washington to augment it. Only in these unusual cases is actual cash used. All the other checks are "canceled out" without any actual cash changing hands.

This description enables us to tell by what path a check will travel if it is sent out of town. If a man in Hanover, N. H., sends his check to a life-insurance company in Hartford, Conn., the life-insurance company will endorse the check and deposit it in a Hartford bank. The Hartford bank and the Hanover bank are in the same district—the first or Boston district. The Hartford bank endorses the check and sends it to the Federal Reserve Bank in Boston and receives credit for it there. The Federal Reserve Bank in Boston debits it to the account of the Hanover bank, endorses it, and sends it

to Hanover, where the check is canceled and returned to the man who drew it with his bank statement at the end of the month. If, however, the man in Hanover had ordered a bushel of oranges from Florida and had sent his check to the grower in Eustis, Fla., the bank in Eustis would be in the sixth or Atlanta district while the Hanover bank on which the check was drawn would be in the first or Boston district. The orange grower would endorse the check and deposit it in his Eustis bank. The Eustis bank would endorse it and send it to the Federal Reserve Bank in Atlanta.[1] The Federal Reserve Bank in Atlanta would give credit to the Eustis bank and would notify the settlement fund in Washington of the total amount due from the Boston and from the other districts. It would send the check, after endorsing it, to the Federal Reserve Bank in Boston. This latter bank would endorse the check, debit it to the account of the Hanover bank, and send it on to the Hanover bank, where it would finally be paid. This sounds much more complicated than it really is. Instead of going through a long chain of correspondent banks, the check goes from the man who drew it to the payee, to the payee's bank, to the Federal Reserve Bank of the payee's district, to the Federal Reserve Bank of the drawer's district, and back to the drawer's bank. And by the procedure that we have described it becomes possible to cancel most of the checks, settling them without any transfer of funds at all. Thus, the system of deposit and loan, in conjunction with the system of clearings, makes it possible to use the existing monetary stock more efficiently, doing much more business with a given amount of money than would be possible if transactions typically had to be settled by an exchange of cash. To illustrate, during the boom year of 1929, the total money in circulation in the United States was 4.7 billion dollars. The total deposits in commercial banks amounted to 57.9 billion dollars, so that we see that there was far more "money" on deposit in the banks than there was in the entire monetary stock. The bank clearings in 37 cities of the United States for 1929 amounted to 686.2 billion dollars.[2] In 1944 the Federal Reserve Banks cleared some 1.7 billion checks. These figures serve to make again the point that we have emphasized over and over again; namely, that deposits are not money in the bank and that payments by check are not payments of money. In fact, the clearings systems that we have described are set up largely to make it possible

[1] As has been said before, the Eustis bank might be given a "direct sending privilege" to send checks directly to the Federal Reserve Bank in Boston if it had a daily volume of checks on the first district that would warrant it. Such a situation would be unusual, however, and the ordinary procedure would be that described in the text. If the bank had the direct sending privilege, we would merely cut out one step.

[2] These figures are not strictly comparable. The figure for money in circulation is for June 30, 1929; the figure for deposits is described as "of June 30 or about June 30," 1929. The clearings figure is for the entire year. Despite the fact that the figures are not entirely comparable, they do serve to illustrate the point.

to dispense with money in checking transactions. Most checking transactions involve the use of credit rather than cash.

20 • 8. Federal Reserve Banks and the Business Cycle. We have not yet described the business cycle,[1] but we shall take it for granted that the reader is familiar with the recurring periods of "prosperity" and "depression," which have followed one another throughout the time of recorded history. The authors of the Federal Reserve Act felt that these periods, which make up the *business cycle,* were in some measure the result of financial and monetary causes, and that it should be possible to control them to some extent and to reduce their fluctuations, even if not to eradicate them entirely, by judicious use of the banking system. For that reason, certain powers were given to the authorities of the Federal Reserve System[2] with the hope that they might be used to such an end. The three major approaches to the problem are through control of the rediscount rate, through control of the reserve requirements, and through open-market operations.

20 • 9. Rediscount Rate. We learned in Sec. 19 • 11 how and why member banks rediscount commercial paper. If their reserves are running low and they wish to continue extending credit, they "sell" some of their commercial paper to their Federal Reserve Bank through the process of rediscount, thus getting funds which they can use as reserve behind additional loans. If the Board of Governors feels that a period of speculative expansion is under way and is likely to lead to future difficulties, it may wish to curb the extension of credit. Under such circumstances the board might order the Federal Reserve Banks to raise their rediscount rates. We can see at once that this will reduce the profitability of rediscount for the member bank.[3] As the rediscount rate gets higher and higher as compared with the member bank's discount rate, the profitability of rediscounting gets smaller and smaller. If the rediscount rate rises to a point where it is equal to the discount rate, there is no profit at all in rediscount; and if the rediscount rate is larger than the discount rate, there is an actual loss in rediscount. On the other hand, if the Board of Governors feels that a period of financial slump is under way and that businessmen are contracting their output and employment through fear, they may well decide that steps should be taken to make borrowing easier and to encourage the giving of credit to businessmen. Under such circumstances they would be likely to order the Federal Reserve Banks to lower their rediscount rates, making lending more profitable.

[1] See Chap. 22.

[2] Most of these powers date, at least in their present form, from amendments to the Federal Reserve Act, which were part of the Banking Act of 1935.

[3] The student might well go back at this point to the detailed description of rediscount in Sec. 19 • 11, where the effect of rediscount on the member bank's balance sheet is shown with examples.

Member banks will then be encouraged to lend on commercial paper until they reach the limits of their reserves, rediscount to augment their reserves, and then start lending again. The theory is perhaps easiest to understand in terms of a simple example. Suppose that a bank's discount rate is 5 per cent and the Federal Reserve Bank's rediscount rate is 1 per cent. Let us suppose that the member bank discounts a 90-day note for $4000. It makes a profit of $50 on the transaction.[1] If it now rediscounts the note immediately at the Federal Reserve Bank, it gets $3990 of its reserve, and gives up $10 of its profit. It has made a net profit of $40 on the two transactions together, and has lost but $10 of its reserves. Under such circumstances one might expect the member bank to lend the money over and over and over again—which is what the Board of Governors was trying to encourage when it set the low rediscount rate.

The student is likely to think that there is no limit to possible lending by a member bank under such circumstances. As a matter of fact, raising the rediscount rate may act as a rather efficient curb on overexpansion in times of business boom, but experience shows that lowering of the rediscount rate gives very little encouragement to increased lending in times of depression. The reason is easy to understand. To be sure, the bank that we used as an illustration in the preceding paragraph found it profitable to lend, rediscount, and lend again *if the firms to which it lent the money were good credit risks.* It made $40 on the transaction in our illustration. But, suppose that the first firm, which borrowed the $4000, is unable to repay at the end of the 90 days. The Federal Reserve Bank will try to collect and will be unable to do so. But the endorsement of the member bank appears on the commercial paper that was rediscounted, and the member bank will now have to make it good. Instead of a $40 profit, it finds itself with a $4000 loss! A bank cannot afford to lend with the idea of rediscounting unless the borrower is a good credit risk. And in times of business depression few firms are good credit risks. For that reason we discover in practice that reductions in the rediscount rate have little effectiveness in stimulating business in times of depression. For a considerable period during the depression of the 1930's the rediscount rates of the Federal Reserve Banks were 1 per cent or below, yet the member banks could not find what appeared to them to be credit risks good enough to make rediscounting worth while.

20 • 10. Reserve Requirements. In Sec. 19 • 10 we related the facts with regard to the reserve requirements for member banks. We discovered that these requirements, usually left to the discretion of bankers in other countries, are set by law in the United States. For banks that are members of the Federal Reserve System, the Board of Governors is given power to set the rates within limits. The minima had for many years been set by law at

[1] See Sec. 19 • 6.

7, 10, and 13 per cent for different classes of banks, but the Board of Governors now has the power to set the reserve requirements anywhere between these figures and double these amounts. If the board thinks that credit is being overexpanded, it can raise the reserve requirements toward or to the upper limits. Banks will then be able to extend less and less credit against a given amount of cash reserve. If, on the other hand, the board wishes to stimulate the extension of credit, it may lower the reserve requirements. Member banks will then find that they have more cash than is required. But banks do not like to hold idle funds unnecessarily. They make profits by lending money. Therefore, now that the reserve requirements have been lowered, they have good reason to extend more credit if they can find good credit risks to whom to lend. One would expect to find the reserve requirements high when an effort was being made to curtail the extension of credit, and low when the board wished to increase the amount of credit outstanding. Again, there are some difficulties with the operation of this control in practice. Even when the reserve requirements have been set at their highest, the banks have ordinarily had rather large amounts of *excess reserves*—reserves over and above those required by law. Thus on June 21, 1946, the required reserves of member banks were 15.2 billion dollars, while the actual reserves were 16.1 billion. This gave excess reserves of 0.9 billion dollars, which could have been used, of course, to support more than five times that amount of additional credit.[1] If we go further back, we find that late in 1940 and early in 1941 the excess reserves were almost as large as the required reserves. Banks were then in a position to double the amount of their deposits without overstepping the legal reserve requirements.

20·11. Open-Market Operations. The Board of Governors has power to influence the extension of credit by means of the rediscount rate and the reserve requirements. Directly it is not responsible for the third mechanism of control, the open-market operations. These are under the jurisdiction of the Open-market Committee. This is a committee composed of 12 men, of whom five are selected by the Federal Reserve Banks and the other seven are members of the Federal Reserve Board. Therefore the Federal Reserve Board has a clear working majority of the Open-market Committee if the board members stick together.

Open-market operations consist in the purchase or sale of securities on the market. In practice, the securities bought and sold have always been government bonds. The Open-market Committee can require the Federal Reserve Banks to buy or sell government bonds on the open market if such

[1] Since member banks were at that time required to hold reserves of but 20 or 14 per cent behind their deposits, as shown in Table 19·2, they could lend from five to seven times as much as the amount of their cash reserve. Therefore the 0.9 billion dollars of excess reserves could have been used to support perhaps 5 billion dollars of deposits, or thereabout.

purchases and sales seem desirable. The purpose of these operations, like that of fixing the rediscount rate or the reserve requirements is to influence the extension of credit. If, in the opinion of the Open-market Committee, business is expanding too rapidly, the Open-market Committee instructs the Federal Reserve Bank to sell bonds on the market. If we sell a million dollars' worth of bonds and collect the million dollars, we have reduced the funds against which banks can lend. If we sell the bonds directly to banks, the funds are taken immediately from them but if we sell the bonds to private individuals, they will in most cases have to draw the funds from banks in order to pay. Thus, in either case, the funds in the banks will be depleted, and the bank's reserve ratios will be reduced, reducing in turn the amount that they can lend. On the other hand, if the authorities feel that credit is too "tight"—that it is too difficult to borrow funds—they can instruct the Federal Reserve Banks to buy bonds on the market and pay cash for them. If the bonds are bought from banks, this will increase the cash holdings of the banks at once. If the bonds are bought from other people, most of the sellers will deposit their receipts in banks. In either way, the cash holdings of banks will be increased, their reserve ratios will be increased, and their ability to lend will be increased. Since banks find it profitable to lend money rather than to keep it lying idle, they will under such circumstances increase their lending if they can find prospective borrowers who are good credit risks. Buying bonds puts money into the banks and stimulates lending; selling bonds takes money from the banks and reduces lending.

It is fair to ask, of course, how can the government sell bonds whenever it wants to, or how can it be sure that it can buy bonds? The answer is, that the government can always sell bonds if it offers them at a low enough price, and it can always buy them if it offers a high enough price for them. If it offers to sell cheap, it can sell; if it offers to buy at a high price, it can buy. But will not the government lose money on its open-market operations if it buys at high prices and sells at low prices? The answer is that, even if the government does lose money on its open-market operations,[1] these operations may still be successful. Their purpose is not the making of a profit, but the reduction of the swings of the business cycle. If, through open-market operations or in any other way, we could rid ourselves of the tremendous social losses of recurring "prosperity" and "depression," we would gladly spend millions of dollars to do so; and if we lose a few thousand or a few million dollars on open-market operations, it is money well spent *if it does accomplish its purpose.* Again we must record the fact that, like the other means of control, open-market operations seem to be more effective in stemming the tide of overexpansion in prosperity than in stopping the ebb

[1] In fact the government does not by any means always lose on such operations, but what gains it makes are fortuitous by-products.

of contraction in depression. Buying bonds and putting money into the banks encourages banks to lend only if there are good prospects who wish to borrow.

Before we leave the subject, perhaps it is worth pointing out again that the effect of open-market operations is magnified by means of the reserve system. If the reserve requirements for member banks are, say, 10 per cent, then the purchase of a million dollars' worth of bonds will put a million dollars in the banks, increasing their lending capacity by 10 million dollars. On the other hand, the sale of a million dollars' worth of bonds will take a million dollars out of bank reserves, reducing the banks' lending capacity by 10 million. The banking system is one that multiplies or magnifies the effect of money, making $1 do the work of $5, or possibly of even $10 or more.

20 · 12. Federal Advisory Council. In addition to the bodies already mentioned, the Board of Governors of the Federal Reserve System is assisted by a committee of 12 members, one from each Federal Reserve District chosen annually by the board of directors of the Federal Reserve Bank of his district. This committee is known as the Federal Advisory Council and, as its name implies, its functions are purely advisory. It has no power save that of conferring once each quarter with the Board of Governors, of calling for information, and of making recommendations.

20 · 13. Federal Reserve System and Foreign Central Banks. So far we have confined our remarks to a description of the domestic activities of the Federal Reserve Banks—their relationships with member banks, with each other, and with the various overhead supervisory bodies. Although we shall turn our attention to the subject of international trade in later chapters,[1] we should mention the fact, while describing the Federal Reserve System, that it does provide for certain contracts with the central banks of foreign countries, such as the Bank of England, the Bank of France, and the Reichsbank. The Federal Reserve Bank of New York has been by far the most active in this kind of work. It has opened accounts with most of the important foreign central banks and has acted as their agent in buying for them government securities and bankers' acceptances. It has purchased gold abroad and has "earmarked" gold for the foreign central banks. It has also handled foreign bills of exchange for them.

20 · 14. Safety of the American Banking System. No kind of business, no government, has ever been completely safe. Safety is always a relative matter—one of degree. It is also a difficult thing to measure or appraise. Yet, on any basis of measurement, it must be confessed that American banking has been riskier than that of most countries. The number of bank failures, the proportion of deposits lost through failure, etc., have been larger here than

[1] See especially Chaps. 23 and 24.

in other important commercial nations. The actual magnitude of the situation in the United States is difficult to visualize but perhaps can best be depicted by the data of Table 20 • 1 and the corresponding presentation in Fig. 20 • 4. The student will note the sudden increase in failures starting with the business collapse of the early twenties, and the sudden decrease following the emergency banking legislation of 1933. The very large American failure

Table 20 • 1. Numbers of Bank Suspensions in the United States by Years, 1904-1945*

Year	Number of suspensions	Year	Number of suspensions
1904	122	1925	612
1905	79	1926	956
1906	45	1927	662
1907	41	1928	491
1908	156	1929	642
1909	69	1930	1345
1910	34	1931	2298
1911	59	1932	1456
1912	63	1933	4004
1913	46	1934	57
1914	117	1935	34
1915	124	1936	44
1916	54	1937	59
1917	42	1938	55
1918	27	1939	42
1919	43	1940	22
1920	49	1941	8
1921	501	1942	9
1922	354	1943	4
1923	648	1944	1
1924	776	1945	0

* For 1904–1931, see "The Banking Situation in the United States," p. 34, National Industrial Conference Board, New York, 1932; 1932–1941, "Statistical Abstract of the United States, 1942," p. 303; 1942–1944, *Federal Reserve Bulletin*, August, 1945, p. 795.

rate is despite the fact, alluded to earlier, that in the United States the law requires banks to hold certain minimum cash reserves, while in most countries the reserve ratio is left to the judgment of the banker. It is evident that legal specification of reserves has not been enough in itself to ensure safety of the banks or of deposits in them. Let us see why the failure rate is so high.

There may be many reasons for the unfortunate experience of American banks. Some people would ascribe it to a certain rashness of the American

temperament—to that rough frontier spirit of adventure and daring which led men in all walks of life to take chances despite the possibility of heavy loss. Probably most important among the causes of bank failures has been the fact, described in some detail in Sec. 20 • 3, that the United States is typically a country of small banks. Every tiny fresh-water hamlet feels a thrill of civic pride in having a bank of its own and, if it can have two banks or five banks, so much the better. Allied with local pride as a cause of small banks

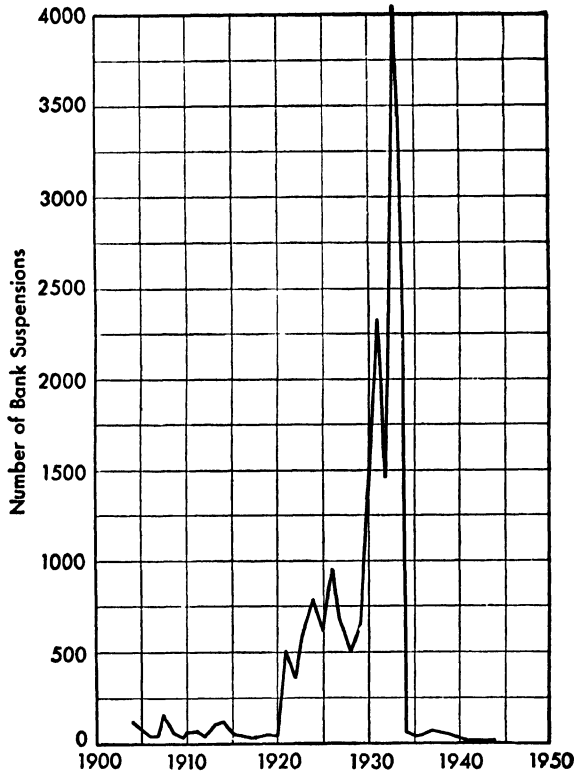


FIG. 20 • 4.—United States bank suspensions, 1904-1945.

is the American fear of monopoly, especially the fear of "Wall Street" and the "money power." This fear has led Americans to prefer many small banks to a few large ones. The facts are evident even in the state statutes. In Oklahoma, for example, the law provides that in towns under 500 population banks must have capital of at least $10,000. Missouri, Nebraska, and other states make special legal provisions for banks in towns with populations below 1000. In several states, the laws provide for the establishment of banks with capital as small as $10,000. The extreme to which the situa-

tion has gone is shown in Table 20·2, which gives the numbers of commercial banks with various amounts of deposits in towns of various sizes at about the time of the outbreak of the second World War. It will be noted that 4694 of the commercial banks, or 32 per cent of the total, fell *both* in the class of banks with deposits below 1 million dollars and the class of banks in towns of under 1000 population. Seventeen per cent of all commercial banks were in villages with populations below 500 people!

Now the facts are that one can hardly expect a bank to operate safely on a very small volume of business, or in a very small nondiversified community. The reason is a simple one. In the first place, if a given loan "goes

Table 20·2. Distribution of All Commercial Banks in the United States on June 30, 1941, According to the Population of the Place Where the Bank Was Located and According to the Amount of Deposits in the Bank*

Population of place where bank is located	Total number	Amount of deposits, millions of dollars									
		$0.1 or less	$0.1–$0.25	$0.25–$0.50	$0.50–$1	$1–$2	$2–$5	$5–$10	$10–$50	Over $50	None, or not available
All.....	14,304	456	2214	3183	3160	2340	1675	571	465	154	86
Under 250.....	758	169	378	161	41	5	1	—	—	1	2
250–500.....	1655	137	806	554	138	17	2	—	—	—	1
500–1000.....	2400	91	692	1042	485	81	8	—	—	—	1
1000–2500.....	3036	24	273	1054	1254	384	45	1	—	—	1
2500–5000.....	1799	2	33	242	730	624	157	5	2	—	4
5000–10,000....	1472	7	6	67	322	674	363	28	4	—	1
10,000–25,000 ..	1232	5	8	23	93	335	606	131	26	—	4
25,000–50,000..	560	4	5	8	26	64	219	161	60	—	13
50,000 and up...	1392	17	12	32	71	156	274	245	373	153	59

* "Banking and Monetary Statistics," p. 43, Board of Governors of the Federal Reserve System, Government Printing Office Washington, D. C., November, 1943.

bad," the proportionate effect is much greater in a small than in a large bank. In the second place, the small bank in the small town finds it impossible to get that diversity of investment which is necessary for safety in any investing business. It may be that the bank is located in a town in a cattle region, in which case almost all the loans are apt to be made to cattlemen; or it may be that the loans are concentrated in cotton, in wheat, or in the manufacture of textiles. Certain it is that they will usually be concentrated in one or two industries if the bank is a small and localized one. If, now, the particular industry runs into temporary difficulty—low prices of cattle, cotton, wheat, or textiles—the borrowers will have trouble repaying their loans, and the bank will find itself also in trouble. If, on the other hand, we consider a large bank which is lending huge sums over a wide territory,

it is likely that the industries involved will be diverse. When one of the industries is slack, another will be booming. No single industry will account for any large proportion of the business, and what is lost in some industries is gained in others.

The relationship between the location of a bank and its safety is strikingly shown by the figures in Table 20 • 3, which covers 5642 bank failures in the United States in the years 1921-1929, inclusive.[1]

20 • 15. Branch and Chain Banking. In most countries, as we have seen, this problem is met by having a small number of very large banks, each with many branches scattered throughout the territory. Thus, in Sec. 20 • 3 we spoke of England's five banks with over 8000 branches, and of Canada's four with some 2000 branches. Bank failures were practically nonexistent in these

Table 20 • 3. Number of Bank Failures in Communities of Various Sizes, United States 1921-1929

Population of Community	Number of Failures
Under 500	2278
500–1000	1133
1000–1500	567
1500–2500	537
2500–5000	413
5000–10,000	239
10,000–25,000	180
25,000 and over	295
Total	5242

countries even in the depths of the depression. We can contrast this with the 14,000 commercial banks in the United States—one bank for every 10,000 people! Yet the number of banks in this country is smaller than it was a few years ago, partly because of the large number of bank failures during the depression of the thirties, and partly because of the tendency toward amalgamation that has been felt in banking circles just as it has elsewhere in business.

In many of the states of the United States, banks have been prohibited by law from setting up branches; in others, branch banks have been strictly limited. Some states have allowed state banks to set up branches, but only within the state or the city. When, in 1921, the Comptroller of the Currency authorized national banks to establish branches under certain restrictions in states where state banks already had the power, the U. S. Supreme Court ruled that he had no authority to give such authorization. As a result of this ruling, the comptroller authorized national banks to open additional *offices*

[1] Data from "The Banking Situation in the United States," p. 40, National Industrial Conference Board, New York, 1932.

within the same city, on the supposition that offices differed in some way from branches. Offices were supposed merely to receive and pay out deposits, and not to carry on other banking functions. In 1927, Congress gave national banks in any state the same powers to establish the branches that were given to state banks in the same state, and the present tendency is for state laws on branch banking to be liberalized. Yet, it is still true that branch or chain banking is much more closely restricted than are similar activities in most other industries.

Banks have themselves often wished to establish branches for many different reasons. To some extent they have felt that a business operated on a larger scale would be able to make operating economies that would increase the profit. In part there has been mere pride in size. The increasing congestion of traffic in large cities has made it hard for customers to reach the bank from all parts and has caused banks, where they could, to open branch offices at various convenient points throughout the city. As branch banking becomes country-wide, it tends to increase the mobility of the nation's credit resources, by shifting them quickly and easily from the places of excess to the places of deficit. Probably most important of all is the argument that branch banking would bring the diversity that is necessary if we are to reduce the number of bank failures, which even in the prosperous years from 1924 to 1928 averaged 750 failures per year.

On the other hand, those who oppose branch banking fear the development of a monopolistic money power and the increasing concentration of that power in the hands of the few to the detriment of the many. They fear that branch banks would tend to drain the resources of small communities into the great financial centers, and that the great banks would be less understanding of and sympathetic to local needs. It is argued that the small country bank cannot well develop branches, and that as a result the tendency will be for them to be taken over as branches of great city systems. Much of the opposition to branch banking has come from small country bankers, as well as from those who fear concentration of economic power in general. Finally, the opponents would say that the safety of the banking system may be obtained by other and less dangerous methods, particularly by a system of insuring deposits in banks. We shall turn to a description of such insurance in the next section but, first, we should say that the consensus in banking circles seems to be that branch banking and chain banking would bring certain added efficiencies to the banking system and that the tendency toward them is certain to continue.[1]

20 · 16. Guarantee of Bank Deposits. Bank failures, which had occurred in considerable numbers ever since 1921, increased tremendously with the

[1] Chain banking, as distinct from branch banking, is the control of a number of separate banks by means of a holding company. Holding companies are described in Sec. 25 • 4.

coming of the thirties. In the three years, 1930-1932, there were some 5000 bank failures in the United States, and in the single month of October, 1931, there were 522 suspensions, involving 471 million dollars of deposits. In the hectic days of early 1933, the number of failures had soared to 4000 (see Fig. 20·4). It was against this dark background of inexcusable inefficiency and irretrievable loss that Congress decided to take some action protecting depositors. In 1933, they set up a temporary organization, made permanent in the Banking Act of 1935, called the "Federal Deposit Insurance Corporation." This agency is commonly called the FDIC, and we shall follow that practice hereafter.

The FDIC is managed by a board of directors of three men: the Comptroller of the Currency acting ex officio and two others appointed by the President of the United States. The original capital was subscribed by the United States Treasury and the Federal Reserve Banks. The Treasury subscribed 150 million dollars, and each Federal Reserve Bank subscribed an amount equal to half its surplus. The latter subscriptions ran to a total of approximately 140 million dollars, making the total original capital about 290 million.

Every bank that is a member of the Federal Reserve System is required by law to participate in the "insurance"[1] program, and nonmember banks may be admitted by the FDIC if they meet the requirements. On Dec. 31, 1945, 13,297 of the 14,011 commercial banks holding 121.8 of the 124.0 billion dollars deposits were covered by insurance. Thus, the plan covered 95 per cent of the commercial banks which held 98 per cent of the total deposits. However, as we shall see in a moment, the entire amount of these deposits is not covered by insurance.

Each insured bank is required to pay a semi-annual assessment equal to one twenty-fourth of 1 per cent of the amount of its deposits—one-twelfth of 1 per cent each year. These amounts are added to the original capital of 290 million dollars and invested in government securities. If an insured bank fails, the FDIC is required to pay the depositors as soon as possible the full amount of their deposits up to $5000. In other words, only the first $5000 of each deposit is insured, and depositors who have larger sums than that on deposit must wait to collect, if they are able to collect at all, from the assets of the failed bank. It is to be noted that the bank is assessed on the total amount of its deposits, even though but part of the deposits will be insured.

[1] The FDIC does not actually "insure" deposits as the term is properly used. No actuarial tables are used, and there is no computation of risk or variance of premium to offset variations in risk. The system is technically one that "guarantees" deposits rather than "insuring" them. However, the law mistakenly refers to it as insurance, and we shall continue to use that name in the discussion here.

To be sure, the vast majority of deposits are smaller than $5000 but, to the extent that a bank caters to large depositors it is assessed to assist those banks which have small depositors. Strong and efficient banks are assessed to protect weak and inefficient ones. This has been one of the major objections to the present system. Although there is without question some merit in this objection, and although the system has been in operation for too short a time so far for us to give any final appraisal of it, nevertheless it appears at present that one can state with reasonable safety that the FDIC program has so far been successful. At least, it did restore confidence in banks at a time when confidence has been badly shaken. Figure 20·4 indicates that the bank failure rate fell sharply after the imposition of the system. The student should be warned that there is no reason to suppose that the FDIC is in any large measure the cause of that decrease in the bank failure rate, but it is certain that it did play some part in it. Also, it should be emphasized that the purposes of the FDIC are not confined to stepping in after banks have failed, but that they include also the taking of whatever measures seem wise to keep banks in good condition so that failures will not take place.

The FDIC has tremendous resources with which to repay depositors of closed banks. Not only does it have the original capital of some 290 million dollars and the semiannual assessments from members and the interest on the government bonds in which the assessments are invested, but also it has the right to borrow up to approximately 1 billion dollars on its own notes or bonds. The corporation's investments in government securities on Mar. 31, 1945, amounted to 818 million dollars. This can be compared with total deposits of all failed banks from 1934 to 1945 of less than 145 million dollars. The resources available appear to be sufficient to meet any ordinary demands. Yet, there are some authorities who doubt the ability of the FDIC to meet another serious depression. They argue that, as we have already pointed out, the system does, in effect, tax the strong bank for the benefit of the weak. They fear that, if another major depression should bring large numbers of bank failures, many of the strong banks would find it wise to withdraw from the insurance system rather than bear the expense. For state banks, this would mean giving up membership in the Federal Reserve System; for national banks it would mean giving up the bank's charter and becoming a state bank. Therefore, these critics fear that the FDIC is so organized that the effect of a serious depression would be to disrupt the Federal Reserve System and the whole system of national banks by making it profitable for banks to withdraw. There seems to be no way of judging the likelihood of such an event in advance.

20·17. Banking and Confidence. Banking, as we have discovered in this and the preceding chapter, is a peculiar sort of business—one in which

people talk constantly in terms of money without really handling very much actual cash. Depositors think that they have money in the bank when they have none. Men paying for goods by check think that they have paid cash when they have not. The total deposits in all the commercial banks in the United States on Dec. 31, 1945, amounted to 150 billion dollars, and the total amount of money in circulation in the country on that same day was but 28 billion. Yet, although much of the work of a bank is carried on with "fictitious money" rather than with real money, it is important to realize that banks are dependent in a very direct way on actual cash. The bank does have to have enough cash on hand so that it can pay on demand those depositors who desire payment. No commercial bank has enough cash on hand to pay all its depositors, even though they do have the right to demand payment in cash. The insured commercial banks in the country had, on Dec. 31, 1945, demand deposits of 74.7 billion dollars. The depositors could have demanded payment of this amount if they had wished. Yet these same banks on that day had a total of but 1.8 billion dollars cash in their vaults. Although the depositors had a right to demand 74.7 billion dollars, it is transparently clear that they could not have procured it.

A concerted demand for cash on the part of the depositors of any commercial bank will always wreck it. The bank is legally obligated to pay, yet it is unable to do so. To be sure, the bank has assets enough, but it has invested them in forms that cannot be converted immediately into cash. It has, for example, funds lent out for 30, 60, or 90 days. If the depositor were willing to wait until the loan came due, the bank could collect the funds and pay him. But he has a right to demand payment at once. If enough of the depositors in any bank become worried about its ability to pay and start to withdraw their funds in anticipation of a bank failure, they will force a bank failure. When depositors lose confidence and start a "run"—a concerted withdrawal of funds—the bank is immediately in trouble no matter how carefully and conservatively and honestly it has been operated. On the other hand, if people have confidence in a bank and do not withdraw their funds, then it needs no cash at all. We have the peculiar situation that, if people believe that a bank is in good condition, it is reasonably safe; but, when people think that the bank is in bad condition, it can look for trouble ahead. Confidence is peculiarly necessary in the banking business. As President Franklin Roosevelt said in his first inaugural address, in the midst of the mounting wave of bank failures that dwarfed all precedents, "The only thing we have to fear is fear itself!" The famous "bank holiday" that followed was of some importance in giving the banks time to convert at least part of their assets into cash, but by far its greatest importance was in giving people time to regain their lost confidence. Confidence is the most important single asset of a commercial bank.

Questions

1. What were the major functions of the two United States Banks?
2. Why, in each case, were the charters of the United States Banks not renewed?
3. What were the major accomplishments of the two United States Banks?
4. What was the Independent Treasury System?
5. What were the wildcat banks, and what were their peculiarities?
6. What were the greatest weaknesses in the American banking scheme just before the introduction of the National Banking System?
7. What were the major accomplishments of the National Banking System?
8. What were the major shortcomings of the National Banking System?
9. What is the present status of bank notes issued by state banks?
10. What were the major advantages and disadvantages of the national-bank notes?
11. Explain what is meant by pyramiding bank reserves.
12. Why do we have 12 Federal Reserve Banks instead of one? List all the advantages for such an arrangement. What disadvantages are there?
13. Why might the officers of a bank prefer not to be members of the Federal Reserve System? Why might they think that membership was advantageous?
14. "The Federal Reserve Banks are 12 separate privately owned banks." The foregoing statement is technically correct, yet by itself it would be apt to mislead an uninformed person. What needs to be added to the statement to ensure adequate comprehension?
15. On July 31, 1945, the 12 Federal Reserve Banks had a total of 23.3 billion dollars of Federal Reserve notes outstanding. On the same date the total deposits in these Federal Reserve Banks were 16.9 billion dollars. How large an amount did the banks need in gold certificates as a reserve?
16. On July 31, 1945, the 12 Federal Reserve Banks actually held gold certificate reserves of 18.0 billion dollars. What was the Federal Reserve ratio? (The Federal Reserve ratio is the percentage which the reserves are of notes and deposits.)
17. If a man in Omaha, Neb., buys something from a man in Memphis, Tenn., and sends his check in payment, through what hands will it pass?
18. What endorsements will the check of question 17 bear when it is returned to the drawer?
19. What are correspondent banks?
20. What is the interdistrict settlement fund, and for what is it used?

21. Explain in simple terms what is meant by the loose statement that "most checks are not paid in cash, but are canceled out."

22. How is it possible to clear checks that run to a greater total amount than all the money there is?

23. Just what steps would be available for use by the authorities of the Federal Reserve System if they wished to curb what seemed to them an overexpansion of credit? Explain just how each of the steps would operate.

24. Why is it that the methods enumerated in answer to the preceding question seem actually to work better in curbing overexpansion than in stimulating business in times of contraction?

25. What are excess reserves? Of what importance are they?

26. How do you explain the fact that bank failures have been very common in the United States and almost unheard of in many other countries?

27. Do you believe that the restrictions on branch banking should or should not be relaxed? Discuss.

28. What is the difference between insuring bank deposits and guaranteeing bank deposits?

29. Explain why confidence is the most important single asset of a commercial bank. Does it appear on the balance sheet?

21. THE GENERAL LEVEL OF PRICES

21 · 1. General Price Level. It is a matter of common observation that prices are not stable but, on the contrary, are constantly shifting and changing. We have spent much of our time in preceding chapters in describing some of the more important factors affecting prices of commodities. Now we note the fact that sometimes the prices of many, even of most, commodities start to move together in the same direction. Instead of haphazard and unrelated movements, with some prices rising, some falling, and some staying the same, we find a noticeable general drift of prices upward or downward. We speak of times of "high prices" or of "high cost of living," and we realize that, while we are not thinking of any particular commodity by itself, nevertheless the sum total of all prices seems to us to be higher than usual.

Of course, if some prices rise and some fall, it is hard for us to say with assurance whether prices "on the whole" have risen, fallen, or stayed the same. It is necessary for us to devise some instrument by which we can measure the *average* movement and determine what has happened to the *general level of prices*. This is done by means of *index numbers*. Index numbers are complex things, about which learned men have written long books, and it is quite impossible in a general textbook on economics to consider all the refinements of technique that go into their computation. But we can say enough about them to give some idea of the nature of the problem, and particularly to allow the student to see how the end products—the index numbers themselves—are used.

Let us suppose that we have but five commodities that are bought and sold on the market: milk, meat, houses, clothing, and toothpicks. We will not complicate our problem here but will assume that the meat is all of uniform quality instead of being part sirloin and part tripe, and similarly that the other commodities are uniform. Let us now suppose that the prices of these five commodities in 1945 and in 1947 are those given in Table 21 · 1. We notice that three of them (milk, meat, and housing) have fallen in price, one (toothpicks) has risen in price, and one (clothing) has remained un-

changed. Should we say that the 1947 prices have risen, fallen, or remained unchanged as compared with the prices of 1945?

We note first that prices are high or low or medium as compared with those of some other time. The prices of one year alone tell us nothing. We need some other set of prices with which to compare them. In our problem we have assumed that someone wants to use the 1945 prices as the basis for comparison. In such a case, we say that 1945 is the "base period," which means merely that it is the period selected, more or less arbitrarily, as the basis for comparison. After we have computed our index number, we shall indicate the base period by stating that 1945 = 100. Sometimes we see index numbers given with an accompanying statement of such a kind as "1910-1914 = 100," which means merely that the statistician is comparing the prices at other times with the average prices for the years 1910-1914. What

Table 21·1. Basic Price Data for Index Numbers

Commodity	Price per unit	
	1945	1947
Milk, per quart	$00.14	$00.12
Meat, per pound	00 58	00.43
House rent, per month	55 00	50.00
Cost of a new suit	42 00	42 00
Toothpicks, per box	00.05	00 20

we want in our particular simple problem is to find the index number of prices in 1947 with 1945 = 100. Although there are many different ways of finding such an index number and although these ways differ in some rather important respects, we shall leave the technical problems to the statistician and dodge all the difficulties. One simple way, obviously, in which we could form an opinion as to the change that had taken place in prices in general would be to find for each commodity what percentage the 1947 price was of the 1945 price. Thus, for example, the 1947 price of milk (12 cents) is 12/14, or 86 per cent of the 1945 price. If we make a similar computation for each commodity we get the following relative prices in 1947: milk 86, meat 74, rent 91, clothing 100, and toothpicks 400. Each of these *relative prices* shows the 1947 price as a percentage of the price in the base period, 1945. If we now take the simple arithmetic average of these relative prices we get the simplest form of index number. In our case, we find that

$$\frac{86 + 74 + 91 + 100 + 400}{5} = \frac{751}{5} = 150.2$$

We can then say that the 1947 index number is 150.2 "on a 1945 base."

21 · 2. Weighting Index Numbers. The index number just computed, however, is so crude as to be useless. The student will note right away that we have treated all the price changes as being equally important. We have pretended that a fourfold increase in the price of toothpicks was just as important as a 26 per cent reduction in the price of meat. Actually, we realize at once that changes in the price of meat and milk are much more important to consumers than are changes in the price of toothpicks and, if we wish to use our index number to show what has happened to the prices of things that consumers buy, we should not treat all the price changes alike. We should give them differing *weights*.[1] Thus, if we feel that the cost of clothing is 25 times as important to consumers as the price of toothpicks, that milk is 40 times as important, that meat is 45 times as important, and that rent is 60 times as important,[2] we could give the commodities the weights of 1, 25,

Table 21·2. Computation of Weighted Index Number from Data of the Text (Based on Table 21·1)

Commodity	Relative price	Weight	Product
Milk	86	40	3440
Meat	74	45	3330
Housing	91	60	5460
Clothing	100	25	2500
Toothpicks	400	1	400
Totals	...	171	15,130

Index number = 15,130 ÷ 171 = 88 5

40, 45, and 60. Now we would compute our index number by means of the method illustrated in Table 21 · 2. Here we see that each price relative has been multiplied by the corresponding weight, the products have been added together, and their sum has been divided by the sum of the weights. The answer, 88.5, is the weighted 1947 index number of prices on a 1945 basis.

21 · 3. Interpretation of Index Numbers. We shall have to assume that the statisticians who have computed any index numbers that we use took

[1] The word "weight" is used here, not in the sense in which one uses it in physics, but in the sense in which one uses it when he says, "I give a good deal of weight to his opinion." It means "importance."

[2] We might find the relative importance of these different items, for example, by studying consumer budgets to find how much average families spend on each of the commodities. If we found that they spent 25 times as much per year on clothing as on toothpicks, then we might give them the weights described above.

proper precautions in their work and selected original data and methods adapted to yield information of the type that we need. If we can assume that this was true in the case of the computation just completed, what does the result mean? How do we interpret the figure 88.5? First of all, let us remember that every index number is a percentage and also an average. It shows us what percentage the prices in any given year are, on the average, of the prices in the base period. Thus in the present case, we learn that the prices in 1947 were, on the average, 88.5 per cent of the 1945 prices. This is true, of course, only if we have given the commodities their proper weighting —their proper relative importance. If that is so, we can say that the changes in price that actually took place had about the same effect on consumers as if all prices of all commodities had dropped to 88.5 per cent of their former amount. We can say that prices in 1947 are, on the average, 88.5 per cent of what they were in the base period, or that prices have, on the average, dropped 11.5 per cent. (They have dropped from 100 to 88.5 per cent, which is a drop of 11.5 per cent.) We can say that, on the average, consumers in 1947 have to spend only 88.5 cents to buy what cost them $1 in 1945. Any of these interpretations will serve our purpose reasonably well, and any of them will show what we mean when we talk about changes in the general price level.

To be sure, in practice the statistician would not be content to measure the changes in but five commodities. Hundreds of different commodities are used. In fact, the U. S. Bureau of Labor Statistics publishes index numbers of the wholesale prices of "all commodities" based on well over 800 different items. In such a case, one can be reasonably well assured that the resulting index number does indicate, at least roughly, the general level of prices.

21 · 4. Price Level and the Value of Money. Sometimes, as we have said, the prices of most commodities begin to rise. As the prices get higher and higher, a dollar will buy less and less. People say, "The dollar doesn't go so far as it used to." We see that the change in the price level can be thought of as a change in the value of money. We do not say that the value of money changes *because* the price level changes. They are not cause and effect, but the same thing. If prices rise to double what they used to be, it is evident that a dollar will buy only half as much as before. Doubling prices is exactly the same as cutting the value of money in two. The general price level and the value of money are reciprocals of each other. They vary in exact inverse proportion. Thus, economists sometimes say that prices tend to be high in wartime; but they sometimes alternatively make the statement in the form that the value of money tends to be low in wartime. It is important for the student to get in mind the fact that the general price level and the value of money are inseparably connected, and that when prices rise the value of money falls.

21 · 5. History of Changes in the Price Level. Figure 21 · 1 and Table 21 · 3 show the changes that have taken place in the general level of wholesale prices in the United States between 1798 and 1946. The student will be struck at once by the tremendous variation in them. He will note the great upswings in prices that have accompanied all major wars, and he will note also the smaller but nevertheless pronounced intermediate swings between times. He will note the long drop in prices after the War of 1812, which reached a low point in the 1840's. Similarly there is a deep trough between

Table 21·3. Index Numbers of Wholesale Prices in the United States, 1798-1946; 1926 = 100*

Year	0	1	2	3	4	5	6	7	8	9
1790	..	..	..	..	..	..	..	..	84	86
1800	88	97	80	81	86	97	92	89	79	89
1810	90	86	90	111	125	116	103	103	101	86
1820	73	70	73	71	67	71	68	67	66	66
1830	62	64	65	65	62	68	78	79	75	77
1840	65	63	56	51	53	57	57	62	56	56
1850	58	57	60	66	74	75	72	76	64	65
1860	64	61	71	91	132	127	120	111	108	103
1870	92	89	93	91	86	81	75	73	62	62
1880	68	71	74	69	64	58	56	58	59	55
1890	56	56	52	53	48	49	47	47	49	53
1900	56	55	59	60	60	60	62	65	63	68
1910	71	65	69	70	68	69	86	118	131	138
1920	155	98	97	101	98	103	100	95	97	95
1930	86	73	65	66	75	80	81	86	78	77
1940	78	87	98	103	104	106	121			

* Computed from 1910–1914 index numbers in "Farm Economics," Cornell University, pp. 1586-1587, 1634, and scattering thereafter, by multiplying by 0.6846.

the Civil War and the first World War, reaching a low point about 1896 or 1897. At least up to the time of this writing, prices had been kept much more closely in control during and after the second World War than during preceding wars, but the possibilities of great postwar inflation are still present as we go into 1947.

21 · 6. Effect of Price Changes. If all prices changed equally and simultaneously, we would not need to worry particularly about the fluctuations depicted in Fig. 21 · 1. The potato farmer need not worry to find his shoes and gasoline and electricity doubled in price if his potatoes, too, are selling for twice as much. In fact, however, prices do not all move together nor by the same amount. We learn from experience that, when the general price level is

rising or falling, some prices tend to lead the way and some to lag behind. Usually, for example, wholesale prices move up or down before retail prices, and the wages of unskilled labor before those of skilled labor. Salaries are particularly inflexible and slow to move. As a consequence of these *price lags,* manufacturers are likely to make unusually high profits when prices are rising. They buy raw materials at a given price but, before they have finished manufacturing them and getting them ready for sale, the price has risen. As a result their selling price keeps a jump ahead of their buying price, and their profit margin increases. But times of rising prices are apt to be hard for

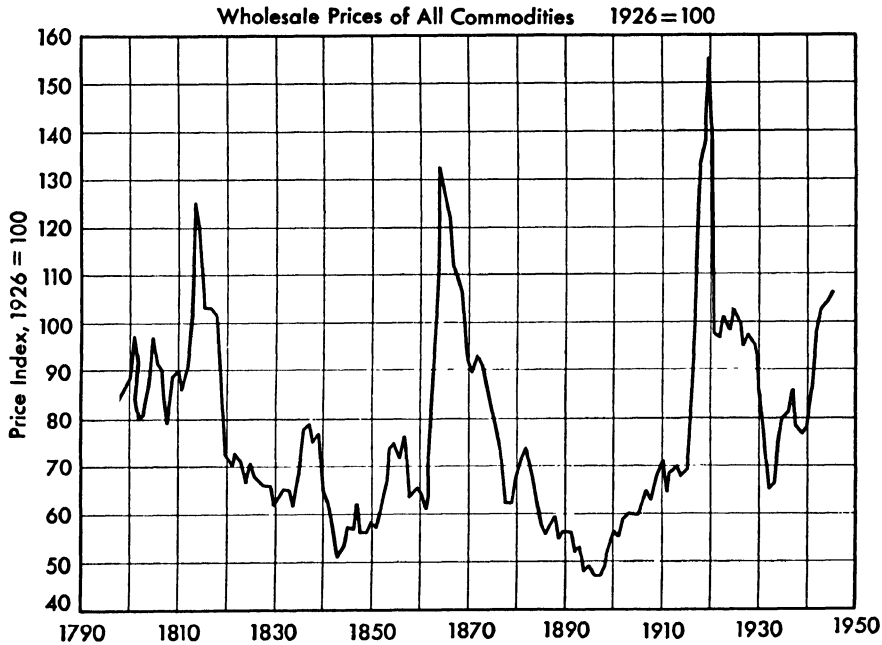


FIG. 21 • 1.—Wholesale prices of all commodities, 1926 = 100.

skilled workers and salaried people. They find the prices of the things that they buy rising, while their incomes lag behind.

Changing price levels are felt particularly in those cases where money is used as a store of value or as a standard of deferred payments.[1] In these cases the lapse of time involved makes it possible for the value of money to change. If a coal-mining corporation borrows $10,000 to buy new equipment, thinking that the increased output of coal will make it possible to repay the loan over a period of, say, 20 years, and if the 20-year period is one of steadily falling prices such as that which occurred from 1864 to 1884, the company will find it harder and harder to repay the loan. If the contract

[1] See Sec. 18 • 21.

calls for a payment of 6 per cent interest each year, the company will be forced to pay $600 annually but, since the price of coal is presumably falling along with everything else, the company will have to mine and sell more and more coal each year to meet the $600 interest bill. The bill calls for the same amount of money, $600, each year; but it calls for more and more coal each year. Reference to Table 21·3 indicates that prices fell during the period 1864-1884 to less than half their original level. It would be necessary for the coal company to mine and sell twice as much coal at the end of the period as at the beginning to pay the same interest bill.[1] And at the end of the 20-year period, when they came to repay the $10,000 that they had borrowed, it would take twice as much coal to raise the money as they had anticipated when the loan was contracted. Debtors always suffer a loss from falling prices and find the burden of their debt growing greater and greater. A period of *deflation,* as we call any period of generally falling prices, hurts debtors but helps creditors. The creditor is receiving back the same amount of money that he lent, but it will buy more at the new low prices than it would have bought at the high prices that existed at the time he lent it. The man who has money coming to him from his deposit in the savings bank or from his life-insurance policy always gains by deflation, just as the debtor always loses at such a time.

On the other hand, during periods of *inflation,* as we call any period of generally rising prices, debtors gain. The coal company of our example finds that it takes smaller and smaller amounts of coal each year to raise the necessary money, because the price of coal has risen. But creditors, receiving a fixed annual sum of money, find that it buys less and less as prices rise. A man who had retired on a fixed income in 1897, for example, would have found greater and greater difficulty in getting along during the next 20 years. Reference to Table 21·3 or to Fig. 21·1 indicates that prices jumped from 47 in 1897 to 155 in 1920. In the latter year prices were, on the average, 3.3 times as high as in 1897. The man with the fixed income would be able to buy less than a third as much in 1920 as he had bought in 1897. The widow cashing in an insurance policy in 1920 got much less for her money than the widow whose husband died a dozen years earlier or a dozen years later. Inflation benefits the debtor and hurts the creditor, or anyone else who has money due. Deflation benefits the creditor and hurts the debtor.

An understanding of these facts helps us to understand many episodes in history which would otherwise be inexplicable. For example, no one can understand the agricultural unrest, the Populist and Granger movements, the Free-silver bimetallist movement, and the other symptoms of discontent

[1] This is assuming, of course, that coal prices moved during the period approximately as the average of all prices.

in the Middle West during the eighties and early nineties unless he recalls (1) the fact, shown in Fig. 21 · 1, that prices were falling and had been falling for several decades and (2) the fact that the agricultural groups of the country were at that time, by and large, debtor groups who were suffering from the deflation—from something which they did not understand and could not control and for which they were not responsible. The great swings of prices depicted in the figure have accounted for many of the most important social and economic movements of their times.

21 · 7. Causes of Price Movements. The causes of the major movements of prices about which we have just been speaking are complex, yet they can be summarized easily under a relatively few headings. Suppose that we have a group of five men, each of whom has one dollar. Each of them buys something from one of the others and spends his dollar. The five one-dollar bills have done five dollars' worth of business. But we might as well have done all the business with one one-dollar bill. Mr. A could have bought something from Mr. B, and paid him the dollar. Mr. B could then have bought from Mr. C and handed the dollar along. Mr. C could have bought from Mr. D, Mr. D from Mr. E, and finally Mr. E from Mr. A. In each case the buyer would hand along the dollar that he had just received, and at the end we would have done five dollars' worth of business by spending the single dollar bill five times. It makes no difference whether we spend five one-dollar bills once, or one one-dollar bill five times. The total amount of business is the same. If we have 10 one-dollar bills and spend each of them four times, we have done $40 worth of business. It is evident that the total amount of business carried on with money will be equal to the amount of money multiplied by the average number of times that each unit is spent. If we let M represent the amount of money and let V represent the velocity of circulation (that is, the average number of times that each dollar of money is spent), then we can readily see that the total business done with money will be MV.

But we have learned that far more business is done with credit, especially with bank deposits, than with money. Here again we have to consider not only the amount of deposits but the rate at which they are spent—the velocity of turnover or of circulation. If Mr. A has a bank deposit of $100 and gives his check to Mr. B, we have learned that the bank will reduce Mr. A's deposit and increase Mr. B's. We still have $100 of deposits, but they have been spent once. If Mr. B now gives his check for $100 to Mr. C, the total deposits stay at $100, but they have been spent twice, and we have done $200 of business with $100 of credit. If the deposits are thus spent ten times in a year, the $100 of credit does $1000 of business. If we represent the amount of bank deposits and other noncash means of payment by M' and the velocity of circulation by V', then the total business done on credit is

$M'V'$, just as the total business done with cash is MV. The total business done in the community will, of course, be the sum of the cash business and the noncash business, or $MV + M'V'$.

If I buy a pair of shoes for \$8, you buy a loaf of bread for 10 cents, and John Doe buys a package of tobacco for 15 cents, there have been three transactions at an average price of \$2.75. The total amount of business done has been $\$8 + 0.10 + 0.15 = \$8.25$; or we could as easily have found the total by saying that there were three transactions at an average price of \$2.75 and that the total business was $3 \times \$2.75 = \$8.25$. If the number of transactions is T and the average price per transaction is P, then the total volume of business is PT. But we have already noted at the end of the preceding paragraph that the total volume of business is $MV + M'V'$. We can represent the total volume of business either way. Since this total volume is equal to $MV + M'V'$, and also equal to PT, it is evident that

$$MV + M'V' = PT$$

This is known as the "equation of exchange" and is due to one of America's ablest economists, Dr. Irving Fisher. If we recall the meanings of the symbols, we shall see that the equation merely states the truism that "the total value of the monetary transactions plus the total value of the non-monetary transactions equals the value of all the transactions." Since we are interested in the general level of prices P, and the factors that affect it, we may restate the equation in the following form:

$$P = \frac{MV + M'V'}{T}$$

This helps us to see that the general price level, or the average of all prices, varies directly with the volume of money and credit and with the velocity of their circulation and inversely with the volume of trade or the number of transactions.

The equation of exchange is of no actual use in computing changes in the general price level. Its advantage is that it summarizes compactly the forces that affect the general price level, and enables us to see what kinds of effects we can anticipate as a result of changes in any of its elements. Thus, for example, we may be interested in knowing the effect of changes in the amount of money in circulation, which we have represented by M. We note from the equation that, if we increased the amount of money only and changed none of the other factors on the right-hand side of the last equation, the general price level P would have to rise. If the increase in the volume of money were accompanied by a reduction in the speed with which it was spent, reducing V as M increased, it is quite possible that the product MV might stay the same or even fall. We cannot tell the effect of increasing

the volume of money unless we know also what will happen to the rate at which it is spent. In addition, of course, we realize that changes in M may well affect M'. If we increase the amount of money in circulation, we increase the cash reserves of banks. Under these circumstance banks are able to lend more, to increase the amount of their deposits, and thus to increase M'. Many authorities have assumed that changes in M will, at least in the long run, bring proportional changes in M'—that doubling or cutting in half the amount of money would double or cut in half the amount of bank deposits. It is by no means certain that the relationship between M and M' is so close as this, but certain it is that they tend to vary in the same direction.

The amount of money in circulation is directly under control of the government, and the amount of deposits is indirectly so. The control over deposits is in part exercised through control of the amount of money, but also in part by means of manipulation of rediscount rates, open-market operations, and reserve requirements.[1] But, while the government can control M and M' to a considerable extent, it can do little of a direct nature to control V and V', the rate of spending. This is especially important, because that is the place where prices are most likely to get out of control. A government may feel that it is safe to issue fiat money to finance a war or a depression, for example, and the government fiscal authorities may decide to follow conservative policies and limit sharply the quantity of inconvertible money issued. The increase in M and the consequent increase in M' will presumably raise prices somewhat, but the government authorities feel that they can keep the matter in hand by controlling the quantity of issue. This is where they have often been mistaken, because, if people begin to lose confidence in the money, and fear that it will not be wisely issued or possibly that it will never be redeemed, they naturally start to spend it quickly, so as to get rid of it while it is still worth something. Thus, if you had a ten-dollar-bill today and received what you thought was reliable information that all such bills were to be canceled and become worthless tomorrow, you would probably run right out and spent it. When people are worried about their money (or their bank deposit), they tend to spend it and, as everyone rushes to spend it, the velocity of circulation rises. A spectacular illustration of this tendency occurred during the terrible and unprecedented German inflation following the first World War. Between 1921 and 1923 the velocity of turnover of German marks increased roughly 17-fold.[2] This means that although the German people had about 7 billion times as many marks as they had had ordinarily, nevertheless they were holding each mark only about one-seventeenth as long as usual—that they were spending each mark

[1] See Secs. 20 • 8 through 20 • 11.

[2] BRESCIANI-TURRONI, "The Economics of Inflation," p. 168, George Allen & Unwin, Ltd., London, 1937.

about seventeen times as fast. It is easy for the student to see, if M increases 7 billion-fold and V increases 17-fold, what happens to MV and what is certain to happen to the general price level P. The Hungarian inflation of July, 1946, was even more sudden and extreme.

The significant point for the moment is that, although the government can exercise considerable control over M and M', the values of V and of V' are determined largely by the people's confidence in the government, especially in the money. If they lose confidence, the velocity rises and there is no effective way for the government to control prices. Inflation then is no longer "managed," and those who talk of the advantages of a "managed currency" as a means of controlling the general price level often overlook the fact that in a pinch the "manager" loses his authority.

We can illustrate the operation of the equation of exchange over and over again from actual occurrences. The German inflation and the issue of inconvertible greenbacks during the Civil War are cases in which an increase in the volume of money brought a rising price level. The discovery of gold in California and the subsequent increase in its volume were largely responsible for the reversal of the downward trend of prices that had followed the Napoleonic Wars. A steadily increasing volume of trade after the Civil War, without corresponding increase in the volume of money, was largely responsible for the long trend of falling prices between the end of the war and 1896. Then increased gold supplies from the mines of Alaska and South Africa again reversed the trend and brought the rise that continued until the first World War.

21 · 8. Quantity Theory of Money. Early monetary theorists noted that issues of fiat money, or increased outputs of gold, were accompanied usually by rising prices. They argued that the value of money varied inversely with the supply of it; if we increased the supply, we decreased the value or raised prices. Thus they were led to enunciate the *quantity theory of the value of money,* or, as it is often called, merely the "quantity theory of money." This theory stated that the value of money varied inversely with the quantity of it; or, if we state it in alternative terms, that the price level varied directly with the quantity of money. Increases in the quantity of money meant higher prices.

Today, few if any economists would describe the facts so baldly or so incompletely. It is now well understood that other things besides the quantity of money must be considered. We might perhaps still say that increases in the quantity of money would be followed by rising prices *if other things remained equal,* but the other things are so important that it pays us to investigate them, too. Some theorists would argue that the factors of the equation of exchange are so interrelated that changes in the quantity of money will actually raise prices through its effect on the other factors. They

would argue that, if we increase the volume of money M, we increase bank reserves and thus tend to increase M'. Although this may temporarily reduce the velocity of turnover, since people have more money and therefore spend each dollar less often, nevertheless they will soon find that there are no more goods to be had than before, since increasing the quantity of money does not increase the quantity of goods available for purchase. The volume of trade, T, stays unchanged. Therefore, they would argue, if we have more money and more credit and no more goods, people will soon bid up the price and P will increase. This is, perhaps, the modern statement of the quantity theory; namely, that increases in the quantity of money will affect the other items of the equation in such a way that ultimately the price level will tend to rise. But most modern theorists look upon the equation of exchange merely as a convenient summary of the forces involved and the general direction of their effect.

Table 21·4. Velocity of Circulation of Bank Deposits, Banks of 100 United States Cities outside New York, 1936-1945*

Year	Velocity, times per year	Year	Velocity, times per year
1936	22.4	1941	19.4
1937	22.4	1942	18.4
1938	19.9	1943	17.4
1939	19 4	1944	17.3
1940	18.6	1945	16.1

* *Federal Reserve Bulletin*, vol 32, no 7, July, 1946, p. 770.

The actual sizes of most of the items involved in the equation of exchange are hard to measure. The velocity of circulation of bank deposits M' can be measured with reasonable accuracy, and we find that it varies considerably from time to time. For New York City banks it reached as high a figure as approximately 100 times per year in 1929. In that year, in other words, each dollar of deposits in New York banks was being spent about 100 times a year, or once every 3 or 4 days! In contrast, the figure for 1939 was only 20; the average dollar of deposits was spent 20 times a year, or once every 18 days.[1] For more recent years, the estimates of V' are given in Table 21·4. Perhaps an average figure is a velocity of about 20 times a year of slightly more. It is generally supposed that the velocity of turnover of money V, is somewhat less than the velocity of turnover or circulation of credit V'. Some authors have thought that it is perhaps half as great. But it is much more difficult to get measurements of this velocity, and

[1] *Federal Reserve Bulletin*, vol. 26, no. 1, January 1940, p. 7.

numerical results are largely conjecture. The total amount of money in circulation M, and the total amount of bank deposits M', can be found readily and accurately. They vary quite considerably in amount from time to time, but the figures for the ends of the years 1936-1945 are given in Table 21·5.

21·9. Illustration of the Equation of Exchange. Professor Morris Copeland has attempted[1] to measure the actual changes that have taken place in the items making up the equation of exchange, and perhaps the ideas involved can be made clearer if we illustrate them with some of his figures. For example, if we start with the year 1919 as our base period and represent it by 100, Copeland estimates that by 1925 the index of total transactions T

Table 21·5. Total Money in Circulation, M, and Total Bank Deposits in Commercial Banks, M′, at the Close of Each Year, 1936-1945*

Year	Money in circulation, billions	Deposits in commercial banks, billions
1936	$6.5	$43.3
1937	6.6	42.0
1938	6.9	43.6
1939	7.6	47.8
1940	8 7	54 3
1941	11.2	60 2
1942	15.4	77 8
1943	20.4	94.9
1944	25.3	115 8
1945	28.5	136.1

* Figures on deposits exclude interbank deposits Data are based on Federal Reserve Board releases.

had risen 32 per cent to 132; the index of money and other means of payments, lumping M and M' together, had risen 14 per cent to 114; and the index of velocity of circulation, lumping V and V' together, had risen 2 per cent to 102. What would one expect to happen to prices? The answer is that under these circumstances we would expect to have $MV = PT$, since we are covering both money and credit together in the terms of M and V. Or, of course, $P = MV/T$. But since M is 114, V is 102, and T is 132, we would have $P = (114)\ (102)/132 = 88$. Interpreted, this means about the following: The 14 per cent increase in the amount of money and the 2 per cent increase in the velocity of circulation would, by themselves, tend to raise prices; but they are offset by the large 32 per cent increase in the volume of transactions,

[1] Special Purpose Indexes for the Equation of Exchange for the United States, 1919-1927, COPELAND, MORRIS A., *Journal of the American Statistical Association*, vol. XXIV, no. 166, June, 1929, pp. 109-122.

and the net result is that we would expect prices to drop to about 88 per cent of their 1919 level. To check the result, Copeland gives the actual price figures showing that prices did drop in 1925 to 93 per cent of their 1919 level. In other words, the estimates are not completely accurate and the figures do not check out precisely, but they do give the approximate answer.

If we may illustrate once more, this time with a fictitious example, let us assume that right now we have the following situation:

M is 25 million dollars.
M' is 100 million dollars.
V is 10 times per year.
V' is 20 times per year.

Let us ask what will happen to the general price level if

M increases to 30 million dollars.
M' increases to 120 million dollars.
V decreases to 8 times per year.
V' remains at 20 times per year.
The volume of trade increases 25 per cent.

If we let the present volume of trade T be 100 per cent, we can write our present equation of exchange thus:

$$P = \frac{(25)(10) + (100)(20)}{100} = 22.5$$

After the changes that have been enumerated, the equation of exchange will become

$$P = \frac{(30)(8) + (120)(20)}{125} = 21.1$$

Prices will fall from 22.5 to 21.1, or they will fall to 94 per cent of their original amount. Our answer is, then, that we would expect, as a result of all these changes, that prices would drop 6 per cent, to 94 per cent of their original amount—or that the index number of prices after the changes would be 94 if we used as a base period the time just before the changes took place. In other words, our equation of exchange lets us make a rough estimate of the probable effect of various changes in the situation.

21 · 10. Price Movements and Real Income. It is always dangerous to compare amounts of money at different times on the assumption that "a dollar is always a dollar." If we were to discover that a given man had received a salary of $2500 per year in 1932 and that by 1945 his salary had been "raised" to $3000, we should be cautious about assuming that his income has increased by one-fifth, or even that it has increased at all. To be sure, his *money income* has increased 20 per cent; but, if we look back at the figures of Table 21 · 3, we shall discover that his real income has shown no such

increase. That table shows us that the index number of wholesale prices was 65 in 1932 and had risen to 106 in 1945. As we have learned earlier in this chapter, this indicates that prices in general were only 65 per cent as high in 1932 as they were in the base period, while in 1945 they had risen to 106 per cent of the base level. Prices had risen to 106/65 of their 1932 level, or had risen to 1.63 times their 1932 amount. This is an increase of 63 per cent and, unless our man's money income has also risen 63 per cent, he will find that his real income has fallen—that he is worse off than before and cannot buy so many goods and services with his $3000 in 1945 as he could with $2500 in 1932.

The economist usually tries to get around this difficulty by using what he calls "corrected dollars." When he uses this term, he means that he is making allowance for those changes which occur from time to time in the value of money and is estimating what would have occurred if money had maintained always the same value that it had in the base period. We can make this clear, perhaps, by going back to the hypothetical example of the preceding paragraph.

Our man was receiving an income of $2500 in 1932. Table 21 • 3 tells us that prices in 1932 were, on the average, but 65 per cent of their 1926 level. With these low prices our man could buy more in 1932 with $2500 than he could in 1926, when the prices were higher. How much more could he buy? The answer is simple. If each price was 65/100 as high (since 65/100 is 65 per cent), he could buy 100/65 as much. The $2500 of 1932 would buy as much as (100/65) ($2500) in 1926, or as much as could be bought with $3846 in 1926. This illustration gives us a simple rule: *To "correct" any given amount of money for changes in the price level, multiply it by* 100 *and divide the product by the index number.* In our case, we wanted to "correct" a value of $2500 for the year 1932; so we multipled $2500 by 100 and divided by 65, which was the 1932 index number.[1]

By 1945, our man was receiving an income of $3000. The 1945 index number according to Table 21 • 3 was 106. If we follow the rule again, we find that the corrected value for 1945 is (100/106) ($3000) or $2830. Let us now make this comparison: Our man received money income of $2500 in 1932 and $3000 in 1945; but he received "corrected" money income of $3846 in 1932 and $2830 in 1945. This means that the purchasing power of his income, after we have taken into account both his apparent increase in salary from $2500 to $3000 *and* the rising prices of the period, was actually the same as though his salary had been reduced from $3846 in 1932 to $2830 in 1945. The increase in money income had in fact turned out, because of

[1] Statisticians will understand that it is often difficult to know just what index number to use in correcting monetary figures; but, important as this matter is in practice, it cannot well be discussed here.

rising prices, to be a decrease in real income—in the purchasing power of his income. This man and his family were worse off with an income of $3000 in 1945 than they had been with $2500 in 1932.

Corrected dollars are apt to seem confusing to the student until he becomes accustomed to them; but they are really far less misleading than are the original uncorrected figures which exhibit a superficial precision far beyond their merit. It is the fact of changes in the general price level which forces us to correct our monetary figures. If we could get some stable monetary unit, which kept its purchasing power constant, we could forget about the problem. As it is, we shall have occasion from time to time to compare prices, incomes, costs, taxes, or other monetary figures at different periods of time, and the student must be prepared to offset the changes in the value of money by some appropriate corrections such as those just illustrated.

Questions

1. When we say that the general price level has risen, do we mean that the price of every commodity has risen? If not, what do we mean?

2. If the index number in 1945 is 150 and in 1946 it is 180, have prices risen or fallen? How much?

3. In the preceding question, has the value of money risen or fallen? How much? Explain.

4. What is the purpose of weighting the various commodities used in constructing an index number?

5. Suppose you are told that the index number of prices at present is 130. What does this mean?

6. If, in question 5, you were told that "1940 = 100," what would that mean?

7. What factors should be considered in selecting a base period for index numbers? Or is it a matter of indifference?

8. Using the simple method of Sec. 21 • 2, suppose that in 1950 the prices of the commodities are the following:

Milk	$ 0.20
Meat	0.80
Housing	25.00
Clothing	40.00
Toothpicks	0.10

What is the weighted index number of prices in that year on a 1945 base? Interpret your answer.

9. How much higher did prices go as a result of the first World War than they went as a result of the Civil War? (See Table 21 • 3.)

10. Explain how creditors are affected by rising prices.

11. If you had invested a very large sum of money in bonds, would you prefer to have the general price level fall or rise? Why? Would your answer be different if your investment had been in stocks? Why?

12. What is meant by inflation and deflation?

13. What is the relationship, if any, between the equation of exchange and the quantity theory of the value of money?

14. Why do we consider the effect of velocity of turnover instead of merely the effect of the quantity of money in trying to explain price movements?

15. Explain the meaning of the terms in the equation of exchange.

16. Some pieces of money turn over very slowly, remaining for many years in hidden hoards, while other pieces are spent several times on a given day. Since the velocity of turnover is very different for different pieces of money, what is it that is represented by V in the equation of exchange?

17. Do you understand that the M of the equation of exchange is the total money stock, or merely that part of it which is in circulation? If M and V both refer to the same money, will it make any difference whether it is the total stock or merely the circulating stock?

18. In what ways do you think that changes in the quantity of M would be likely to affect the quantity of M'? Why?

19. In what ways do you think that changes in the quantities of M and M' would be likely to affect the sizes of V and V'? Discuss.

20. It is much easier for a government to influence the sizes of M and M' than of V and V'. Suppose, however, that a government wished to influence the size of V or V'. What steps might it take? Discuss.

21. Would a 10 per cent change in M have a greater or a smaller effect on P than a 10 per cent change in M'? Discuss.

22. Suppose that a man's income fell from $2400 per year in 1868 to $1800 in 1875. Did his real income rise, fall, or stay the same? Base your answer on the data of Table 21 • 3.

23. If a man's money income rose from $2800 per year in 1940 to $3400 in 1945, what happened to his corrected income (that is, to his real income)? Explain. Use the data of Table 21 • 3.

24. Explain what is meant by corrected dollars. Possibly it will be easier if you use it in some such statement as, "John Doe's 1948 income amounts to $3500 as compared with his 1945 income of $3000, *both figures being given in corrected dollars.*"

22. THE BUSINESS CYCLE

22 · 1. Movements in Historical Data. Figure 21 · 1 shows the changes that have taken place in the general level of wholesale prices during a century and a half. We note that the line indicating price changes wiggles through many contortions and convolutions, yet we see a certain pattern or plan underlying the movements. They are not completely random and haphazard. If we were to plot on a similar diagram data showing the volume of output of American industries or the amount of immigration or the average monthly temperature in Kingston, R. I., we would find also that beneath the seemingly aimless movements there lay a plan or pattern. In fact, when we plot data showing changes in natural and social phenomena over a long period of time, we usually find not one, but several interwoven patterns.

One of the commonest sorts of movements in historical data is what we call the "secular trend," or sometimes just the trend. This is a general long-run tendency which continues for a considerable time. Thus in Fig. 21 · 1 we note that prices tended generally to fall from 1864 to 1896. There were interruptions in the movement—temporary countercurrents of rising prices which tended for a year or two to obscure the underlying tendency. But, as we inspect the chart, it is clear that these upward movements were temporary. When prices rose in 1880, and again in 1881 and 1882, people had no way of knowing that the upward movement was temporary. Perhaps they thought that there had been a fundamental change and that the long-time secular trend had been reversed. But, as we look back, we see that this upward movement, like several other less pronounced ones during the longer period, was a temporary aberration. Secular movements may be rising or *positive,* like the price movement from 1896 to 1920; or they may be falling or *negative,* like the price movement from 1864 to 1896. But any long-continued general tendency in numerical data is called a "secular movement" or a secular trend.

In contrast with secular movements, we often find in historical data movements that repeat themselves over and over again, going endlessly

through the same repetitive cycle. Thus, if we were to plot the daily average temperatures at Kingston, R. I., we would find rising temperatures in the spring and falling temperatures in the fall—summer highs and winter lows. Instead of a continued rise year after year, or a continued fall, we run through a *cyclical movement*—a more or less regular repetition. The regularity is not complete, to be sure. The hottest day is not always the same day of the year. The average July temperature is not the same each year that it was the year before. Yet the underlying similarity from one year to another is obvious. Cycles may be a year long, as is the temperature cycle. Many of the cycles of natural and social data do thus correspond with the calendar for reasons that are apparent. Such 12-month cycles are called "seasonal movements" to distinguish them from other cycles that are either longer or shorter than a year. For example, we realize as soon as our attention is called to it that there is another temperature cycle in addition to the annual or seasonal one. This is the daily or *diurnal* cycle. Temperatures tend to be higher at 2 P.M. than at 2 A.M. Although one day does not repeat the preceding one exactly, nevertheless the fundamental fact of approximate repetition is there, and we recognize it as a cycle. There are diurnal cycles in body temperature, in volume of business at a meat market, in volume of traffic on a state highway, and in hundreds of other things.

Just as there are many seasonal and diurnal (and some weekly) cycles, so there are cycles of other lengths not commensurable with the calendar at all. The astronomer recognizes cyclical movements in the amount of light received from Cepheid variable stars, or in the apparent positions of the planets. The physicist finds them in the swinging of a pendulum or in vibrating strings. The biologist discovers seasonal cycles in the growth of plants or longer cycles in the number of rabbits living in a given territory. In every science, both secular and cyclical movements appear. In addition, there are almost always some movements that do not seem to be connected with time at all: they have no time pattern or plan. These we call "random movements." We do not intend to imply that these movements have no cause, or even that their cause is unknown, but merely that their cause is not a time-connected one.

22 · 2. Business Cycles. Cycles, then are not unusual. They are common and are found on every hand. We almost expect to find rhythmic repetitive movements in historical data. Some of them show their high points or peaks in summer, some in winter, and some in spring or fall. For example, Fig. 22 · 1 shows the annual cycles in the output of ice cream and of whisky in the United States. In each case the cyclical nature of the movement is pronounced, but the cycles are almost opposites in their timing. When ice cream output is high, whisky output is low. There is nothing surprising about cycles. What is surprising is that now and again these diverse cycles

somehow seem to get synchronized, so that vast and largely unrelated parts of our economic structure get to swinging up or down in unison. Thus we get periods in which prices, employment, and output are unusually high in almost every industry—or unusually low.

These cycles are very old. As far back as we have historical records we find mention of "fat" and "lean" years; of prosperous and of depressed times. The downswings seem to have been associated usually with factors affecting the quantity of the productive factors available: pestilence, famine,

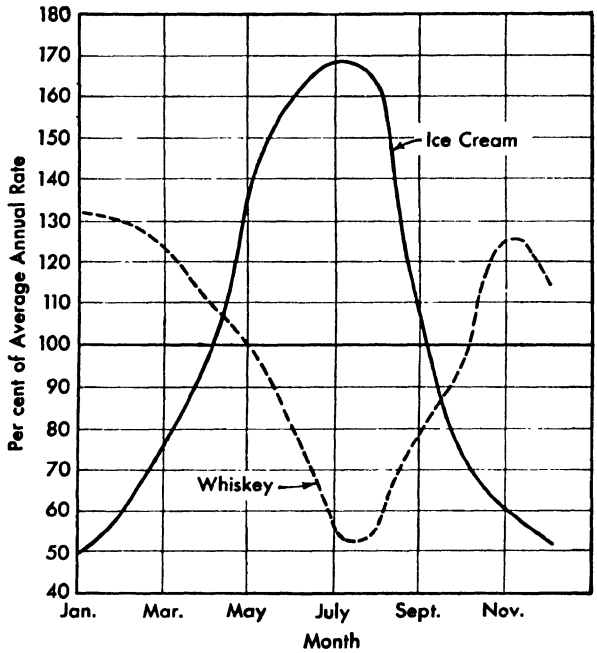


FIG. 22 · 1.—Seasonal variation in output of ice cream and whisky.

fire, earthquake, war, etc. But there were other variations, due not to changes in the amount of the productive factors but to their productivity. We must not let these long cyclical swings in productivity obscure the secular movement upward—the slow and painful progress that men have made in their efficiency since the dawn of time—but we know that there have been long interruptions to this general upward movement. There have been periods of decline and dark ages, when men seem to have slipped back toward the jungle, when the arts and skills and techniques of men's forefathers were lost and the productive factors became inefficient. These have been followed by periods of reawakening, of enlightenment, of renaissance. Thus history records several long cyclical movements in productivity, in addition to the

common swings in productivity that have accompanied changes in the weather, etc.

These old cycles, expressing themselves in changes in the quantity of the factors or in the productivity of the factors, affected all sorts of businesses. They were, to be sure, cycles in business. But they were not business cycles as we use that term technically today. The modern *business cycle* is not so much a change in the quantity of the factors of production, nor in their productivity, as it is a change in the degree of their utilization. In the depths of the great depression of the thirties there were as many men available for work as ever, and the techniques of production were as efficient as ever; but men, machines, and land were idle. The factors of production were incompletely utilized. In the modern business cycle we find changes not only in output but also in employment in the broadest sense—in the employment of all the productive agents. It is in this way that the modern business cycle differs from the old-fashioned cycles of earlier history. When the economist uses the term "business cycle," then, he is not referring to every cyclical movement, or to every cyclical movement in business, or even to every cyclical movement that affects large segments of the economy at once; but to cyclical movements that are evidenced by great swings both in output of industry and in employment of the factors of production.

Statistical investigation seems to indicate that there are three business cycles of different length: what we might call the "long cycles," lasting some 50 years or so; what we might call the "short cycles," lasting a little over 3 years (some students put the time at 40 months); and the "business cycles" proper, lasting approximately 8 to 10 years. The long cycles have been ascribed to variations in the output of gold or other similar monetary causes, and to technical progress and invention and the consequent investment opportunities. These cycles are so long that we do not have adequate data for describing more than two or three of them—not enough data even to be sure that they are cyclical in character. The short cycles seem much more pronounced in the United States than in any other country, present but less pronounced in Britain, and entirely lacking in most countries. They appear more in price data than in production data and are supposed by some authorities to be the result of changes in investment in inventories. Other writers have suggested that they are purely chance phenomena, and have shown that certain types of pure chance data tend to set up cyclical movements. We shall, however, confine our attention to the cycles of intermediate length—those which alone are nowadays referred to when men speak of business cycles.

22·3. Historical Background. Authorities differ as to the dates that they ascribe to business cycles, and even in some cases as to whether or not there were two cycles or one in a given period. If, however, we start with the most

commonly recognized cycles in the United States since the War of 1812, listing the times of the "panics" or "crises" when business fell into the downswing part of the cycle, we might list the years 1819, 1825, 1837, 1847, 1857, 1865, 1873, 1884, 1893, 1903, 1907, 1913, 1920, and 1929. This indicates an average length of cycle of about 8½ years, although the durations were 6, 12, 10, 10, 8, 8, 11, 9, 10, 4, 6, 7, and 9 years. The length of time elapsing between a given point in one cycle and the corresponding point in the next cycle (for example between succeeding high points or between succeeding low points) is called the "period" of the cycle. The height of the wave, from trough to crest, is called the "amplitude" of the cycle (see Fig. 22 • 2). The various business cycles have differed greatly in length, from a minimum of 4 to a maximum of 12 years. They have also varied tremendously in amplitude. Some of the dates mentioned mark minor recessions in business, while others mark major and severe depressions. The

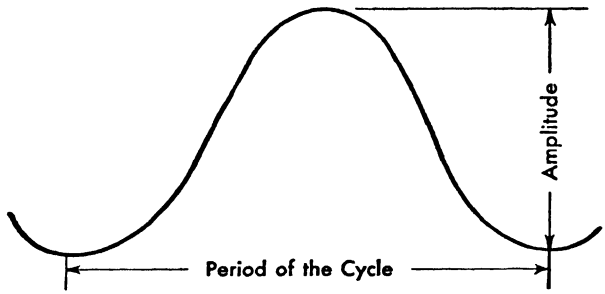


FIG. 22 • 2—Period and amplitude of a cycle.

student must understand that business cycles are not characterized by the same degree of regularity that characterizes, for example, the swings of daily temperature through the season. Although the temperatures of no 2 years are exactly alike, nevertheless the similarities are very great indeed. The longest business cycle in our list is three times as long as the shortest one; the length of time from the hottest day in one year to the hottest day in the next varies, but only by a small percentage. *Business cycles are more alike in the sequence of their events than in the timing of them.*

Of the "crises" listed in the preceding paragraph, the most violent were those of 1819, 1837, 1857, 1873, and 1929. It will be noted that these were coming at intervals of roughly 20 years at first, but that there was an interval of well over a half century between the last two. This is explained by many students as a result of the interaction of the business cycle and the long cycle mentioned in the preceding section. When the downswing of the business cycle occurs during a downswing of the long cycle, the two reinforce each other, and the resulting depressions are unusually long and severe. When the downswing of the business cycle occurs during an upswing of the

long cycle, the depression is apt to be short and of small consequence.

The Panic of 1819 was largely a reaction to the Napoleonic wars. There was an abrupt and catastrophic break in commodity prices, with the cotton industry being especially hard hit. There were many business failures, and foreign trade was sharply reduced. Unemployment became widespread. The agricultural regions of the Central states were particularly plagued. In the East the worst was over by 1821, but distress continued in the agricultural West until 1823. This is said by many students to have been the first international cycle.

The Panic of 1837 seems to have been caused in large part by land speculation in the United States, combined in some measure with overexpansion of investment in canals and railroads. Some students attribute it, at least in part, to the rapid expansion of the state banks, aggravated by the transfer of funds from the United States Bank to the state banks. In May, 1837, the banks of New York were forced to suspend payments, and banks throughout the rest of the country quickly followed suit. Commodity prices declined severely, and foreign trade fell off. Banks failed in large numbers and again unemployment became marked. This time the depression hung on longer, through the following 5 years, and only in 1843 did a slow improvement set in.

The Panic of 1857 was sharp but relatively short-lived. Preceded by railroad expansion and land speculation, the actual crisis was precipitated by the failure of the Ohio Life Insurance and Trust Company. The ensuing difficulties were confined largely to the financial markets, and the industrial East suffered far more than the agricultural West. By 1859, the storm had blown over.

The collapse of 1873 was the deepest and most intense of the nineteenth century and led to the deepest and longest depression. It was international in character, and, like that following 1929, combined depressions in agriculture and industry. Business had been uneasy for a year or more, largely as a result of the postwar effects of Civil War financing. But it was the period of most rapid railroad expansion in history, and men were busy and money was plentiful. The sudden failure of Jay Cooke and Company, which had become involved in unwise railroad financing, was followed by the closing of banks and brokerage houses throughout the east, and finally by the closing of the New York Stock Exchange. There was a short but very severe money panic, which was over before the end of the year. It was followed, however, by a prolonged business depression as prices and business finally dropped to their prewar levels. The drastic and continued fall in prices put debtors in an impossible position (see Sec. 21·6), and failures continued month after month and year after year. Instead of showing signs of improvement, the situation grew slowly worse and worse as the depression intensified,

and the bottom was not reached until perhaps 1878. Not until 1880 were there real signs of recovery.

22·4. **More Recent Developments.** The Panic of 1929 followed one of the most outstanding periods of expansion and prosperity in history—a period in which the rate of output and the average level of living in the United States reached record highs. The ensuing depression was one of the worst in recorded experience, just as the prosperity had been the greatest. This depression was also world-wide in scope. In the United States, which was perhaps the hardest hit of all countries, there seems to have been little recognition of the fact that trouble was brewing. Prices had held unusually stable after the postwar break of 1920. Only when the bottom dropped out of the stock market on October 29, 1929, did people suddenly begin to realize that something was wrong. Within a period of 2 or 3 weeks, the values of listed stocks had fallen over 50 billion dollars—two-thirds of the annual income of the people of the United States! By 1932, the national income had been cut in half, and more than a quarter of the gainfully employed people of the country had been thrown out of work. Industrial output fell off more than 45 per cent, and the general price level fell a third. Building and the accumulation of capital almost ceased. Savings were depleted or entirely dissipated. Farm product prices fell by 60 per cent. The number of unemployed grew to some 13 or 15 million. Interest rates fell, and capitalists found no place in which they could make profitable investments. Finally, after 3 or 4 years of bitter depression, the country passed through a severe banking crisis. In the 4 years following 1929, over 9,000 banks failed in the United States, and the Federal Reserve ratio dropped almost to its legal minimum despite the addition of over 1½ billion dollars to the circulating money stock. This was the first major depression in which governments took an active part. Always before it had been assumed that men's business decisions were primarily their own affairs, but the suffering and despair of the continuing depression led governments to take a hand and led their citizens not only to allow but encourage government interference. There is still strong difference of opinion as to whether or not this government assistance or interference (some people call it one, and some the other) was of any major help. We must remember that society has weathered severe depressions before and has returned to prosperity without government aid, and presumably recovery would have taken place spontaneously again. Those who feel that government aid was wise believe that the process was hastened and the suffering reduced by government action; those who feel that government interference was uncalled for believe that recovery would have taken place perhaps even more quickly and without centralization of power if the government had kept

hands off. As in so many social questions, there is no way to prove either contention.

It is difficult to know how to measure or depict the changes in the economic situation. Some people look upon the price data of Fig. 21 • 1 as indicating fairly well the general conditions. Other people use such indexes as the number of freight cars loaded, the volume of clearings at the major clearing houses, the amount of deposits subject to check, the unfilled orders of the United States Steel Company, the amount of new building construction, the volume of factory pay rolls, the amount of department-store

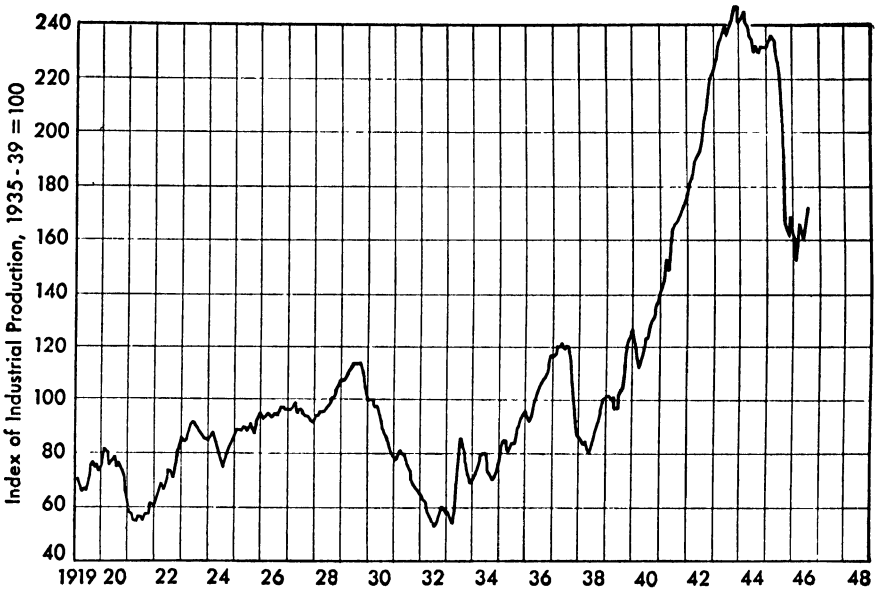


FIG. 22 • 3.—Index of industrial production in the United States, 1919-1946.

sales, or the volume of employment. Each of these shows some aspect of the economic picture. One of the most commonly used indicators is the index of industrial production, published monthly by the Federal Reserve Board. The monthly figures since 1919 appear in Table 22 • 1 and in Fig. 22 • 3. This series unfortunately does not extend back far enough to give us a long-run view of business cycles in the country; it does show well the industrial picture since the close of the World War. Especially noticeable is the tremendous increase in industrial output that accompanied the second World War. In interpreting these data, the student should remember that they cover industrial production only and take no direct account of the agricultural situation or of conditions in transportation, the professions, etc.

22·5. A Typical Business Cycle. As we have said, the most characteristic thing about a business cycle is the sequence of events that make it up. Some-

Table 22·1. Index of Industrial Production in the United States, 1919-1947. Seasonal Movements Have Been Eliminated* (1935-1939 average = 100)

Year	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Annual average
1919	71	68	66	67	67	72	76	77	75	75	74	75	72
1920	82	82	81	76	78	79	76	77	74	72	66	62	75
1921	58	57	55	55	57	57	56	58	58	62	61	61	58
1922	63	66	69	67	70	74	74	72	76	81	84	86	73
1923	85	86	89	91	92	91	90	88	87	86	86	85	88
1924	86	88	86	83	80	77	75	78	81	82	84	87	82
1925	89	89	89	90	90	89	91	90	88	92	94	95	90
1926	93	94	95	94	94	95	95	97	98	98	97	97	96
1927	97	97	99	96	97	97	95	95	94	92	92	93	95
1928	95	95	96	96	97	98	99	101	102	104	106	107	99
1929	108	108	109	110	112	114	114	114	113	110	105	100	110
1930	100	100	98	98	96	93	89	87	85	83	81	79	91
1931	78	79	81	80	80	78	76	74	70	68	67	66	75
1932	65	63	62	58	56	54	53	54	58	60	60	58	58
1933	58	57	54	58	68	78	86	82	77	73	69	70	69
1934	72	75	79	80	80	79	73	72	70	71	72	77	75
1935	83	85	85	82	82	84	84	87	89	93	94	96	87
1936	95	92	94	99	101	103	105	107	108	109	113	116	103
1937	116	117	120	120	121	119	120	120	115	107	95	87	113
1938	86	84	84	82	80	81	86	90	92	95	100	101	89
1939	102	101	101	97	97	102	104	104	113	121	124	126	109
1940	122	116	113	114	118	123	123	126	129	132	136	140	125
1941	143	147	152	149	160	164	166	167	169	172	174	176	162
1942	181	183	186	189	191	193	197	204	208	215	220	223	199
1943	227	232	235	237	239	237	240	242	244	247	247	241	239
1944	243	244	241	239	236	235	230	232	230	232	232	232	235
1945	234	236	235	230	226	220	210	186	167	162	168	163	203
1946	160	152	168	165	159	170	172	177	179	181	183	182	170
1947	188	188											

* Data from the *Federal Reserve Bulletin*, August, 1940, and scattering thereafter.

times these events follow each other rapidly, and the cycle is a short one; sometimes the process is long drawn out, and the cycle is a long one. To be sure, the events do not always happen in exactly the same order; yet the similarity is close enough to give a real cyclical character to the phenomena.

When we talk of the business cycle, we are thinking of rhythmical changes in the rate of economic activity—in employment, output, wage rates, etc. Figure 22 • 4 gives a rough preliminary picture of the situation. If the broken horizontal line that crosses the middle of the diagram represents normal or average rates of activity, the business cycle swings sometimes above and sometimes below it. It is customary to divide the cycle into two main divisions, and then to divide each of these two further into two subdivisions, making four stages of the business cycle altogether. When economic activity is increasing, as shown by a rising curve, we have what we call a period of "expansion." If we start at the bottom, this expansion may continue for some time before activity is even up to normal. The part of the expansion

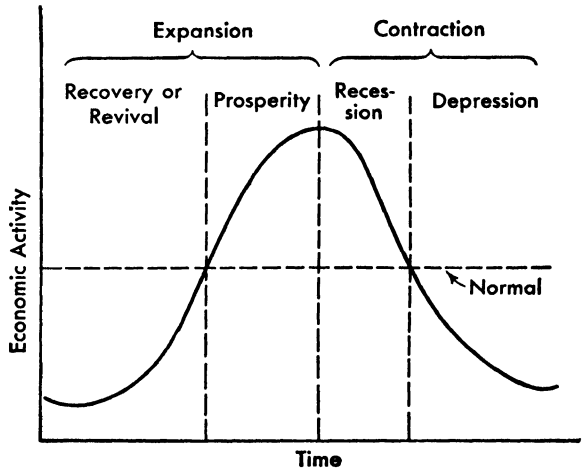


FIG. 22 • 4.—An idealized business cycle.

that lies between the low point and normal activity we call the period of "recovery" or "revival." Experience shows us that expanding economic activity does not ordinarily stop when it reaches a normal level. It continues on above normal. The period of continued expansion above normal is called "prosperity."

The long downswing of the cycle, when economic activity is slowing down, is called the period of "contraction." Just as expansion runs from the trough of the wave up to the crest, so contraction runs from the crest of the wave down to the trough again. The first part of this contraction is merely a return to normal activity. When activity is above normal but contracting, we say that we are in a period of "recession" or "crisis" or "panic." The latter term is correctly applied only if the period is a short and sudden one—if the fall of the curve is very steep. Again we learn from experience that the contracting movement does not typically stop when

normal is reached. Contraction continues and activity becomes subnormal. The part of the period of contraction that continues after a normal rate of activity has been reached and passed is called "depression." Thus the two major parts of the cycle are expansion and contraction, and they are subdivided into recovery, prosperity, recession, and depression. If we distinguish these terms carefully, we shall use them as they are depicted in Fig. 22·4, or in accordance with the following classification:

1. Expansion. Economic activity is increasing.
   a. Recovery or revival. Activity increasing but still below normal.
   b. Prosperity. Activity increasing and above normal.
2. Contraction. Economic activity is falling.
   a. Recession or crisis or panic. Activity decreasing but still above normal. The word panic should be used only for a short, sharp, sudden recession.
   b. Depression. Activity decreasing and below normal.

It is quite obvious that, when we are close to the turning points, it may be hard to determine whether we are in the period of depression or of recovery, whether we are in the period of prosperity or recession. Sometimes the change is sudden, striking, and dramatic, so that no one can mistake it, as in the case of a real panic. Thus Sept. 24, 1869, is still known as "Black Friday" from the sudden series of business failures and catastrophic losses which followed the cornering of the gold supply by Jim Fisk and Jay Gould; and the break in the stock market on Oct. 29, 1929, left no one in doubt as to the turn of events. Sometimes, on the contrary, there is such a slow imperceptible change from prosperity to recession, or from depression to recovery, that students cannot tell even afterward just when the change occurred. Yet we do find it useful, in describing the business cycle, to break it more or less arbitrarily into these four divisions, and we shall now turn to a more minute description of the events of each of them.

22·6. Prosperity. The period of prosperity is characterized by high employment, high productivity, and persistent optimism. Wages are high, and we have what is called "full employment." Anyone who is willing and able to work can find a job. The spirit of optimism leads businessmen to expand and to borrow for that purpose; and the same spirit of optimism encourages lenders to lend them the money for such expansion. Thus, debts are usually rather large, and as the period progresses the interest rate tends to rise because of the large borrowing. As wages and prices rise, businessmen find their costs rising and tending to catch up slowly with their revenues. Thus, toward the end of the period profits, which had been high, are tending to shrink somewhat. As the banks lend more and more, their reserves tend to become low, and the optimism breeds growing speculation. Commerce and trade are extensive and, as a result, bank clearings are large. Business failures

are very few. Inventories are usually large, and there are heavy investments in durable capital goods.

22·7. Recession. It should not be supposed that any one of the stages of the business cycle is uniform from beginning to end, culminating in a sudden change to the next stage and then no change at all until the following stage is reached. On the contrary, slow changes are taking place throughout each of the stages. As the period of prosperity draws toward its close, the higher costs are tending to reduce profits, and at least some businessmen and financiers are beginning to worry about the overextension of credit, the low bank reserves, or the fact that the better opportunities for investment have been taken up and only more risky ones remain. A feeling of hesitation begins to replace the confident optimism, and lenders start to demand more security. Sometimes, things are precipitated by an important and spectacular business failure, as when the failure of Jay Cooke and Company in September, 1873, marked the beginning of the sudden panic and the subsequent long depression. Be that as it may, in the period of recession the feeling of optimism gives way to a feeling of hesitation, doubt, and ultimately fear. Businessmen try to protect themselves by selling property in order to raise money to get out of debt. These forced sales, in increasing quantities, tend to cause sharp reductions in prices and, as prices fall lower and lower, businessmen are forced to sell more and more in order to meet their debts. Credit is suddenly sharply contracted. Business expansion stops. Orders are canceled and workmen laid off. Where men had before taken profits confidently for granted, they are now filled with a haunting fear of loss—a fear that is all too often realized as dropping prices reduce profit margins to the vanishing point. Business failures increase, investment stops, unemployment rises sharply. Money incomes are reduced both by unemployment and by reduced wage rates for those who still work. Prices tend to fall even more rapidly than costs, thus tending to reduce profits even more. Bank credit is contracted. There is a tendency for men to convert their holdings into cash, which they hoard. Bank reserves increase as the banks call their loans or refuse to renew old loans and make new ones. Discount rates ordinarily rise as bankers fear for the safety of their funds. Bank clearings drop as business declines. Optimism has by now given way to fear—"nameless, unreasoning, unjustified terror which paralyzes needed efforts to convert retreat into advance."

22·8. Depression. The period of depression is well named. It is a time of pessimism. The same people who, in prosperity, had supposed that they were living in a "new era" and that Utopia had finally been reached, now forget that depressions have come and gone often in the past and assume that they are experiencing something new as well as something awful and terrible and uncontrollable. As the depression wears on, the feeling becomes one of

numbed and calloused resignation or of unreasoned and frenzied rebellion. Debts have been liquidated, and men have lost what they can lose. They have been "through the wringer." Business is stagnant, and unemployment is large. The low wages and large unemployment reduce incomes to very low levels; and the low incomes reduce purchasing power, thus reducing the demands for goods. The poor prospects for sales tend to encourage businessmen to reduce output, lay off more men, and reduce wages even further; and thus the deadly cycle of reduced wages and falling demand continues. The level of production is cut far below normal, and businessmen work on a day-to-day basis, ordering only what is necessary to meet current needs. Investment in capital goods practically ceases. Bank reserves are at a very high point, since businessmen are afraid to borrow and bankers are afraid to lend. Discount rates tend to be rather low as bank reserves pile up. Bank clearings are low, and business profits are small. Inventories of goods get lower and lower as men hesitate to produce in advance of orders. The bankruptcy rate is high as the less efficient firms are eliminated. As wages and prices drop, business costs reach new low levels. Hoarding continues and tends to increase.

22·9. Recovery. With costs low, inventories low, and with many unfilled demands, with indebtedness low and with bank reserves high, the stage is set toward the end of the depression for business revival. The more courageous and imaginative businessmen begin to see profit opportunities. Sometimes an outside force appears to give the needed stimulus: a new invention, unusually profitable crops from a freak of the weather, or a war. Sometimes, the change can be explained only by the fact that men have to live, and to live they have to consume, and to consume other men have to produce. Be the cause what it may, business slowly picks up again as employers take advantage of the low wages and the low costs to produce for the backlog of unfilled demand. Employment slowly rises, and with it incomes rise. At first, it is possible to hire more workers without raising the wage rate, since there are many persons involuntarily unemployed. But, as the slack is taken up, it ultimately becomes possible to hire more men only if wage rates are raised, and now both the increased employment and the higher wage rates tend to increase buying power and stimulate demand. With bank reserves high and discount rates low, businessmen are encouraged to expand their operations. Cash balances are reduced as people stop hoarding and put their money to work. As prices of finished goods start up, costs tend to lag behind as employers take up the slack of involuntary unemployment and unused plant capacity. Men start to hope again, and their expectations of profit rise higher and higher. Industrial output rises as bank reserves slowly decrease and bank clearings increase. Business failures become uncommon, and businessmen begin to invest in plant and equipment to replace

worn-out or obsolete capital. Realizing that prices are rising, businessmen tend to stock up on materials in advance, and thus the extra demand as inventories rise raises prices even more rapidly. And so the economy moves back toward the period of prosperity and another repetition of the cycle. Although it need not be true that every cycle contain every one of these characteristics in exactly the order here named, nevertheless the sequence

Table 22·2. Summary of Characteristics of the Various Stages of the Business Cycle

Characteristics	Prosperity	Recession	Depression	Revival
1. Employment..	High	Sudden drop	Low	Slow rise
2. Industrial output.....	High	Decreasing	Low	Slow increase
3. Wage rates...	High, but lag behind prices	Falling, but lag behind prices	Low	Rise, but lag behind prices
4. Prices.....	High	Rapid drop	Low	Slow rise
5. Bank loans...	High	Cut sharply	Low	Expanding
6. Bank reserves.	Getting low	Sudden rise	High	Slow drop
7. Bank clearings	Large	Sudden drop	Small	Gradual rise
8. Discount rates	Rising toward end	High	Low	Fairly low
9. Costs of production...	High	Falling	Low	Rising
10. Profits	High but falling	Disappear	Appear toward end	Rise
11. Business failures	Few	Sudden rise	Many	Few
12. Speculation...	Widespread	Little	Very little	Increasing
13. Business inventories...	High	Falling	Low	Rising slowly
14. Building construction..	Much	Sudden stop	Almost solely public works	Slow rise
15 State of feeling	Optimism	Hesitation and fear	Pessimism, despair	Cautious hope

is one that has become familiar from the continued repetition of its major elements.

22·10. Summary of Characteristics of the Business Cycle. Table 22·2 summarizes the conditions that exist in the various stages of the business cycle. The student should remember that there is usually a slow change from one stage to the next, and that in a summary table one is forced to enter short catch phrases that cannot give a complete picture. The table should not be used alone, but in conjunction with Secs. 22·5 to 22·9 which describe the occurrences in greater detail.

22·11. Causes of the Business Cycle. When the economist comes to state the cause of the business cycle, he is not embarrassed by lack of explanations; on the contrary, he is embarrassed at the large number of different explanations that are offered. Everyone seems to have an explanation of his own, and no single explanation gathers anything like a majority of support. We shall content ourselves here with stating what some of the suggested causes are and with noting some similarities and some differences, without suggesting categorically that any one of them is the right one.[1]

We may well start by dividing the theories into two main groups. In the first place, there are many students who feel that the causes of the business cycle are *external causes,* lying outside the economic system itself. These students feel that noneconomic forces impinge repeatedly on the economic system, setting up cyclical movements just as a man standing outside a bathtub with a paddle can set up waves within the tub by application of force at the right time. On the other hand, there are many other students, equally competent, who believe that the causes lie within the bathtub itself—that the economic system is so made up that economic forces themselves give rise to pulsations of a rhythmical nature. These students believe that the business cycle is the result of *internal causes,* which are part and parcel of the economic system itself.

Some of the earliest and most interesting of the explanations of the business cycle attributed it to external causes. Sir William Herschel, English organist and brilliant astronomer, suspected that sunspots affected the weather. W. Stanley Jevons, English logician, economist, and statistician, showed that the sunspots were cyclical in nature and that their cycle corresponded in length with the business cycle. He reasoned that sunspots affected weather, weather affected crops, and the bounty of the harvest affected business. In this way it seemed reasonable to him that sunspots caused business cycles—that the cause was astronomical. In more recent years, one of America's most brilliant mathematical and statistical economists, Henry L. Moore, pointed out that there is a marked cycle in rainfall in the major agricultural regions of the United States. This rainfall cycle was, in his opinion, the cause of the business cycle. Moore also believed that there was some connection between the rainfall cycles and the planet Venus, which every 8 years passes between the earth and the sun. Thus we have at least two "explanations" of the business cycle which look at it as a *natural*

[1] When it comes to explaining what causes the business cycle, the author is constrained to quote Boltzmann's paraphrase of Goethe's "Faust,"

"So soll ich denn mit saurem Schweiss
Euch lehren, was ich selbst nicht weiss!"

True, this quotation is an apt one, applicable at many points in the field of economics, but nowhere more so than in business-cycle theory, where so many people, differing among themselves, are each sure that he is right.

phenomenon, based on astronomical causes in just the same way as the annual cycle of temperature. Both of these explanations associate the external force with the economic system through its influence on crops and agriculture. One would expect, if this were true, that business cycles would be less important in nonagricultural areas and that, as a nation became more completely industrialized, the business cycle would disappear or at least diminish in importance. Quite the contrary seems, however, to be true.

Some men have thought that the causes of the business cycle were external but casual happenings; such occasional and haphazard affairs as earthquakes, hurricanes, a chance very good or very bad harvest, the discovery of new territory or new resources, and the like. These people would look for a different but external cause for each cycle. Under such an explanation, it is difficult to see why the resulting movements would be rhythmical, although in fairness one must concede that the order of events and not their duration constitutes the business cycle.

When we turn to the internal explanations, we find that they are far more numerous. Some students have thought that the business cycle is the result of maldistribution of wealth—that our present organization of economic society is such that wealth is bound to be distributed unequally, giving more buying power to some and less to others than they can reasonably use, and thus causing maladjustments within the economic machine. They feel that the flow of income tends to be irregular, piling up temporarily in certain segments of the economic system and running short in others. Other students feel that the cause is that any system using capital[1] finds it necessary to foresee human wants in advance—to become to some extent a prophet.[2] Because men are not able to foresee their own wants accurately, they are constantly making errors in judgment. They invest in an industry, expecting that the product will be wanted and later find that it is not in sufficient demand for them to recover their investment; on the other hand, they may suddenly find that there are areas of demand that cannot be filled because no advanced preparations were made. These maladjustments continue and pile up until finally they become serious enough to force a depression.

Monetary economists are likely to see the causes of business cycles as essentially monetary—perhaps the result of defects in the monetary or banking system, or in the flow of gold or other sorts of money. Experts in international trade and foreign relations ascribe them to various restrictions on free trade among the nations of the world, to tariffs, trade quotas, and nationalistic efforts for advantage. Some authorities see the cycle as caused

[1] This includes not only what we call "capitalistic" systems, but all systems in which capital is an important element in production.

[2] This point was discussed in Sec. 13 • 5.

primarily by price movements, especially by the tendency, as they see it, for certain prices to get "out of line" with others, upsetting the normal "price structure." In recent years, it has been the style to explain the business cycle in terms of divergences between savings and investments, or "oversaving"; or on the assumption that the interest rate is higher or lower than the actual earning power of capital. Some people have argued in terms of "overproduction" or "underconsumption," or the balance and lack of balance between production and consumption. One of the current eminent authorities ascribes the business cycle to innovations and inventions, showing at considerable length that there is a good reason to suppose that such inventions will themselves not be made uniformly but by fits and starts. Marxists have tended to blame the business cycle on the fact that laborers do not receive the full value of the product of their labors, but that a "surplus value" is retained by the employer, thus forcing maladjustment of buying power and of consumption. On the other hand, some students favor a more eclectic view, explaining one cycle in terms of one of these causes and another in terms of another. They feel that there is no one cause of all cycles—no acceptable monistic explanation; so they choose for each individual cycle the special explanation that seems best in the light of the conditions of the time.

If the student is confused to find so many "explanations," it is only fair to say (1) that the proponents of each of these doctrines have accumulated a large amount of both empirical and theoretical material in support of their belief, and (2) that the problems involved are of so complex a nature that there will evidently be a great deal of difficulty in developing human understanding to the point where men will fully understand each other in the field. The problem is of such striking and immediate importance, however, that one may be sure that wise men will continue to make the attempt at understanding and explanation. At the best is must be admitted now that, no matter how certain individuals are that they know the whole story, there is no explanation that is accepted by any considerable proportion of competent students in the field. If the student of an elementary course is confused at the wealth of supposed explanations, it may make him somewhat easier to know that the experts suffer from the same confusion.

22 · 12. Proposed Remedies for the Business Cycle. One is naturally in a much better position to find cures for things if he knows their causes than if he does not; but, if the student is surprised to find us discussing remedies for the business cycle when we have but just finished saying that the causes are not known, he should remember that quinine was known to be useful in treating malaria many years before it was known that malaria was caused by a protozoan organism in the blood or that it was transmitted by certain species of mosquitoes. Sometimes the best that we can do is to treat symptoms, pending the discovery of the cause of the disease. There-

fore, no apology should be necessary for discussing proposed remedies for the business cycle despite the unsatisfactory state of knowledge concerning its cause. To be sure, if and when the cause is determined, it should be easier to find a cure.

First of all, we should ask whether or not one should try to eliminate the business cycle. Is it a natural phenomenon, like the seasonal swings in temperature, about which we could do nothing even if we wished? Of course, if it is a natural phenomenon, caused by sunspots or transits of Venus, there is little we can do to stop it; even then, we might well wish to treat the symptoms and prepare for what is coming. The fact that the seasonal swing of temperatures is beyond human control does not stop men from heating their homes or going to Florida and California in winter, or going to the mountains and the beach, or using air conditioning in the summer. The costs of a depression are incalculable. Not only is there the monetary and physical cost of goods that are not produced, but there are even more important spiritual costs that accompany unemployment and reduced income. Businessmen are forced into bankruptcy through no fault of their own, farmers lose their farms, crime increases, working conditions deteriorate, and life savings are wiped out. Even in what we call "prosperity" there is suffering on the part of people with fixed incomes who find it harder and harder to get along as the price level soars and as the buying power of their income dwindles. It is perhaps unfortunate that the names "prosperity" and "depression" have been given to these two periods, giving the impression that one of them is desirable and the other undesirable. Some people gain and some people lose in either period. It is the shift from the one condition to the other that brings undeserved gain to some and undeserved loss to others. The economic and spiritual damage of the business cycle is largely a result of the changes, not of the position of the curves themselves at any moment. This damage is so great that there seems to be unanimous consent that something should be done about it if we can find anything effective to do; and, when the situation is at its worst, there are always many people who are willing to jump at any suggested "remedy," no matter how fantastic, on the assumption that at least it cannot make thinks any worse. Many of the proposed remedies have been things of this kind—ideas born more of desperation than of cerebration.

Since we do not know for certain what the causes of business cycles are, we are driven largely to treating the symptoms. We do know pretty well what these symptoms are and have summarized them in Table 21 • 7. For example, we see that bank loans are high in prosperity and low in depression. It is commonly suggested that we should adopt policies that will discourage lending after recovery had passed a "normal" stage of activity, and should encourage it when recession has fallen below "normal." We have noted that

the Federal Reserve Board is given certain powers over reserve requirements, rediscount rates, and open-market operations to this end.[1] Many students think that, when a business boom gets under way and there are signs that business is overexpanding, these controls should be used to cut off the expansion before it gets out of hand—to stop the spree before there is danger of a hangover. As a matter of practical politics, this is very, very difficult to do. The political pressures against it are enormous as businessmen and workers object to policies that curtail business expansion. At this stage of the business cycle, everyone is a "rugged individualist," loudly proclaiming his right to make his own decisions and his willingness to take the consequences. In the period of depression when bank loans are low, it is advocated that the powers of the banking authorities be used to encourage extension of credit. When the depression has continued long enough, it seems easy to convince people to give over altogether too much power to the government, and the difficulty is then to restrict government interference from going too far. The erstwhile individualist who wanted to take the consequences of his own decisions now blames the government for indecision and inaction, and throws the rascals out to make way for other groups who are certain that they ought to do something but not at all certain what it is. We have already noted the fact that controls over credit and banking seem to be more effective in curbing expansion than in stopping contraction. Yet these controls are among the weapons most commonly relied on for eliminating or diminishing the business cycle.

Another school of business-cycle theorists look at the price symptom rather than at the bank-loan symptom. They note the changes in the general price level that accompany the business cycle and think that, if these price changes could be eliminated, the situation could be kept in hand. We have already seen that great groups of people are hurt whenever prices rise or fall and, if we could eliminate the price movements, we would at least have overcome this part of the difficulty. Some persons have advocated the leveling of prices by government subsidy in times of depression, or by a program in which the government would buy up large stocks in times of plenty, thus raising the price of goods at that time, and sell them again at times of scarcity, thus lowering the prices at the latter time. Experience seems to indicate that such programs cannot be successful unless the government also takes over control of the amounts produced, of foreign trade, and of the distribution of surpluses. In other words, such methods of control are not compatible with the continuation of a free economy.

Other students who advocate control of prices have suggested that it be done by manipulation of the monetary system. Most noted of these suggestions, and one which has had a great deal of support, was made by

[1] See Secs. 20 • 8 through 20 • 11.

Professor Irving Fisher. This calls for a "commodity dollar" or a "compensated dollar," or (as its opponents call it) a "rubber dollar." Fisher points out that, when the country was on a gold standard, we had a dollar of fixed weight but of variable purchasing power. The weight was 25.8 grains of gold 9/10 fine. This was a dollar and for many years remained a dollar. Its weight was constant, but its purchasing power varied over a wide range.[1] Fisher has advocated a dollar that would so vary in weight as to maintain a constant purchasing power. If the index number of prices showed that prices were beginning to rise a little, we would know that the value of money was falling and would increase by a small amount the quantity of gold in the dollar—increasing the weight just enough so that the dollar would retain its former value. If, on the other hand, our price index showed that prices were beginning to fall, we would know that the dollar was beginning to increase in value, and we would reduce its weight slightly to keep the value constant. Under such a system, gold coins would not circulate, but people would use gold certificates. These gold certificates would be redeemable in varying amounts of gold—always in just enough gold to keep the purchasing power constant or nearly so. Thus, if the index number rose from 100 to 104, we would know that prices had risen 4 per cent,[2] and we would increase the gold in the dollar by 4 per cent, from 25.8 to 26.832 grains. Students often ask if this would not introduce a speculative element into our monetary system, if people would not try to profit from expected changes in the gold content of the dollar. The answer is that they almost certainly would but that they already speculate on the changes in the general price level, and we would merely substitute one kind of speculation for another, while relieving people who do not wish to speculate from the necessity of running risks on changes in the price level which they now have to do. At present, with price levels changing, people are forced to speculate every time they take out an insurance policy, every time they put money in the bank, buy a farm, build a house, or enter into a contract that calls for a future payment of money. Fisher's compensated-dollar proposal assumes, of course, a return to the gold standard. Although such a return seems in many ways desirable, it is by no means certain as this is written.

[1] The student will recall that whenever the general price level rises, the value of the dollar falls. He can then visualize the changes in the purchasing power of money, or the value of the dollar, by looking back at Fig. 21 • 1, which shows the changes in the price level. Every time the line on that chart rose, the value of money fell; every time the line fell, the value of money rose.

[2] In practice it would be possible to make adjustments in the weight of the dollar long before the prices had risen this much. One could not keep the value of the dollar precisely constant, because no adjustment in the gold content could be made until prices had changed enough to influence the price index. But this is immaterial. The system could, if it worked at all, prevent the major changes in prices and reduce them to minor and trifling fluctuations.

We have mentioned proposals to reduce the business cycle by control of credit and of prices. Other people look at still other symptoms. They note, for example, the great fluctuation in incomes during the business cycle and advocate measures to equalize and stabilize them. Their methods may involve income and inheritance taxation, or social insurance. They may involve planned public works, which has for generations been one of the leading suggestions for government aid in times of depression. The thought behind these latter proposals is that the government should carry on the major part of its programs of building highways, dams, and public buildings during times of depression, thus paying out enormous sums and increasing incomes when they would otherwise be low and that in times of prosperity such building should be reduced to the smallest practical minimum, so as to reduce income payments at a time when they are otherwise high. Such a scheme would have the added advantage that the government would be building at times when costs were low rather than at times when they were high. It would be an economical method of carrying on government projects. For this reason, it is also often urged that private corporations should follow the same procedure and that a reasonable improvement in business forecasting on the part of businessmen would do a good deal toward evening out the cycle without any need for government interference.

Those who pay especial attention to the employment changes during the business cycle are apt to advocate measures that are intended to reduce the variations in employment and unemployment. They think that the business cycle could be reduced by the use of employment services of various sorts. Since they feel that private employment agencies do not cover a wide enough territory and are open to certain other serious objections, they usually advocate governmentally controlled employment services that will find jobs for men out of work and help them to shift to those places where employment is available. Anything that can be done to increase the mobility of labor is, of course, a step in this direction.

People who rely on the government to settle their problems for them often advocate increased amounts of government "planning"—what they call a "planned economy." To be sure, every economy is planned in some way or other, even if we merely plan to let nature take its course and let every man make his own decisions. Everyone agrees that some participation by the government in economic activities is necessary and desirable, but there are vast differences in the extent to which different people are willing to go in this direction. The issue is, of course, how much liberty the individual is willing to give up and should give up in return for the hope that the government can plan things better than they would have been without its intercession. People differ on this for two main reasons: (1) because they differ decidedly in their opinion as to the amount of liberty that men actually

have under our system and how valuable it is to them (some people think that the liberty is entirely illusory—merely liberty to be unemployed and to go hungry; while other people think that individual liberty is real and important and worth maintaining even at some expense in other ways) and (2) because they differ in their opinion as to how much better the government can plan than can individuals (some people think that the government can get the benefit of the most expert advice and can get coordination that brings results impossible when men work alone and at cross purposes; while other people think that government is in general inefficient and quite as much subject to human failing as are single individuals, and that there is also the constant danger that government power will be used for selfish personal and party ends rather than for the general welfare). With such underlying differences in the appraisal of the situation, it is natural that there is sharp disagreement over proposals for government planning. The same comments can be applied to those proposals that the business cycle be eliminated by adopting some entirely different sort of social and economic organization—by giving up capitalism and adopting communism, fascism, or some other alternative.

To summarize, then, the major proposals for correcting the business cycle have been directed at particular symptoms and have included such things as

1. Control of credit and banking.
2. Control of prices.
3. Control of incomes.
4. Economic planning by government.

Each of these proposals has many ramifications and subdivisions, depending on the methods suggested to control the prices, the credits, etc. Every one of these proposals has been tried, as have many others, and there is no agreement as to their efficacy even after the fact. As was pointed out in Chap. 2, there is no way to experiment in the social sciences—no way to try a given method and check it against control cases where the method was not used or, if the method turns out badly, no way to go back and start over again. During the great depression of the thirties, many expedients were tried, and no one has any way of knowing whether recovery came more quickly with them than it would have come if none of them had been attempted. Conclusions in the field are unfortunately matters of opinion and, despite the fact that many of the opinions are very firmly held, they differ sharply even among competent and well-informed people.

22 • 13. The Cobweb Theorem. One particular type of production cycle, which appears in actual statistics in the cases of several agricultural products and possibly also in the buildings trades, has been explained in terms of analysis strikingly similar to yet

fundamentally different from the supply and demand analysis of Chap. 15. Let us assume that the two solid curves of Fig. 22 • 5, labeled *D* and *S*, represent the demand for and the supply of some commodity. Under competitive conditions the price will be *OA* and the quantity *OB* will be exchanged. If, now, the demand increases to the position of

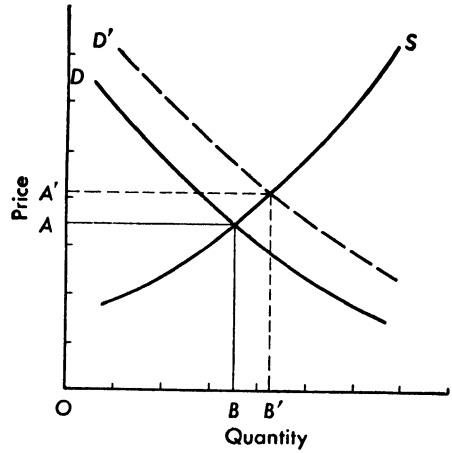


FIG. 22 • 5.—Instantaneous adjustment of price and output to changing demand conditions.

the broken line *D'*, our former analysis would tell us that the price will rise to *OA'* and the quantity exchanged will increase to *OB'*.

Let us ask, however, by what process the adjustment takes place. It has been sug-

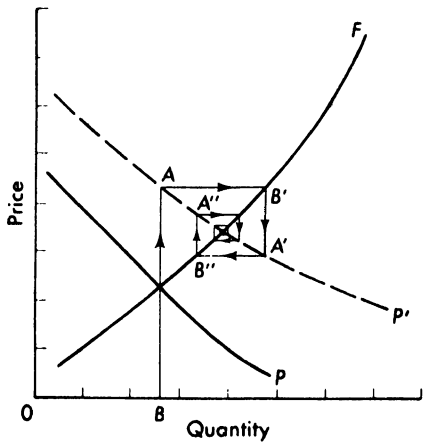


FIG. 22 • 6.—Converging cobweb cycle.

gested that the adjustment is not immediate but works itself out through a series of steps. In the first place, it is assumed that larger quantities of a product can be sold only at lower prices, or that reductions in the quantity for sale make it possible to

sell at a higher price. This relationship between the quantity offered for sale and the price at which it can be sold is represented by the solid line labeled *P* in Fig. 22 • 6. This we shall call the "price line" or "price curve," to distinguish it from the ordinary demand curve. The demand curve shows the quantities that would be taken at various prices; the price curve shows the prices that would be paid if we offered various quantities for sale. The two curves might coincide, although in practice they might not.

Next, it is assumed that production cannot be adjusted immediately to changes in the price. If the price rises, the production cannot be increased for another year; if the price falls, it will likewise take a year to reduce the output. In Fig. 22 • 6 the curve labeled *F* shows the quantity that will be produced and offered for sale next year in response to any particular price this year. We shall call it the "forecast curve." A supply

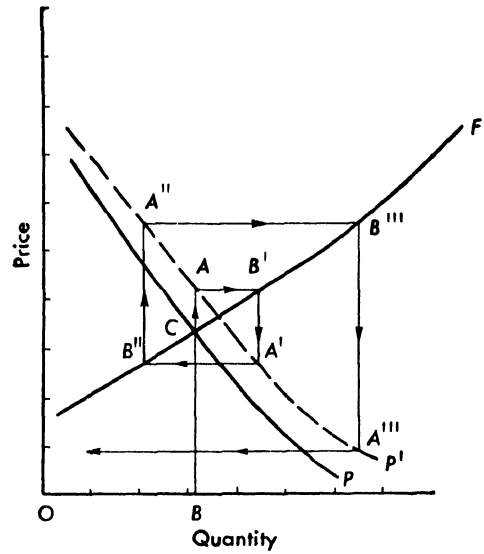


FIG. 22 • 7.—Diverging cobweb cycle.

curve shows the quantities that would be offered right now at various prices; the forecast curve shows what quantities will be offered next year in response to various prices today.

Suppose that we have been in a condition of equilibrium, producing quantity *OB* in Fig. 22 • 6. This is an equilibrium situation in the sense that this year's quantity will elicit a price that is just sufficient to keep next year's production at the present level. The situation will stay the same from year to year. But, if something happens to shift the price curve to the position of the broken curve *P'*, it will be possible to sell the quantity *OB* at a price *OA*. And the price *OA* will stimulate producers to increase their output next year to *OB'*, which will force next year's price down to *OA'*. But this low price will influence producers to reduce their output the following year to *OB''*, which will raise the price to *OA''*. Thus *OB* and *OA* are the quantity and price for one year, *OB'* and *OA'* are the quantity and price for the following year, *OB''* and

OA'' are the quantity and price in the third year, and so on. The student will see from the diagram that a sequence of alternating high and low outputs (or, alternating low and high prices) has been set in motion, slowly approaching an equilibrium at which the price and output will correspond to the crossing point of the curves F and P'. We have a cycle, but a *converging* one, so that the amplitude of the swings slowly grows less and less. Because of the appearance of the diagram in Fig. 22 • 6, with its line slowly approaching the center like the web of a spider, this explanation of cyclical movements is called the "cobweb theorem."

Yet there is no certainty that cycles of this sort will converge toward a new equilibrium. Let us look at Fig. 22 • 7, in which the curves have the same meaning as in Fig. 22 • 6. This time the original equilibrium output is OB at a price OC but, when

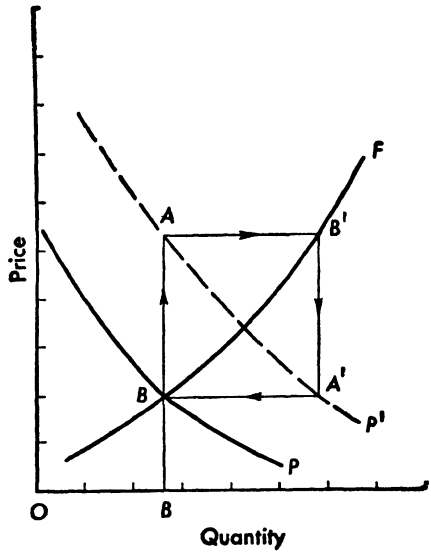


FIG. 22 • 8.—A cobweb cycle of constant amplitude.

the price curve shifts to the broken position P', we shift as before to a first-year price OA, second-year output OB' at second-year price OA', third year output OB'' at third-year price OA'', and so on. Again we get a cycle of alternating high and low prices at alternating low and high volumes of output, but this time the cycle is *diverging*, with the swings getting bigger and bigger. Under such conditions, no new equilibrium would ever be reached. In fact, although we can draw such curves in theory, the student will see that there are limits beyond which this sort of fluctuation would not be likely to occur in practice. Converging cycles are more reasonable than diverging ones.

Finally, let us look at Fig. 22 • 8. Following the same analysis as before, we see that, if the equilibrium of the two original curves is once upset, we never reestablish it but rather alternate year after year through a regularly recurring *constant* cycle, neither converging nor diverging. The price the first year rises to OA, the next year the quantity is OB' and the price OA', the next year the quantity returns to OB and the

price to OA, and we then alternate prices from OA to OA' to OA to OA', indefinitely, and quantiites from OB to OB' to OB to OB', indefinitely.

Cycles of the "cobweb" sort may obviously be converging (approaching a new equilibrium), diverging, or constant. Which they are in practice will depend, as the student may discover for himself by experimenting with diagrams of his own, on the relative steepness or elasticity of the two curves. When the forecast curve is the steeper, the cycle converges. When the price curve is the steeper, the cycle diverges. If the two curves have the same slope, the cycle repeats itself indefinitely with constant amplitude.[1]

Although some actual cycles similar to the cobweb have been found in real life, it should be pointed out that the cobweb analysis is based on certain assumptions which are probably contrary to fact. In the first place, it is assumed that the crop that a man

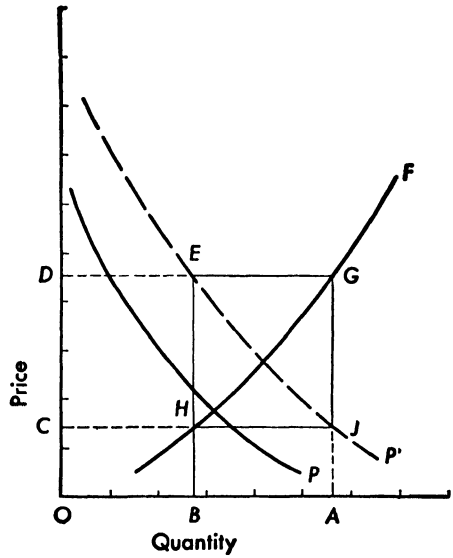


FIG. 22 • 9.—Profit and loss with a constant cycle.

plants in 1950 depends solely on 1949 prices. As a matter of fact, this is not true. Whereas 1949 prices do undoubtedly influence decisions as to the 1950 crop, producers are also influenced by their expectations. Men's production decisions depend not only on the backward look but on the forward guess. If last year's price was high, producers are apt to foresee some reaction to the high price and anticipate larger outputs by their competitors.

Next, we should notice why divergence would not continue indefinitely in practice. Let us suppose that the diagram of Fig. 22 • 7 represents the conditions in the production of potatoes. The curves show us that, once the equilibrium is upset, we fall into a series of diverging cycles. But common sense tells us that this cannot continue, because the forecast curve is certain to get very, very steep as we move across the diagram toward

[1] For a further discussion of this point the student is referred to Note 13 of the Mathematical Appendix, p. 905.

the right. The only way we can produce more and more and more potatoes is to withdraw land and other productive factors from other uses. The diagram as drawn suggests that this would continue without limit, until all the productive factors of society were concentrated in potato production. We know on the contrary that as potatoes became more plentiful and as other products became scarcer, the forecast curve would become very steep indeed. Even a very considerable increase in the price of potatoes would not shift much more land out of the production of other crops and into that of potatoes. Hence, in practice, the shape of the curves is such as to make continued divergence impossible.

There is even reason to suppose that the constant cycle could not continue forever. This is clear if we look carefully at Fig. 22 • 9. Let us suppose that the industry is freely competitive, so that the price is equal to the average cost. Let us note the amount of profit and of loss in the constant cycle of Fig. 22 • 9. In the first year, we produce quantity *OB*, our income is *OBED*, and our expenses are *OBHC*. Our net income is *CHED*. In the second year, our production is *OA* and our gross income is *OAJC*. Our expense is *OAGD* and our loss is *CJGD*. Producers experience alternating years of profit and of loss, but the losses always exceed the profits. Bankruptcy would bring an ultimate end to a cycle of this sort. We can assume, then, that insofar as cycles of the cobweb variety appear in practice they are either converging cycles that tend toward a new equilibrium, or temporary affairs limited by the demands for competing goods or by the ultimate bankruptcy of people in the industry.[1]

Questions

1. Mention several secular movements with which you are familiar. Several cyclical movements. Of the latter, which are and which are not seasonal movements?

2. The term depression is used very loosely in common speech. If you had to tell exactly what it means, how would you define it?

3. Do you think that the present is a time of prosperity, recession, depression, or recovery? What evidence do you have to substantiate your conclusion?

4. What were some of the political results of the Panic of 1873 and the following long decline in prices?

5. Which do you think would be worse: to have business slightly depressed for 10 years or badly depressed for 2 years? Discuss.

6. How many separate cycles do you discover in the data of Table 22 • 1 and Fig. 22 • 3? What were the dates of the crests and of the troughs?

7. Describe the period of recession.

8. Some authorities have said that the feelings of optimism and pessimism in business are the results of the business cycle, while other people have felt that they were the cause. How might the latter be true?

[1] In this section on the cobweb theorem, the author has leaned very heavily on the exposition by Oscar Lange.

9. It has been suggested that changes in sunspots and other such things bring about a cyclical change in the average state of the health of people. Would this influence at all your answer to question 8? How?

10. Do any of the suggested causes of the business cycle, listed in Sec. 22 • 11, seem more reasonable to you than others? If so, try to rank them from the most to the least reasonable. Explain why some of them seem unreasonable to you.

11. If you were to attempt to explain the business cycle in terms of the errors in forecasting which businessmen make in a system that uses capital, would you think that such errors would be more common in times of depression or of prosperity? Why? What effect would it have on the cyclical nature of the resulting movements?

12. If you were forced to treat symptoms because you did not know the cause of the business cycle and had to direct your efforts at only one symptom, which of the items listed in the first column of Table 22 • 2 would you pick out to attack? Why would you pick this particular one?

13. Suppose that when the index number of prices is 100, the dollar contains 15 grains of gold. If the index number now changes to 95 and you are following the principle of the compensated dollar, what new weight will you select for the gold dollar? Why?

14. "Economists know so little about the business cycle, we might just as well stop fooling with it." Comment.

15. Draw original large-scale diagrams showing convergent, divergent, and constant cobweb cycles.

16. The price of potatoes per bushel and the production of potatoes in millions of bushels in each year from 1930 to 1941 are shown in the accompanying table. Prices are in "1926 dollars"; that is, they have been corrected for changes in the value of money and show what the price of potatoes would have been each year if the dollar had retained a constant purchasing power equal to what it had in 1926.

Year	Corrected price	Production, million bushels	Year	Corrected price	Production, million bushels
1930	$1.05	344	1937	$0.61	376
1931	0.63	384	1938	0.71	356
1932	0.59	375	1939	0.90	342
1933	1.25	343	1940	0.90	376
1934	0.60	406	1941	0.93	356
1935	0.74	379	1942	....	370
1936	1.42	324			

In Fig. 22 • 10 each dot represents the production for a given year and the corresponding price at which potatoes sold in that year. A smooth curve has then been drawn on the diagram to show the average relationship between the production any

year and the price the same year. This curve, which corresponds roughly to what we called the "price curve" in Sec. 22 • 13, has been labeled *P*. In Fig 22 • 11 each dot represents the price of a given year and the production in the following year. Again

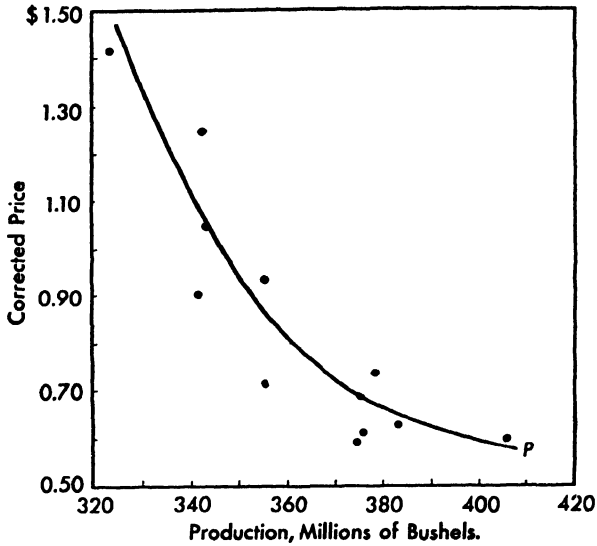


FIG. 22 • 10.—Production of potatoes and corrected prices, United States, 1929-1941.

a smooth curve has been drawn on the diagram to show the average relationship between the price of any year and the production in the following year. This curve,

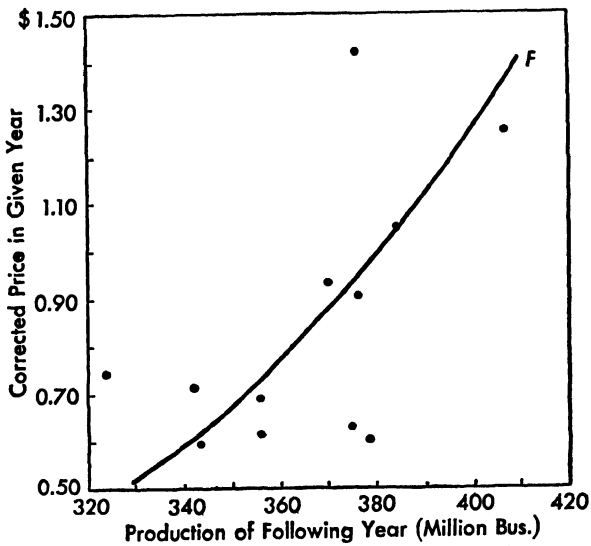


FIG. 22 • 11.—Price of potatoes each year and production of following year, United States, 1929-1941.

which corresponds roughly with what we called the "forecast curve" in Sec. 22 • 13, has been labeled *F*.

In Fig. 22 • 12 the two curves of Figs. 22 • 10 and 22 • 11 have been put on the same diagram. Thus we have the basic raw material for cobweb analysis. Evidently,

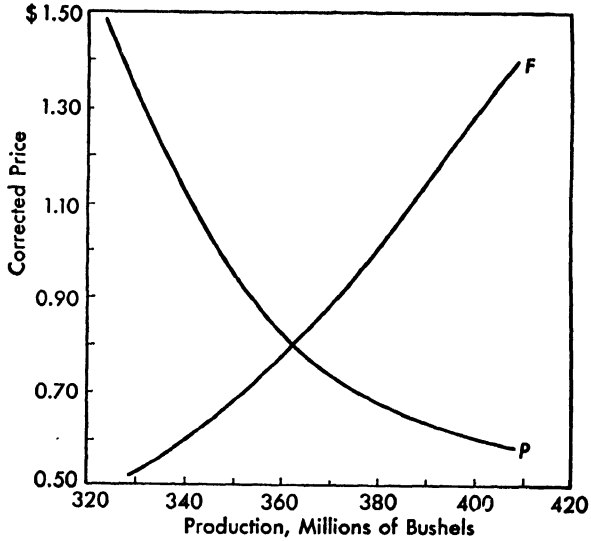


FIG. 22 • 12.—Price and forecast curves for potatoes, United States, 1929-1941.

the equilibrium output would have been approximately 362 million bushels at a price of 79 cents per bushel.

If for any reason the actual production turned out to be only 340 million bushels, what would be the sequence of events in accordance with the cobweb theorem? Is the situation converging, diverging, or constant? Suppose that the production of 340 million bushels occurred in 1950. Give the production and the price in each of the next 5 years.

17. Under what circumstances do cobweb cycles converge? Diverge? Remain constant?

18. What reasons can you give for believing that actual cobweb cycles are more apt to converge than to diverge or to remain constant?

19. Assume that the figures in the accompanying table show the production (in millions of tons) and the price (in dollars per ton) of a given commodity in the years

Year	Price, dollars per ton	Production, million tons	Year	Price, dollars per ton	Production, million tons
1950	$80	61.6	1954	$58	64.0
1951	42	65.8	1955	54	64.5
1952	63	63.5	1956	56	64.2
1953	50	64.8	1957	55	64.4

enumerated. Plot the data on a diagram and find the approximate location of the price line and of the forecast line. Plot these two lines on a separate sheet of graph paper.

20. Using the final graph made in question 19, compute the equilibrium price and the output.

21. Continuing with the final graph of question 19, assume that in 1950 the production had been 68 million tons. What would have been the price and the production in each year 1950 to 1955, inclusive? Is the cycle converging, diverging, or constant?

23. INTERNATIONAL TRADE

23 · 1. Purposes of International Trade. In Chap. 10 we discovered that there are definite gains from specialization and trade. We found that, if areas specialize in the production of the commodities in which they have the greatest inherent advantage and import other commodities, they will in the end have more of all the commodities concerned than if they had tried to produce them all at home.[1] This gain through specialization and trade is exactly the same whether or not the trading areas happen to be separated by a political border. Two adjacent farms may find it advantageous to specialize and trade. If a township line separates them—or a county line, a state line, or an international boundary line—the basic situation is unaltered. Trade between people in different nations has the same fundamental advantages as between neighbors in the same country. It is carried on because differences of soil or climate or resources cause differences in production costs in the different areas and, with these differences in production costs given, men would be foolish not to take advantage of their own peculiar local superiorities. Geographical specialization and trade enable countries to obtain materials that are not to be found at home, and to dispose of surpluses that could not be used at home. As we have seen, both countries involved in an exchange of goods may gain even though the money values of the goods furnished by them are equal, because of differences in the utility of the goods in the countries. International trade, like domestic trade, springs up voluntarily because the buyers and the sellers involved find it to their mutual advantage. The tendency of such trade is to force labor and other resources from inefficient to efficient industries. If all the productive factors were completely mobile, the productive advantages of particular areas would quickly disappear; but natural resources have almost no mobility, and labor is only very imperfectly mobile. Even in the case of capital, which is the most mobile of the factors of production, it is commonly true that men prefer to invest their capital at home rather than abroad, partly

[1] See especially Secs. 10 • 9 through 10 • 11.

because of the greater ease of oversight and completeness of understanding and partly for supposed patriotic reasons. Since the factors of production do not move freely, the differences in cost of production persist, and the advantages of specialization and trade persist likewise.

In talking of matters of foreign trade we fall easily into the loose habit of speaking of trade between countries—of the imports "of the United States" or of trade between "the United States and Canada." It is important to remember that foreign trade, like domestic trade, is usually[1] trade between individuals and business firms. The imports "of the United States" are not purchases by our government but purchases by firms in the United States from firms abroad. The trade between "the United States and Canada" is not trade between the two governments but trade between private individuals who happen to be separated by an international boundary line. It is true that in recent years we have witnessed a tendency to return to increased governmental interest in and interference with international trade of the kind that was common in medieval times but, in the main, the trade is still individual trade and has the advantages that characterize trade between individuals within the same country.

If we think about the economic problems of the individual, we realize that he is constantly importing (buying) and exporting (selling) goods. He finds it convenient to use money in effecting these transactions but, if we look beneath the surface, we realize that he is actually exchanging goods for goods. He sells his surplus goods for money and uses the money to buy goods that he needs. If he does not produce surpluses for sale, he has nothing with which to buy. The surpluses that he has produced are of little value to him except as they enable him to get goods from other people. The surpluses are not an end in themselves, but a means to an end. The potato farmer who grows 6000 bushels of potatoes would have wasted his time if he could not exchange them for shoes, grain, housing, and washing machines. Students commonly overlook the fact that the same situation exists in foreign trade. Surpluses of goods for export are of little use to us except as we can get imports with them. The exports are not an end in themselves, but a means to an end. Exporters and importers find it easy to effect their transactions in monetary terms but, if we look beneath the surface, we realize that they are actually exchanging goods for goods. Surpluses produced in the United States, which are of little value to us, are exported in order to establish credits abroad with which we can command supplies of tin, rubber, sugar, and silk for import. If we do not export, we have nothing with which to buy abroad. The cotton farmers of the country, who have produced far

[1] This would not be true, for example, in time of war, when lend-lease or other government trade might make up most of the foreign trade; and it may not be true of such countries as Russia, where foreign trade is a state monopoly. Aside from such cases, the statement is correct.

more than can be used within the country, have wasted their time unless the excess cotton can be exchanged for tin, rubber, sugar, and silk. *Imports, not exports, are the goal of foreign trade. Exports are but a means of getting imports.* This fact needs emphasis and thought, because men so often take only the superficial view that leads them to think that exports are to be preferred to imports—that exports are somehow "better" than imports. As a matter of fact, exports and imports are two sides of a transaction. Every purchase is a sale, and it is impossible to maintain that purchases are better than or worse than sales since they are both parts of the same transaction. Similarly imports and exports are irrevocably tied together. We cannot have

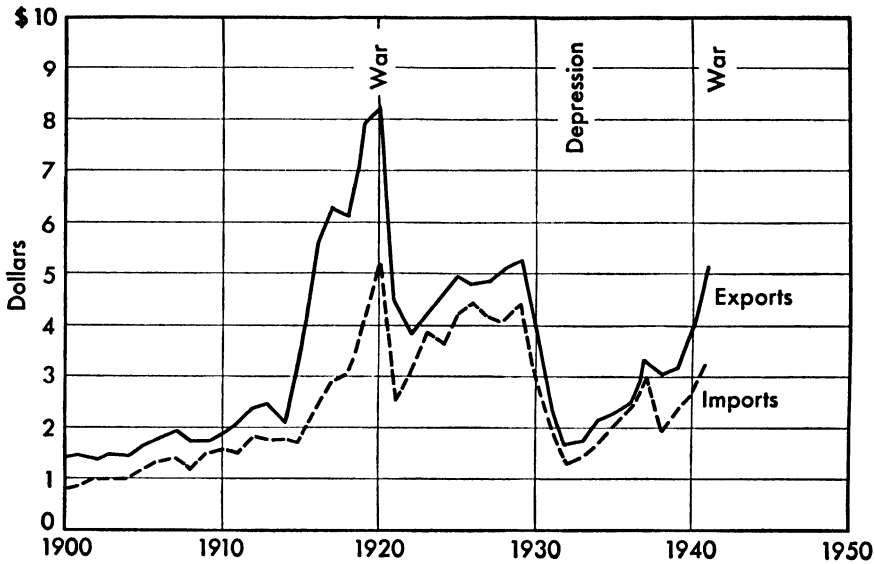


FIG. 23 • 1.—Value of merchandise exports and imports, United States, 1900-1941.

the one without the other, except in the cases (which of course cannot continue) where men get something and refuse to pay for it.

23 • 2. Balance of Trade. The relationship of imports and exports needs, however, some further comment. One might draw the conclusion from what has just been said that a nation's imports and its exports must always be equal, or at least that they must be equal on the average over a period of several years. Yet Fig. 23 • 1 shows that the exports of the United States have exceeded the imports each year since 1900 and, if the chart showed data for earlier years, we would discover that exports from the United States have exceeded the imports into the United States in every year without exception since 1893, and in only 2 years since 1875 have the imports been as large as the exports.

This situation we can again understand best by taking an analogy of trade between individuals. If we have two farmers, one of whom produces apples and the other corn, and they exchange surpluses, the value of corn given up ought to equal the value of apples given up if the trade is to be fair. Of course, this need not be true in each separate year. The apple grower might give up the greater value this year, and give credit to the potato farmer for the difference; but in such a case the situation should be reversed in some future year. In the long run, the value of goods given up by either farmer should be equal to the value received if the exchange is a fair one.

But, suppose we consider the relationship between the potato farmer and his doctor. It is quite likely that the trade will appear now to be all in one direction. We find the potato farmer giving up perhaps $50 worth of potatoes each year, year after year, and getting no commodities back at all. Yet we would not necessarily call the situation unfair, because we realize that the doctor has given services in place of goods. His advice is valuable, and the farmer is willing to pay for it. Payments need not always be made in goods. If we evaluate merely the goods that change hands, it is quite possible that we shall mislead ourselves. In the example just taken, we would have said that the potato farmer was "exporting" more than he "imported." His "exports" amounted to $50 and his "imports" to nothing at all. An individual or a nation that exports more than it imports is playing a losing game, unless there are some unseen items going in the other direction to offset the seeming discrepancy.

Because people commonly think superficially, men have come to speak of a situation like that depicted in Fig. 23 • 1, when the value of the exports exceeds the value of the imports, as a "favorable balance of trade." On the other hand, they have spoken of an excess of imports over exports as an "unfavorable balance of trade." We shall continue to use these terms in this accepted manner, although it is important for us to emphasize these points in connection with the terms.

1. A favorable balance is not any more desirable than an unfavorable balance. In fact, if we were to choose either one, we would probably select the unfavorable balance, since it would mean that we were getting goods without paying for them.
2. Neither a favorable nor an unfavorable balance can continue if we include all the goods and services involved. If we consider only merchandise transactions and neglect the services, it is possible to have a continued balance always in the same direction. The data for Fig. 22 • 1 are data on merchandise transactions only, as are most data on imports and exports.

23 • 3. Invisible Items in the Balance of Trade. It is relatively easy to determine the value of physical goods imported and exported. Such goods

(save for cases of smuggling) pass through regular ports and are inspected by customs inspectors. In every country careful records are kept of them. These items of trade are called the "visible items" and are the ones on which figures are usually given. In addition to the silk, rubber, and coffee that Americans import from abroad, there are many "invisible services" which they obtain—satisfactions which they get from foreign countries and for which they have to pay to people in those countries. These invisible imports are like any other imports, in that they add to the satisfaction of Americans and require payment of American money or goods to foreigners. Thus, for example, when American tourists go to Europe to visit Stratford-on-Avon or the Swiss Alps, to drink French wine and eat German cheese, the goods and services that they receive do not cross a frontier and are not checked by a customs officer. Yet they are "imports" from the standpoint of Americans and do have to be paid for. Just as the potato farmer who received no physical goods from his physician was not really being wronged but was receiving full value in other forms than in goods so, if American producers shipped abroad a larger value of goods than they imported, the difference might be made up by travel expenses of Americans abroad, by remittances of Americans to relatives in Europe, or by the use of foreign ships to carry commerce, etc.

With nations, as with individuals, the trade tends in the long run to balance if we include all the valuable items involved; and if we count up and evaluate all the goods and the services and if imports still exceed exports, then we shall have to make good the difference in money. On the contrary, if our exports exceed our imports after we have included the invisible items along with the visible ones, then we should receive money to make up the deficit. *Exportation of goods and services tends to stimulate importation of money; importation of goods and services tends to stimulate the exportation of money.* The name "favorable" was given to an excess of exports by men who thought that money was an end in itself. They realized that an excess of exports would tend to stimulate an inflow of money (although they did not realize fully, as we do today, that this need not be true if only merchandise trade were considered), and they did not realize that goods rather than money are the end of life.

The national real income of any nation—the goods and services that its people have to consume—is made up of the domestic production plus the imports minus the exports. If we put this in the form of a simple algebraic equation we have

$$\text{Real income} = \text{production} + \text{imports} - \text{exports}$$

In all of these, services must be included along with physical goods. It is evident that real income rises if we increase our imports or decrease our exports

or both. A so-called "unfavorable" balance of trade would indicate (if we had included all the invisible items in computing it) that our real income was larger than it would have been with no international trade at all. A so-called "favorable" balance of trade would indicate that our real income had been diminished by international trade. The value of imports and exports can differ only if someone is getting goods for which his is not paying. Such an arrangement is obviously of greater advantage to the receiver than to the unpaid provider, more advantageous to the importer than to the exporter. But such an arrangement can continue only as a matter of charity or compulsion, not as a matter of voluntary trade.

23·4. Differences between Foreign and Domestic Trade. In all the things mentioned so far, foreign trade is like domestic trade. The terminology may be somewhat different, since we do not talk of imports and exports in the case of domestic trade; but we do talk of purchases and sales, which mean the same thing. Yet, while domestic and foreign trade are alike in their essentials, nevertheless there are some differences that are worthy of note.

In the first place, as soon as we shift from domestic to foreign trade we run into new and troublesome monetary problems. Two men carrying on domestic trade have the advantage of a common money, acceptable to both; but two men carrying on foreign trade use different sorts of money. The buyer of goods has no money of a kind acceptable to the seller; the seller of the goods cannot use those kinds of money ordinarily available to the buyer. Therefore, in cases of foreign trade we have the problem not only of the exchange of goods but also of the exchange of money. This latter problem we call the problem of "foreign exchange," and we shall discuss it in the following chapter. When we talk of foreign trade or international trade, we are talking about the exchange of goods and services; when we talk of foreign exchange, we are speaking of the exchange of different sorts of money. This latter problem is peculiar to foreign trade.[1]

Although the problems that arise because the buyer and the seller use different kinds of money are the most important ones that distinguish foreign from domestic trade, one should not overlook the problems arising from differences in language, customs, and laws. An American dealing with another American finds it relatively easy to reach an understanding. Even when the buyer and the seller talk the same language and are accustomed

[1] To be sure, a similar problem can arise in domestic trade and did actually arise much more often before the institution of the Federal Reserve System with its system of clearings. Business firms occasionally insist on payment with funds payable in their own city, or in New York. The advertisement may say, "Price $5, payable in New York funds only." In such a case a man in Connecticut would have to arrange to obtain funds payable in New York, instead of payable at his own bank; and we would have a case of domestic exchange similar to the problems of foreign exchange. This situation is fortunately becoming rather uncommon.

to the same weights and measures and usages of trade, there are possibilities of misunderstanding, as any lawyer can testify from personal experience. But, when the parties to an agreement talk different languages and are familiar with different sets of customs and laws, the problem becomes more complex and the possibilities of misunderstanding are multiplied. Then we have to add the fact that governments often set up artificial impediments to international trade—impediments which either do not exist at all or which exist only in a comparatively minor degree in the case of domestic trade. These may take the form of absolute prohibitions of some kinds of foreign trade, or of special taxes paid on foreign trade from which domestic trade is exempt; or they may be "health and sanitary regulations" which have to do with health and sanitation in name only, and which are actually used to stifle foreign trade. Although such things can be found also in domestic trade, they are much less common, much less open, and much less effective there than in foreign trade.

Finally, we find that there is much more government interest in and interference with foreign trade than is the case with domestic trade. Governments have long looked upon foreign trade as a special source of revenue and have tended to think about it as more subject to regulation than domestic trade. If one party gets the better of the other in domestic trade, it is still "all in the family"; but, if foreigners get the better of our own citizens, then our people are poorer and the foreigners are wealthier. Moreover, much of the thought on these matters is sadly out of date, going back to the mercantilistic ideas of a century and a half ago.[1] Government interference in many cases arises because poorly informed people believe that exports are desirable and imports undesirable. Therefore, governments still often follow policies intended to encourage exports and discourage imports of goods. To be sure, such policies are often followed, as far as feasible, in internal trade. Chambers of commerce and merchants' associations often try to convince local people to "buy at home"—do not patronize the chain store or the mail-order house, but "keep your money at home." This idea is as provincial and as foolish in domestic as in foreign trade; but in foreign trade it carries the weight of long though thoughtless acceptance in high places.

23 · 5. Importance of Foreign Trade. The importance of foreign trade varies tremendously from country to country. Naturally, the larger the country, its population, and its productivity, the greater its foreign trade is likely to be. One would expect a country with 50 million inhabitants to carry on more foreign trade than a country of 1 million. Although large countries might be expected to carry on more total foreign trade than small ones, paradoxically enough, the larger and more varied the resources of the

[1] See Sec. 3 · 12.

country, the less important foreign trade is likely to be, relatively. If the country covers a vast area, with varied climates and soils and resources, most of the things that its inhabitants want can be produced efficiently somewhere within its own territory. If, on the other hand, the area is small and homogeneous, few things can be produced within it and most things must come in from outside. Thus, the relative importance of foreign trade is somewhat artificial, depending on the almost chance location of national boundaries. The foreign trade of the United States is small as compared with its domestic trade, probably only about one-twentieth as large. If the 48 states of the United States were carved up into as many independent countries, each of them would find its foreign trade greatly increased in relative importance. The Vermont farmer sending maple sirup to Boston would be carrying on exactly the same trade as before, but now it would be listed as an "export" and the codfish that came back would be "imports." The smaller and more homogeneous the area, the larger will its foreign trade be relative to its domestic trade and the smaller will its total trade be.

The student will readily understand from these general statements the situation of the United States in foreign trade. As one of the largest countries in the world, its foreign trade looms large absolutely; yet its foreign trade is small compared with its domestic trade. Various sorts of estimates have been made, based on various kinds of evidence, but it seems to be true that in ordinary times the foreign trade of this country is only about 5 per cent as large as its internal domestic trade. The foreign trade of most countries is a much larger proportion of their domestic trade. It has been estimated, for example, that prior to the second World War Great Britain's foreign trade amounted to 25 per cent of its domestic trade, while the corresponding figure for Germany was 20 per cent. These countries, smaller than the United States and with less varied resources, found it necessary to depend more heavily on foreign sources of supply.

Yet it is not entirely fair to judge the importance of foreign trade solely on the basis of its "percentage importance" by saying that it makes up but 5 per cent of our domestic trade. The heart of an adult human being weighs perhaps a half pound, or maybe one-third to one-quarter of 1 per cent of the entire weight of the person. Yet, no one would measure its importance in such terms. When we start to ask how easily we could get along without it, we begin to realize that it, like many other organs, is absolutely necessary to the continued operation of the entire being. Similarly, although international trade makes up a relatively small part of the value of our total trade, it is nevertheless a part of considerable importance to the functioning of our economic organism. To be sure, if we had to give up either our domestic or our foreign trade, we would unhesitatingly relinquish the latter; but under such circumstances it would be impossible for our people to continue

to lead the kinds of life to which they had been accustomed. We would lose our customary foreign markets for a half billion dollars' worth of petroleum and another half billion dollars' worth of machinery each year. Our foreign sales of cotton, although subject to considerable fluctuation from year to year, usually have run to another half billion dollars or more. Automobiles and accessories have amounted to only a little less, while we have also relied on foreign markets to absorb millions of dollars' worth each year of iron and steel products, chemicals, tobacco, wheat, fruit, and textiles. Any of these industries would have to undergo radical reorganization if foreign outlets were cut off. On the other side of the ledger, without foreign trade the American people would have to get along without, or with greatly reduced supplies of, rubber, sugar, coffee, paper, silk, tin, hides and skins, etc. These are all important articles of import. To be sure, we could live without any one or even without all of them; but it would be a very different sort of life. Many of the products that we have been accustomed to regard as products of domestic manufacture are in fact made in whole or in part from raw materials brought to us from abroad. Therefore, while it is important for us to understand that, in a great and richly endowed country like our own, foreign trade is far less important than domestic trade, nevertheless it is just as important that we should not draw the false conclusion that our foreign trade is a matter of indifference to us. As in so many cases, simple percentages fail to tell the whole story, and one needs to study the entire situation in some detail before he is qualified to draw conclusions based on proper insight.

23·6. Problems of Measurement in International Trade. At the time this is written, it is too soon to predict the nature of the postwar international trade of the world. The last available figures not influenced by wartime conditions are those for 1938. Even those figures are difficult to interpret, as figures on international trade always are. This is true for several reasons; first, because of the differences in monetary units in the different countries, so that the trade of the United States is measured in gold dollars (or inconvertible paper dollars) while that of Great Britain is in gold pounds and that of India in silver rupees. To add to the difficulty, we have to remember that the monetary units of the various countries themselves change as the countries go on or off the gold standard, or revalue their currency, as the United States did in 1933. Thus, if we consider trends in foreign trade over a period of years, it is not sufficient to reduce the data for other countries to dollars. We have to tell whether they are in "old" or "new" dollars. Finally, we are faced with the difficulty that changes in the value of foreign trade are often misleading because of changes in the general price level. Thus, for example, if the value of foreign trade had doubled from 1910 to 1920 in terms of dollars, there would have been an actual decline in the quantity of

goods exchanged, since prices more than doubled in those 10 years.[1] Money is always a poor measure when applied in cases involving a lapse of time, and in drawing conclusions about trends in foreign commerce it is important for us to make corrections for the monetary units involved and for changes in the general price level if we are not to be misled.

23 • 7. Recent Trends in World Trade. International trade, like domestic trade, fluctuates violently with the changes of the business cycle. Figure 23 • 2 shows how the foreign trade of the countries of the world rose through the general expansion period of the twenties, and disappeared quickly with the onset of the depression in the thirties. Of course, foreign trade of the ordinary kind was almost wiped out with the onset of the war, giving place to lend-

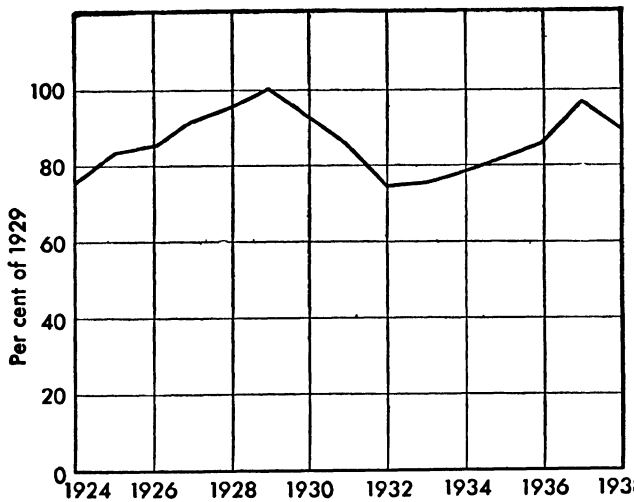


FIG. 23 • 2.—Index of physical volume of world trade, 1924-1938. 1929=100.

lease and other war-directed and governmentally sponsored trade. Figure 23 • 2, by the way, illustrates well one of the points made in Sec. 23 • 6. In 1932, the value of world trade was but 39.1 per cent of the 1929 value; but the physical volume in 1932 was 74.5 per cent of the 1929 physical volume. The volume had dropped about 25 per cent while the value was dropping over 60 per cent. Very evidently the great reduction in value was in large part the result of falling prices, and only in a lesser degree was it a reduction in the quantity of goods handled.

In world trade in general, roughly a quarter of the commerce is in foodstuffs, about two-fifths in manufactured goods, and just over a third in raw materials (see Fig. 23 • 3). These proportions fluctuate somewhat from year to year, but the variations are not of any considerable size. If, however,

[1] See Table 21 • 3.

we study individual countries, the year-to-year fluctuations are naturally much more marked. Up to the outbreak of the second World War there had

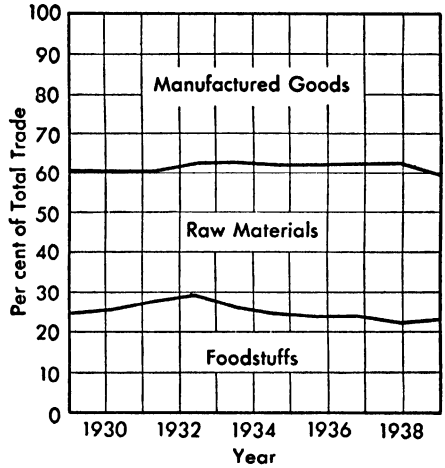


FIG. 23 • 3.—Percentage make-up of world trade, 1929-1938.

been marked trends in the exports of the United States, with a noticeable shift toward greater and greater exportation of manufactured products and

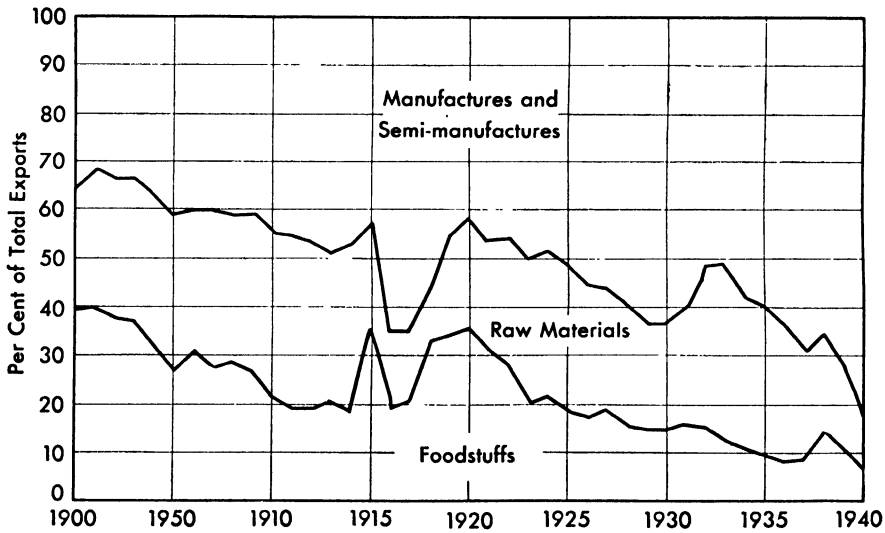


FIG. 23 • 4.—United States exports, 1900-1940.

away from the exportation of foodstuffs. As the country was becoming less agricultural and more industrialized, the exports were altered accordingly (see Fig. 23•4). The imports, on the other hand, showed much less

definite trends. Between 25 and 30 per cent of them were foodstuffs, and about 40 per cent were manufactured or semimanufactured articles. The remaining 30 to 35 per cent were raw materials (see Fig. 23 • 5). The effect of the first World War is very noticeable on both the exports and the imports of the United States in the years 1914 to 1918.

The world trade of a century or a century and a half ago was largely between countries in the same latitudes, but countries that differed in the stage of their economic development. The older and more industrialized ones were sending finished products to their colonies and bringing back in return the raw materials that they needed. In the first quarter of the nineteenth century, for example, roughly two-thirds of the exports of the United

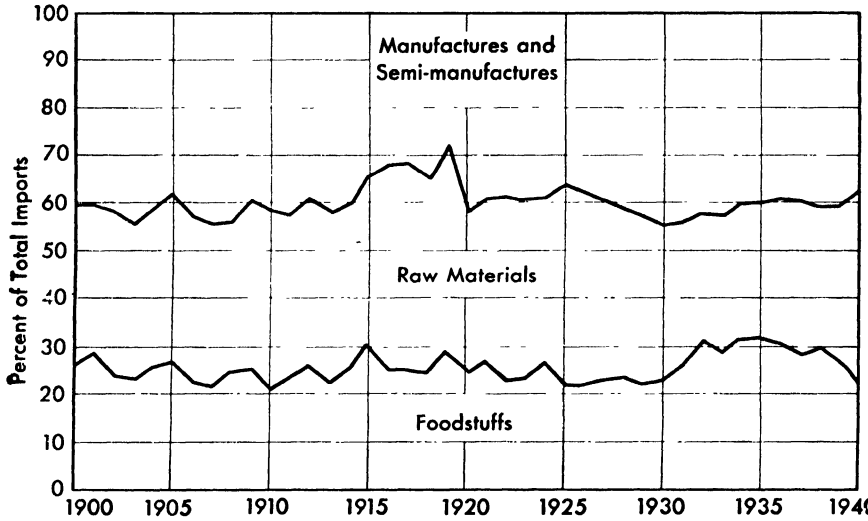


FIG. 23 • 5.—United States imports, 1900-1940.

States were crude raw materials, and almost 70 per cent of the imports were either manufactured foodstuffs or other finished manufactured goods. With the passage of time and the shifting of both populations and capital, these differences in the stage of economic development have tended to disappear. The products of the United States have come to be more and more nearly like those of England and the industrial parts of continental Europe, and as a result the advantages of trade between these areas have become less and less. As these technological differences have come to be less important, differences in climate, soil, and natural resources have come to be more important, and the old east-west trade has tended to give way to a north-south trade. The trade of the temperate zones with the tropics has come to be the important trade between nations, and there are indications that it is likely to continue of major importance. Europe, which took over 80 per

cent of our exports just after the Civil War, has dropped to less than 50 per cent. At the earlier date, some 55 per cent of our imports came from Europe; now the figure is somewhere around 30 per cent. During the same period our growing imports of rubber, silk, and tin have swollen our Asiatic imports from an original 10 per cent to a present 30 per cent, and our exports to Asia have jumped from 1 to 2 per cent to over 15 per cent. Similarly many of our other major import items, such as coffee, come from tropical or semi-tropical areas.

The United States has come to be the world's foremost importer of raw materials, and one of the two leading nations in world trade. Table 23·1

Table 23·1. Leading Countries in World Trade, 1929 and 1938, Showing for Each Country Its Percentage of Total World Trade in Each Year*

Country	1929	1938
United Kingdom	13.05	13.84
United States	13.84	10.66
Germany and Austria	10.47	10.00
France	6.19	4.69
Canada	3.68	3.46
Japan	2.86	3.21
Belgium	2.73	3.19
Netherlands	2.78	2.87
India and Burma	3.05	2.96
Italy	2.83	2.41
Australia	1.89	2.20
Sweden	1.40	2.09
Union South Africa	1.27	2.08
China without Manchuria	2.13	1.98
Argentina	2.52	1.88

* Data from "Review of World Trade, 1938," p. 21, League of Nations, 1939.

shows the proportion of the total world trade carried on by each of the leading countries in 1929 and 1938. Data are given for every country that accounted for 2 per cent or more of the total trade in either year.

Among the other trends in world trade to be mentioned is one that was particularly noticeable in the years between the two great World Wars. This was the tendency for national political domination of international trade, and the use of such trade for partisan and nationalistic purposes rather than for ordinary economic advantage. Countries came more and more to restrict certain segments of their foreign trade and to encourage others. Especially, it was true that trade tended to be confined to a greater extent to political groups of nations—the British Commonwealth of Nations, the trade between Germany on the one hand and central Europe or South America on the other hand, the trade between the "yen bloc" of Asiatic

nations, etc. Instead of allowing the course of foreign trade to follow the natural lines of comparative advantage, nations tried increasingly to force trade to follow the lines of supposed political advantage. Tariffs, trade quotas, exchange quotas, and various other sorts of restrictions came to be increasingly common as the advantage of individual citizens was sacrificed for real or fancied national interest. As a result, the total volume of world trade was considerably smaller than it otherwise would have been.

23·8. United States in World Trade. The exports of the United States are greater than those of any other country, and her imports second only to those of Great Britain. Her foreign trade is unusually diversified, as to both commodities traded and areas with which trade is carried on. Such tropical products as cane sugar, rubber, jute, hemp, coffee, tea, cocoa, and vegetable oils usually make up more than a quarter of her imports, while the tropical countries of Latin America send her mineral oils, ores, and metals. Thus the tropical regions between them usually account for more than half of the import trade. It has been suggested[1] that the United States tropical imports represent the labor of some 15 million workers, or many times the labor necessary for producing a corresponding value of exports. The tropical countries, however, are considerably less important in our export trade.

The total trade of the United States is so large that she is a major buyer or seller of many of the commodities that enter into foreign trade. In the world import market, the United States accounted for 69 per cent of all imports of silk in 1938, 56 per cent of the bananas, 49 per cent of the coffee, 40 per cent of the tin, 38 per cent of the rubber, 31 per cent of the sugar, 23 per cent of the manganese ore, and 20 per cent of the bauxite. Of all the iron and steel scrap exported in 1938, 58 per cent was exported by the United States, which also exported 42 per cent of the corn, 38 per cent of the tobacco, 38 per cent of the cotton, 33 per cent of the gasoline, 23 per cent of the unworked copper, and 21 per cent of the crude petroleum.[2] The United Kingdom took 16.8 per cent of her exports, Canada 15.1 per cent, Japan 7.8 per cent, and no other country as much as 5 per cent. On the other hand, 13.2 per cent of her imports came from Canada, 6.5 per cent from Japan, 6.0 per cent from the United Kingdom, 5.7 per cent from British Malaya, 5.4 per cent from Cuba, and 5.0 per cent from Brazil.[3] Thus we see that the trade with Canada is by far the most important part of the foreign trade of the United States.

[1] "The Network of World Trade," League of Nations, Geneva, 1942, p. 52. This book contains a vast amount of suggestive statistical material describing the world trade of the United States and also of other nations in the period just before the second World War.

[2] *Ibid.*, p. 32.

[3] These figures are for 1938, the last fairly normal year before the second World War. They are taken from the League of Nations publication, "International Trade Statistics: 1938," p. 109.

In 1938, the most important articles of import were coffee and tea, which made up 9.6 per cent of the total imports, paper and paper pulp 9.5 per cent, sugar 7.2 per cent, and rubber 6.9 per cent. Among the exports machinery accounted for 28.0 per cent (automobiles and other vehicles alone amounting to 13.0 per cent), fuels of various sorts accounted for 14.8 per cent, textile fibers including cotton for 7.9 per cent, cereals for 6.4 per cent, iron and steel for 5.8 per cent, and tobacco for 5.5 per cent. These few enumerated classes of commodities accounted for well over two-thirds of all our exports.[1]

23 · 9. Interferences with International Trade. As has been mentioned before, foreign trade and domestic trade are fundamentally alike, yet there are some important points of difference. Although even domestic trade has never been entirely free and unrestricted, it has been far freer than foreign trade. From the earliest times, governments have taken an especial interest in the regulation of foreign commerce, sometimes with the aim of raising revenue but often with the idea of restricting the trade itself. Such restrictions are traceable in some cases to the old mercantilist idea that a country benefits by increasing the amount of gold and silver within its borders. Such a belief leads men to adopt policies that encourage exportation and discourage importation of goods, so that the balance will be paid in money. On the other hand, restrictive policies are often the result of selfish pressure by persons who expect to benefit from them. It is evident that a domestic businessman and his employees may gain by policies that put foreign competitors at a disadvantage. Even if the general buying public suffers a loss by the restriction of foreign competition, the domestic producer may make a real gain. It is natural for "home" producers to urge consumers to patronize "home" industry. The philosophy may take the form of urging buyers to favor American-made products rather than foreign goods; or it may, when addressed to a more localized audience, urge them to buy New England products, Kansas products, or Chicago products. The local merchant in the small town tries to persuade his fellow townsmen to "trade at home" and not send their money to the neighboring cities or the mail-order houses. Such a doctrine of provincialism or isolationism is obviously of advantage to the local merchant, but it may well be disadvantageous to the buyer who can obtain goods of higher quality or at lower prices farther afield.

A quotation, often erroneously ascribed to Abraham Lincoln, expresses well a popular misconception on the subject of trade and interferences with trade. The quotation runs, "I don't know much about the tariff, but I do know this much: If I buy steel rails in England, we have the rails and they have the money; but if I buy steel rails in the United States, we have both the rails and the money." This aphorism has the convincing quality that goes with simplicity and at least superficial "logic." But it will pay the student

[1] "International Trade Statistics," *op. cit.*, p. 110.

to examine it a little more carefully and try the same sort of "logic" in more extreme cases where the nature of the misconception is more obvious. Suppose, for example, that a man living in Alaska says, "I don't know much about the tariff, but I do know this much: If I buy cotton from Texas, we have the cotton and they have the money; but, if I buy cotton grown here in Alaska, we have both the cotton and the money." One sees at once that the statement is entirely true, but that the cotton which one has in the latter case will be very small in amount and very expensive. To be sure, Alaskans can grow cotton if they wish to invest heavily enough in greenhouses, but cotton obtained that way is very costly indeed. Alaskans can get more and better cotton by producing things in which they have a local advantage (or in which their local disadvantage is less marked than it is in the case of cotton) and by trading those things for cotton. It is just another case of comparative advantage, with which we are fully familiar from earlier study.[1] If the foreign territory is less efficient in production than we are, we need not be protected against them. We can already produce at lower costs and undersell them. If the foreign territory is more efficient in production than we are, then we make a real gain by letting them produce the things in which they have the advantage while we in turn specialize in the production of other goods. If, in the case of the quotation attributed to Lincoln, American steel rails are cheaper and better than English steel rails, there will be no need for protection of our home industries. We shall not be importing rails anyway. But, if steel rails can be produced more cheaply in England than in the United States, we can reword the quotation to say, "I don't know much about the tariff, but I do know this much: If I buy steel rails in England, we have more and better rails than we could otherwise get and they have the money; but, if I buy steel rails in the United States, we have to get along with poorer and more expensive rails for our money."

23·10. Kinds of Interferences with Trade. Governmental interference with foreign trade takes many forms. Sometimes they are open, obvious, and acknowledged; sometimes they are hidden, obscure, and disclaimed. They may be entirely informal and unofficial, relying on the pressure of public opinion rather than the force of law. They have often paraded in the guise of health regulations which purport to protect the public while actually protecting domestic manufacturers. By far the most common of the open efforts to restrict foreign trade have been tariffs. An import tariff is a tax levied on goods brought into a country, and of course the necessity of paying the tax tends to discourage the importation. Export duties are prohibited in the United States by the Constitution, yet absolute prohibition of exportation is allowed and occasionally practiced, as when the United States forbade

[1] See Secs. 10·10 *ff.*

the export of gold in 1933. Import duties, on the contrary, have been common, and will be discussed in detail in Sec. 23 • 11.

We discovered in Sec. 17 • 6 that businessmen sometimes find it advantageous to sell goods in foreign trade at prices below the full production cost, covering only the variable expense on their foreign sales. *Dumping on a foreign market* is the practice of selling goods abroad at prices lower than those charged in the country of origin. This particular sort of international trade is often regulated by special antidumping legislation, which forbids the sale of imported goods at prices below those charged in the country of origin. This is, of course, a very special sort of interference with international trade and does not affect commerce except insofar as the practice of dumping is involved.

The years between the two great World Wars witnessed the invention and use of several new and effective methods of controlling international trade. Some countries set up trade quotas, specifying just how much of various goods could be imported and from what countries or areas. Others accomplished the same purpose by regulation of foreign exchange. As we shall see in the next chapter, purchases of foreign goods require the use of special sorts of documents that we call "foreign exchange," and if the government limits the amount and availability of foreign exchange it regulates indirectly but very effectively the purchase of foreign goods. The tendency during the interwar years was for governments to become more and more powerful, for men to accept more readily the assumption that one should seek the advantage of the nation rather than of the individual citizen, and that the two advantages were in practice separable. The aggrandizement of government was very evident in the growth of fascism and communism, and only less so in various movements within the democratic countries. International trade was one of the most common tools used by all countries in their efforts to gain national advantage, and interferences with the normal flow of trade came to be more common and more onerous than they had been for a century and a half.

Most countries pursue policies that fall somewhere between the two extremes of complete freedom of trade and *laissez faire* on the one hand, and complete nationalism, isolationism, and national self-sufficiency on the other hand. The second World War did much to bring political isolationism into disrepute even with people who continued stanchly and inconsistently to cling to policies of economic nationalism and isolation. Economic and political policies and programs are so intricately interwoven that it is in practice impossible for a country to follow policies that are fundamentally conflicting in the two spheres, and it will be a matter of interest to see whether postwar developments tend to match the present economic nationalism

with an increased political isolationism, or to match the present political internationalism with an increased freedom of trade.

23·11. Kinds of Tariffs. Tariffs are usually classified according to their purpose into two main groups: *revenue tariffs,* which are intended as sources of income for the government; and *protective tariffs,* which are intended to protect domestic producers from competition of foreign products by levying an import duty and thus raising the price at which foreign goods can be sold in the domestic market. Many tariffs actually do both these things. They raise the price of imported goods somewhat, thus giving domestic producers some advantage, and they raise some revenue for the government by taxing the foreign goods that actually come in. Yet it is important to realize that in their fundamentals these two kinds of tariff are almost complete opposites. They are laid on different types of goods and at very different levels. A real good revenue tariff provides little or no protection to domestic producers, and a real good protective tariff yields little or no revenue.

To understand this difference a little more clearly, we might divide all the goods that are consumed by our people into three major classes: (1) goods that can be produced in the United States at considerably lower cost than in competing countries, (2) goods that can be produced in foreign countries at considerably lower cost than in the United States, and (3) goods that are produced in some foreign countries at about the same cost as in the United States, but in no country at very much below our cost. For example, some grades of cotton, automobiles, and farm machinery are illustrations of the first of these groups; tin, silk, and coffee of the second group; and some grades of textiles, wool, etc., of the third group.

If we want to raise revenue from a tariff, we ordinarily select some commodity of the second group and, if possible, one that is characterized by a rather inelastic demand. The reason is apparent. If we lay a small tax on the importation of such a good, people will continue to import it, because there is no local industry to take its place and consumers will still buy considerable quantities even at the higher price that follows the tax. If we set the duty too high, consumption will be curtailed and, if the goods are not imported, there will be no imports to be taxed, and we shall raise no revenue. We therefore keep the tax low and confine it to goods that are almost certain to be imported—goods that can still undersell domestic production even after the tax is paid. Thus, if we wished to raise revenue in the United States from tariffs, we would be apt to lay low import duties on such commodities as tin, rubber, or silk. Some countries have laid revenue tariffs on goods of the third class—goods that are produced at approximately the same cost at home and abroad. In such a case the government, if it is really seeking revenue rather than protection, lays an internal revenue tax on goods

produced domestically which offsets the tariff. For example, the United States might lay a tariff of five cents a pound on the importation of wool and, at the same time, an internal revenue tax of 5 cents a pound on all wool produced within the country. Under such circumstances, the domestic producer would be given no advantage, and importation would not be discouraged to any greater extent than the local product. England for many years followed such a policy of revenue tariffs that were offset by equal internal revenue taxes.

On the other hand, if we want to protect domestic industry by a tariff, we naturally choose some industry needing protection. There would be no sense in laying a tariff on the importation of tin in order to protect or encourage our own domestic tin industry, because we have no such industry and could have none without first discovering domestic sources of that metal. Likewise, if we can produce corn or tobacco far more cheaply than other countries, our producers of these products need no protection, and a tariff will give them none. Even if we pass a tariff on such a commodity, it will have no effect. Without a tariff, we would export such goods, not import them. We do not need a tariff to cut off imports, since we have no imports to cut off. One of the semihumorous parts of our recent tariff history has been the fact that Congress has laid tariffs on a number of products that we were already exporting, in order that the politicians might tell their constituents that they, too, had been given tariff protection. The sad aspect of the affair is that the producers of the export products seem to have been gullible enough in many cases to feel some real sense of obligation. If foreigners have a tremendous cost advantage in the production of an article, we usually have no domestic industry to protect; if we have a tremendous cost advantage, protection does us no good. Protective tariffs are effective only in those cases in which production costs are approximately the same at home and abroad, so that the increase in cost, which the foreigner assumes in the import tax, puts him at enough of a disadvantage so that custom is shifted to domestic producers. It is also clear that the shift will be most marked in the cases of those commodities the demand for which is elastic, since a given increase in the price of the imported goods will reduce markedly the quantity of those which can be sold. Thus, we see that protective tariffs are apt to be high tariffs laid on commodities of elastic demand, which are produced at home and abroad at approximately equal cost. Revenue tariffs, on the contrary, are most successful if they are rather low tariffs laid on commodities of inelastic demand, which are produced abroad much more cheaply than they can be produced at home.

In addition to classifying tariffs according to their purpose into revenue and protective tariffs, we also often classify them according to the way in which the amount of the tax is computed. If the tax is laid according to the

number of units of the commodity imported—such as a tax of $5 a ton, 3 cents per pair, $1.50 per dozen, or 35 cents per bushel—we call it a "specific duty." If, on the other hand, it is laid according to the value of the goods imported—such as a tax of 3 or 10 per cent of the value—we call it an "ad valorem duty," which merely means "according to value." It is evident that ad valorem tariffs and specific tariffs may be considerably different in their effect. For example, if two merchants import the same quantity of cloth, say, 1000 yards each, and if the duty is a specific one, they will both pay exactly the same import tax even if one of them is importing low-quality cloth while the other is importing high-quality cloth. Under these circumstances, however, if the tax had been an ad valorem tax, the importer of the high-quality goods would have paid a higher tax than the importer of the low-quality goods even though the quantity imported was the same. Evidently specific taxes favor the importation of high-priced rather than low-priced goods. Also, it is easy to see that during a time when the general price level is rising, the effect of specific tariffs gets less and less. If we have to pay a tax of, say 25 cents per unit imported, the tax will be proportionately less important when the price of the commodity is $10 than when it is $5. It is often overlooked that times of generally rising prices are periods during which tariffs become less and less important as impediments to international trade, while during periods of falling prices the interference of tariffs becomes more pronounced. This is especially true because of the fact that a large proportion of all tariffs are specific rather than ad valorem.

23 · 12. Effects of a Protective Tariff. The effects of a protective tariff can be visualized easily either by a numerical example or by a diagram. Table 23 · 2 is a hypothetical example showing the demand schedule for a commodity in the United States and also certain facts about the foreign and domestic supply. We note from it that the quantity demanded by the people of the United States falls as the price rises, in accordance with the Law of Demand. We note also that increases in the price stimulate both domestic and foreign producers to offer larger quantities for sale in the domestic market. Of course, if we lay a tariff, we reduce the quantity of foreign goods offered for sale at any particular price but do not affect the quantity of the domestic goods offered. If, as in the example of Table 23 · 2, we lay a specific tax of 20 cents per unit imported, foreign producers will have to get a price 20 cents higher than before to stimulate them to send the same quantity. Thus the table shows that, when the price was 40 cents per unit without any tariff, foreigners shipped 485 units; but with a 20-cent tariff they had to get a price of 60 cents per unit before they would ship the same quantity. The total quantity offered for sale is the sum of the domestic and the foreign offerings.

Let us now apply our familiar methods to the problem, beginning with the case where there is no tariff. If the price of the commodity is 20 cents per unit, 620 units will be demanded and only 190 supplied. If the price is 40 cents per unit, 390 units will be demanded and 745 supplied. The former price is too low, and the latter price too high. If we select a price of 30 cents per unit, we find that 490 units are demanded and 500 supplied. This price, also, is a little too high. Evidently, the price will be between 20 and 30 cents, and almost as high as 30 cents. If we want more accuracy, we may plot the curves that correspond to the demand and supply schedules of the table. Such curves are shown in Fig. 23·6. In this figure are shown the demand curve, D, the domestic supply curve, S_D, the total supply curve when there is no

Table 23·2. Hypothetical Demand and Supply Data to Illustrate the Effect of a Tariff

Price per unit	Quantity demanded	Quantity offered for sale					
		Without any tariff			With specific tariff of 20 cents		
		Domestic	Foreign	Total	Domestic	Foreign	Total
$0.10	820	000	000	000	000	000	000
0.20	620	000	190	190	000	000	000
0.30	490	150	350	500	150	000	150
0 40	390	260	485	745	260	190	450
0.50	300	360	600	960	360	350	710
0 60	230	445	715	1160	445	485	930
0 70	165	525	815	1340	525	600	1125
0.80	110	600	890	1490	600	715	1315
0 90	65	660	960	1620	660	815	1475
1.00	30	700	1010	1710	700	890	1590

tariff, S_{T1}, and the total supply curve when there is a specific tariff of 20 cents, S_{T2}, all taken directly from Table 22·2. If we consider first the case without a tariff, the demand curve and the total supply curve on the American market are D and S_{T1}, These cross at a price of approximately 29 cents and a quantity of approximately 490 units. By our old familiar method of analysis, we find that under competitive conditions the price of the commodity will be about 29 cents and American consumers will buy approximately 490 units. But the American supply curve shows that at the price of 29 cents, American producers will supply only 145 units of the product. If consumers are to get a total of 490 units, the remaining 345 must come from abroad. We can summarize the situation without a tariff by saying that the price will tend to be 29 cents, and that at this price consumers in the

American market will buy 145 domestic and 345 foreign units of product, or a total of 490 units.

If we lay a 20-cent specific import duty, the domestic supply curve stays the same, but the foreign supply curve and the total supply curve are altered. Looking again at Fig. 23·6, we see that the demand curve and the new total supply curve, S_{T_2}, cross, approximately, at a price of 38 cents and a quantity of 400 units. At this new price, domestic producers will supply about 240 units and leave some 160 units to be imported. If consumers pay 38 cents per unit for 400 units, they will spend a total of $152 with the tariff. Before the tariff was laid, they were paying 29 cents per unit for

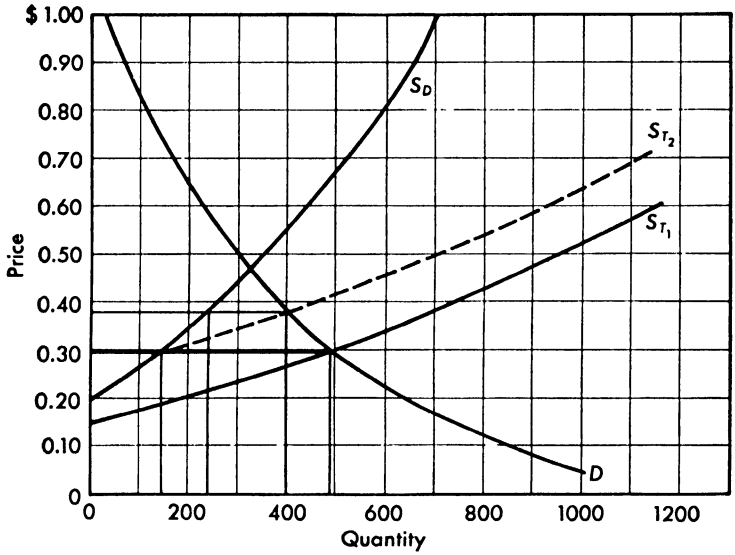


FIG. 23 · 6.—Effect of a protective tariff.

490 units, or a total of $142.90. With the tariff, they spend a slightly larger amount than before but get only 400 instead of 490 units. The government, on the other hand, receives an income of 20 cents on each unit imported, or a total revenue of $32. The tariff reduced imports from 345 to 160 units and increased domestic manufacture from 145 to 240 units. Domestic producers have made a very real gain at the expense of foreign producers *and at the expense of domestic consumers.* It is easy to see why domestic producers of a competitive commodity favor a policy of protective tariffs, but it must be remembered that their gain is made at the expense of the general public.

It should be noted that, although the price of the product was raised by the imposition of the import duty, it was raised by less than the amount of the duty. The duty was 20 cents, and the price rose but 9 cents, from 29 to 38 cents. One might expect the price to rise by the full amount of the

tax, and so it would if there were no domestic sources of supply. But in our problem (as in any case of an actual protective tariff) the rising price stimulated increased domestic output, throwing somewhat larger domestic quantities on the market to replace some of the foreign product that was kept out, and this extra domestic output kept the price from rising as far as it otherwise would have done.

One might experiment with other problems similar to that depicted in Table 23·2 and Fig. 23·6; but he would find that the general tendency in all such cases is, when one imposes a new protective tariff or raises an old one, for

1. The domestic output to rise.
2. The quantity imported to fall.
3. The price of the commodity to rise, but by somewhat less than the amount of the tax.
4. Consumers to buy a smaller quantity of goods and to pay a higher price per unit for them.

23·13. How Tariffs Are Made. In the United States, tariffs are either established by Congress or by agencies authorized by Congress. Tariff bills originate in the Committee on Ways and Means of the House of Representatives, which holds hearings at which interested parties may speak. Such bills are usually very long and complicated, and it is usually true that even most members of the committee are conversant with only some sections. After hearings and discussions within the committee, a bill is drawn up which is usually a compromise of the views of various interests,[1] and this bill is then reported to the House of Representatives. It is now open to amendment and ultimately to either adoption or rejection. If adopted, either in its original or more commonly in an amended form, it is sent to the Senate, where it is referred to the Finance Committee. This committee opens the entire matter to discussion anew, holding new hearings of its own and often rewriting large parts of the bill, so that the measure may be radically different when it leaves the committee from the measure which passed the House. The Finance Committee reports its version of the bill to the Senate, where it is again open to amendment, and then to adoption or rejection. If the bill is adopted, it is almost certain to have been altered in many respects from the form in which it was passed by the House. In such a case, where the House and the Senate have passed different versions of a bill, the two versions are sent to a Conference Committee, made up of members of

[1] For some reason that is difficult to explain, the statement that a bill is "a compromise of the views of various interests" seems to convey a sinister meaning to many people unfamiliar with legislative procedure. The statement here is not intended to be adversely critical. In a democracy, most legislation is the result of compromise, and the legislative process itself is largely organized for the purpose of effecting compromises.

both houses. This committee tries to work out further compromises to satisfy both houses and, if they agree, they report the new compromised bill to their respective houses. It is still possible, of course, for either or both of the houses to reject the report of the Conference Committee; but ordinarily the conference report would be accepted by both houses and the bill in its new compromised form sent to the President for his signature. Only after he has signed it does it become a law. Thus the steps through which the ordinary tariff bill would pass from the time when it was introduced by a member of the House of Representatives until it became a law would be as follows:

1. Introduced in the House by a member.
2. Referred to the Committee on Ways and Means.
3. Hearings held.
4. Reported by the Committee to the House of Representatives.
5. Debated, amended, and adopted by the House.
6. Sent to the Senate.
7. Referred to the Senate Committee on Finance.
8. Hearings held.
9. Reported by the committee to the Senate.
10. Debated, amended, and adopted by the Senate.
11. Sent to a Conference Committee.
12. Reported back to both houses and adopted.
13. Signed by the President.

It will be seen that there is little chance for a measure to go through all these steps unaltered. Usually the bill that is finally adopted differs radically from the one that was originally introduced. Many men with many minds and representing many divergent interests take their part in drafting the bill, and it is a common complaint that the resulting legislation is apt to be a hodgepodge of unrelated details rather than a coherent set of related parts. It has long been urged that we should adopt some more orderly and "scientific" method of drafting tariff legislation, and in the United States several agencies have been set up, one after another, to assist Congress with the intricate and technical problems involved in drawing up a general tariff bill. Most recent is the Tariff Commission, set up in 1916. This is to a considerable extent a research agency, finding facts on which Congress or the President may base action. The most recent general revision of the tariffs of the United States is the Smoot-Hawley Tariff of 1930, one of the highest protective tariffs in the history of this or any other country. The present law makes provision, under what are called "flexible provisions," for reductions in these tariffs by not to exceed 50 per cent, such reductions to be made by the President. Under the Trade Agreements Act of 1934, the administration has taken the initiative in trying to lower tariff barriers by means of

reciprocal trade agreements. Under this plan, the United States seeks to make agreements with other nations, under which we offer to reduce certain of our import duties in return for similar concessions on the part of the other country that is party to the agreement. The effect of this plan is extended over much broader territory by a provision specifying that any reductions made in bargaining with one country shall be applicable to all countries. If we lower our tariff on textiles in an agreement with Great Britain, we must also lower it for textiles coming from France, even though we have entered into no specific agreement with France. The United States has for many years followed a policy of high protection, having tariffs among the highest in the world. The flexible provisions and the reciprocal trade agreements have combined to bring some reduction of tariffs, where the experience with Congress seems to indicate that such reductions would have been extremely unlikely had the matter been left to the legislative rather than the executive branch of the government.

23 · 14. Protection vs. Free Trade: The Case for Protection. The argument between those who favor protection and those who favor free trade is a hardy perennial in American politics. It was long a party and a sectional issue, with the Republican party and the industrial sections generally favoring protection, and the Democratic party and the agricultural regions favoring free trade. Like any other generalization, this one has its exceptions, and as the Democratic party has come to be more closely identified with the industrial workers of the big cities it has become less and less applicable. Today it is difficult to find any important difference between the parties on this issue. Yet the dispute continues.

It should be made clear at the outset that the dispute does not center around the question as to whether we should or should not have tariffs. It is necessary for governments to raise revenue, and one convenient source of revenue has always been the taxation of imports. Even the most enthusiastic free trader may well favor revenue tariffs, which are either laid on commodities not produced at home or offset by internal revenue taxes of equal amount. Our present problem is definitely that of protective tariffs. And the arguments that are brought forward in support of such measures are numerous and diverse. They include some of the most widely accepted fallacies in the whole field of economics, as well as two or three arguments that are generally accepted as valid under special circumstances by most students of the subject.

First, we have the argument of *military necessity.* It is pointed out that, if we specialize in accordance with the Law of Comparative Advantage, there will be many commodities that we shall not produce at all. These commodities will be imported and, as long as we can get them in exchange for our own surpluses, we get an advantage from the trade as was pointed

out in connection with our discussion of that law. But, if our trade with foreign countries is broken off for any reason—strikes, storms, floods, or wars—then we find that we have a surplus of some commodities over and above what we can use and a deficit of other commodities that we had intended to import. War is the most important factor historically in breaking off established trade between nations. Table 23·3, for example, shows the reduction in prewar international trade of various areas which took place as a result of the outbreak of the second World War. We see that these areas lost anywhere from 13 to 80 per cent of their prewar trade. This would be inconvenient at best but, if the commodities which are cut off are materials of war, necessary for the national defense, then the loss may well be fatal. For this reason many prudent people urge that, where necessary, we should stimulate domestic production of such materials as are necessary for

Table 23·3. Proportion of Prewar Foreign Trade Cut Off by Second World War*

Country or region	Percentage of trade lost	Country or region	Percentage of trade lost
Continental Europe	46	Japan and southeast Asia	63
Rest of Europe	35	Rest of Asia	44
United States	42	Africa	54
Rest of North America	13	Oceania	25
Latin America	33	World as a whole	43
U.S.S.R.	80		

* Estimates from *Federal Reserve Bulletin*, vol 31, no. 9, September, 1945, p. 887.

the prosecution of war by protective tariffs rather than rely on foreign sources of supply. This argument is pretty generally accepted, although it should be noted that it will apply to relatively few commodities, especially in a country with the rich and varied resources of the United States. In such cases, we should realize that we are taking a continued loss in times of peace in order to protect ourselves against possible greater losses in time of war. Under such circumstances, it might be better for the government to support the desired industries by the grant of a direct subsidy on domestic production, since then the cost of the program would at least be open and obvious.

Many people also advocate the protection of *infant industries,* their argument being that a new industry, just starting out, is apt to have high costs until it has learned the techniques, trained its workers, established its markets, and built up its volume of output. If such an industry were left unprotected, the argument goes, it would be killed off by the competition of concerns already established abroad; but if it is protected from foreign competition until it has established itself, it may be able to reduce its costs

to the place where protection is no longer needed. This infant-industry argument is also generally accepted as valid, although it is much more important for a young country without industry than it is for an older more diversified and industrialized one. It should also be clear that the infant-industry argument is an argument for *temporary* tariffs. If the industry in question never becomes efficient enough to withstand foreign competition, then this argument is invalid. In practice, it has long been remarked that infant industries never seem to grow up, that they clamor for continued tariff protection and argue that they now have a vested interest in a continuation of the tariff. They assert that they have been built up under protection and that their workers have acquired skill and their owners have invested money, which will be lost if the protection is withdrawn. This outcome should be considered before the original policy is adopted.

Other advocates of protection propose it as a *safeguard against dumping*. This problem has already been described at length (see Sec. 17·6), and we have seen that foreign producers may be able to undersell domestic producers, not because they are more efficient but because they are neglecting their fixed costs. If we could be sure that such an arrangement would be permanent, there would be no reason to object to it. If the foreign consumers can be made to pay the fixed costs so that we can enjoy the commodity by paying only the variable costs, these foreign consumers have been forced in effect to subsidize our consumption of the product. We are gaining an advantage at their expense. But the neglect of fixed costs is apt to be temporary, and we may find, after our domestic producers have been thrown out of business by such competition, that the foreign producers raise the price again. Therefore some people have advocated the imposition of protective tariffs to penalize foreign low-cost goods. It is easily possible, however, to meet this problem without the imposition of tariffs. The United States already has antidumping legislation which prohibits the sale of foreign goods in this country at prices below those charged in the country of origin. Such a law does not penalize the legitimate foreign seller who is not dumping, as tariff legislation would do, but does take care of the dumping aspect of the problem.

Some people have advocated protective tariffs on the ground that we can thus *conserve our natural resources* by artificially turning our productive efforts into channels that do not make use of our own supplies. They tend to overlook the fact that the problem of conservation is largely a domestic one and that all nations could not pursue at the same time a policy of using someone else's natural resources. Other advocates of protection have argued that misunderstandings over foreign trade lead to war and that in the *interests of international amity and peace* we should abolish international trade. It would be quite as easy, and perhaps even more realistic, to

argue that international misunderstandings have been bred from interferences with trade and that tariff misunderstandings have been among the commercial misunderstandings that have been so fruitful of war in the past.

Many of the arguments for protection go back to the sophistry of mercantilism and to the medieval failure to understand the real gains of specialization and trade. Thus, we still run across the argument that importation of goods *sends money abroad* and that it should be stopped for that reason. The people who suggest this argument fail to understand that imports are the real payment for our exports—the only real excuse for sending exports abroad in the first place.

Perhaps the commonest misunderstanding in modern times with regard to the protective tariff is the supposition that it protects high-paid American labor from *competition with cheap foreign labor.* This argument can be heard in almost any political campaign in the United States and is evidently accepted as correct by a very considerable number of workers and employers alike. It has become ingrained among the intuitive and unreasoned political beliefs of the time, along with equal fatuities concerning monetary and labor problems. If the argument held water—if, in fact, high-paid American labor were unable to compete with sweated coolie labor of foreign lands—it is obvious that the United States would be unable to export at all. Yet she is one of the greatest exporting nations of all time, exporting goods made by high-paid labor and selling them in foreign lands in competition with goods made by low-paid native labor. The student who thinks the matter through will realize that high weekly or monthly money wages do not necessarily spell high production costs. If American laborers are paid twice as much, let us say, as Italian laborers but produce three times as much because of greater efficiency, then the cost of production is lower in the United States than in Italy despite the higher American wage rate. American high wages are a result of high American productivity—high in part because of the ingenuity and industry of her workers and in part because of the rich resources with which they toil. These advantages were not given to them by a tariff, nor can they be taken from them by the lack of a tariff. It is a strange commentary on the popular misunderstanding of economic matters that at the very moment when American workers are asking for tariffs to protect them from the pauper labor of foreign lands, these same low-paid foreign laborers are asking their governments for tariffs to protect them from the efficiency of American workers with their rich resources and modern methods! To be sure, it is easy to see that the employer or the employee in a particular inefficient local industry may gain by protection. He continues to find employment for his capital or his labor in a field that would be closed to him in foreign products could come to market freely. But he is at the same time kept from transferring his capital or his labor to

some other industry where it could be employed to even greater advantage. The protection against cheap foreign labor is a real protection in such isolated cases, giving advantage to the people in the inefficient industry to the disadvantage of their fellow countrymen who buy products at unduly high prices.

There are also those who support the protective tariff because it promotes *national self-sufficiency*. There can be no doubt but that the protective tariff does tend in this direction—just as there is no doubt but that the tendency is in ordinary cases an undesirable one. National self-sufficiency always involves national loss and, although that loss may in some cases be offset by a corresponding gain (as in the case of military necessity), such offsets are the exception rather than the rule. If self-sufficiency is desirable for a nation, why not for a state? Would it be desirable for Kansas, Nevada, and Idaho to be self-sufficient? If desirable for a state, why not for a county or a township? Would it be desirable for Duluth or Seattle to be self-sufficient—to grow their own cotton, mine their own ore, and raise their own coffee and rubber? And to carry our *reductio ad absurdum* to its "logical" conclusion, if it is desirable for a city to be self-sufficient, why not for a family? Why is it not desirable for the Carter or the Taylor family to be self-sufficient, producing everything that they consume? The answer is obvious: They would have far less to consume. Specialization and trade increase the amount of the product and, as long as the trade is uninterrupted, there is a net gain thereby for all concerned. Self-sufficiency is an unreasoned fetish that leads men to pursue their own disadvantage. Protective tariffs, to be sure, do increase a nation's self-sufficiency, but this is in ordinary cases an argument against them rather than an argument in their favor.

Of the arguments listed in this section only the military-necessity argument and the infant-industry argument seem worthy of serious consideration. They do have some validity and lead even ardent free traders to realize that "circumstances alter cases," and that there may be times when protection works for the social advantage. Yet these occasions are few and far between. The great and complicated structure of the American tariff system has obviously been built largely on the sophistries of protecting the "full dinner pail" of American workers against foreign low-paid labor and "keeping money at home" and "developing the home market." It is a structure whose architect is a misguided and uniformed Intuition rather than a thoughtful and illumined Reason.

23 · 15. Protection vs. Free Trade: The Case for Free Trade. The case in favor of free trade can be set down in much less space than that given to the case for protection. First and by far most important is the argument based on *comparative advantage*. Duluth and Seattle may be self-sufficient, growing their own cotton, mining their iron, and raising their coffee and

rubber, but they do this at great cost. Their people are poorer because they did not understand the advantages of specialization and trade. Under a system of free trade, the forces of competition tend to drive the factors of production into the most efficient industries and yield more goods for all concerned. The tendency is not toward equality of all peoples concerned, at least if the factors of production are immobile. If labor, capital, and natural resources tend to stay where they are, unequally distributed, then labor will be more productive in some parts of the world than in others and will tend to be higher paid in the places where it is more productive than in the places where it is less so. But, with the unequal distribution of productive factors giving rise to differences in cost of production, there will be advantages to all parties concerned if they specialize and trade. Trade usually springs up voluntarily, because the people who carry it on get an advantage from it. Those who interfere with your trade and mine ordinarily are working to our disadvantage, even if perhaps to their own advantage. (Even the latter is by no means certain, so great are the misunderstandings on the matter of trade.)

There are other arguments also in favor of free trade. There are those, for example, who point to the *political corruption* that arises in the passage of tariff legislation and urge that the social costs are far greater than any gains. Tariff legislation is one of the greatest of the "pork barrels" of American politics, notorious for its logrolling, its "deals," and its interplay of sectional and group interests at the expense of the public interest. The individual businessman sees a chance to get personal benefit, even at the expense of the public welfare; the stakes are great and the temptation difficult to withstand. There are also those who advocate free trade as a weapon against *monopoly*. They argue that the protective tariff is "the mother of trusts" and that, by eliminating foreign competition, the tariff makes it possible for domestic firms to establish themselves in a monopolistic position to the disadvantage of the domestic consumer. This is a hard charge to prove, and it may be as well for us to pass it over without basing the case for free trade too heavily upon it. There are also those who argue that free trade is necessary for *world peace* and that nationalism and isolationism are things of the past, promising a false security, which they cannot yield. We have already noted the impossibility of setting up a system of political internationalism alongside a system of economic provincialism. Between the two great World Wars the world saw a growth of nationalism on a scale that had not been seen for a century and a half. If the promise of intertional understanding and cooperation with which the second World War closed is kept, it is almost certain that the world is entering on an era of greater freedom of trade. If, on the contrary, the nations cling to their

nationalism in economic matters (and pressures in that direction are already strong), there is little sense in talking of political cooperation.

23·16. Brief History of American Tariffs. During its early history, the United States relied on the tariff as its major source of income. The revenue aspects of the tariff were consequently matters of considerable interest. The very first act passed by the new Congress set up under the Constitution was the Tariff of 1789, an act establishing low ad valorem duties at an average rate of about 5 per cent. These rates were increased slightly in later years without any thorough revision of the basic law until the outbreak of the War of 1812. Under war conditions, the government, needing increased revenue, approximately doubled the rates of the existing tariff and, immediately after the war, in 1816, adopted a bill that is often called the "first protective tariff" in our history. It is true that the revenue aspects were still important. As a matter of fact, roughly 90 per cent of the government's income still came from import duties. The people had become prejudiced against direct taxes before and during the Revolutionary War, and the prejudice still lingered. But the Tariff of 1816 raised the rates to an average of perhaps 20 per cent. With foreign sources of supply cut off by the war, new industries had sprung up. Now that the war was over, they clamored for protection against their older and more firmly established competitors abroad. This tariff inaugurated a period of roughly 20 years in which the protective idea was more or less prominent in our tariff acts. In 1818 rates were again raised slightly, and business losses incurred during the crisis of 1818-1819 led businessmen to urge even higher rates. These they got in the Tariff of 1824. It was at this time that the tariff became a party issue, which it has remained more or less strongly ever since. In 1824, the Democratic party espoused the idea of "tariff for revenue only," while the National Republican party[1] stood for protection. It was generally felt that the Tariff of 1824 worked to the advantage of the North and the disadvantage of the South, although strangely enough the people of New England were either lukewarm toward it or in active opposition.

In 1828, Congress made another general revision of the tariff. The matter unfortunately became badly involved in the more partisan aspects of politics, many amendments being offered by enemies of the bill in the hope that they would make it so obnoxious that it would fail of passage, and others in the hope that they would embarrass the President and possibly prevent his reelection. As a result, the final bill was generally called the "Tariff of Abominations" and was unpopular from the start. It brought the level of protection

[1] The present Republican party was still, of course, a thing of the future. It did, when organized, adopt the protective program of the National Republicans and maintained it for two generations or more.

to its highest mark prior to the Civil War, and it was bitterly opposed by the Southern states. In 1830 and again in 1832, there were minor reductions in the rates, but not enough to appease the South, and in 1832 the state of South Carolina finally adopted the policy of nullification, taking the official stand that Congress had no right to adopt tariff legislation affecting the states, and that the law was null and void within the state. The matter became so serious that Congress in 1833 adopted the Compromise Tariff, which called for a gradual decrease of the rates each year for a period of 10 years. In 1842, when the Whigs came into power, they did make some slight increases in rates, but in general the rates were still far lower than those of the Tariff of Abominations.

In 1846 the Walker Tariff[1] attempted to return to the idea of tariff for revenue only. It swept out what had come to be called the "American System"—a system of protective tariffs coupled with internal improvements —for a period of perhaps fifteen years, or until the Civil War. This tariff remained in force for a much longer period than most American tariff measures, partly because the same political party stayed in power and partly because of rising prices, general prosperity, and an ample government revenue. There was a slight additional reduction of rates in 1857, which brought the general level down to about 20 per cent ad valorem.

In 1861, came the Morrill Tariff, which was a return to the idea of protection. In 1862, rates were again raised (to about 37 per cent ad valorem) and, in 1864, to about 47 per cent ad valorem.[2] There were minor adjustments, some upward and some downward, in 1867. There were some reductions in 1872, while in 1875 the need for more revenue as a result of the Panic of 1873 led Congress to make some increases in the rates. But the first real major revision after 1861 was the Tariff of 1883. Congress was embarrassed by an excess of revenue (a type of embarrassment that Congresses have been spared for many years now!) and in 1883 they made a real effort to reduce tariff rates. In 1890, the Republicans again raised the rates in the McKinley Tariff; but the Democrats were soon restored to power under Cleveland, who had advocated sharp tariff reductions. The Wilson-Gorman Tariff of 1894 did reduce duties about 10 per cent, but this was not enough to satisfy President Cleveland, and the bill became law without his signature. This bill is of some collateral interest since it contained a provision for an income tax which was later declared unconstitutional.

[1] Tariffs are commonly named after the chairman of the committee that drew them up. Sometimes both the House and the Senate chairmen are listed in the name of the tariff.

[2] The student will understand, of course, that the rates were very different on different commodities. Some might be taxed 1 per cent and others 10 per cent. Whenever we give figures as to the general level, it is an attempt to state a rough average.

The political character of the American tariff makes it an unpredictable and changeable thing. Cleveland stayed in power for a single term and, as soon as the Republicans came back, they restored the rates of the McKinley Tariff by the Dingley Tariff of 1897. This tariff tended in general to substitute specific for ad valorem duties, and the average rate seems to have been roughly 55 to 60 per cent—the highest ever reached up until that time. Like the tariff of 1846, the Dingley Tariff was destined to last for much longer than the average American tariff measure, aided by a period of rising prices and prosperity that tended to keep businessmen satisfied with conditions without proselyting for more help by the government. In the election of 1908, both parties promised some tariff reduction, and in 1909 the Payne-Aldrich Tariff made a slight move in that direction. It did not seem to be enough to stop a popular clamor that had grown up for tariff reduction—a clamor influenced to some extent by the current unpopularity of monopoly and the trust movement. President Wilson, elected when the Republican party was split by the Progressive or "Bull Moose" campaign of 1912, had compaigned on a low-tariff platform, and in addition he represented the Democratic party, which had long been the champion of "tariff for revenue only." In 1913, under his leadership, came the Underwood Tariff, which brought the greatest reductions in rates since the Civil War. The Republicans, back in power in 1920, passed an Emergency Tariff Act in 1921 which raised the rates again, and the Fordney-McCumber Tariff of 1922 put the rates back to their pre-1913 level. It was in this 1921 act that the flexible provisions were introduced, allowing the President to lower rates by any amount up to one-half in return for similar concessions by foreign countries.

The last major tariff revision was that of the Smoot-Hawley Tariff of 1930. This attempted to appease the farmers by laying tariffs at rather high rates on many farm products not protected before—many of them products that were actually heavily exported and could not be protected. In this sense the tariff was merely a political gesture. But it also raised many other rates and brought the general level of tariffs to one of its highest points in our history and among the nations of the world. The tariff was strongly denounced by economists and businessmen even before its final adoption and many people have felt that it contributed to both the terrible decline in business[1] and the international misunderstandings that followed. It is indisputable that many of the major commercial nations joined in a protest against it, and that its adoption was followed by a large number of retaliatory measures on the part of other countries. Although the Smoot-Hawley Tariff

[1] This decline had already started, of course, before the adoption of the tariff. At the worst, the tariff did not cause it, but magnified and enhanced it.

remains as our basic tariff law, even after many years of Democratic control,[1] Congress did in 1934 pass the Trade Agreements Act, under which the United States has made treaties with many other nations reducing the tariffs on specific commodities. There can be no doubt but that these reductions have been substantial; yet they still leave the United States definitely in the position of a high-tariff nation.

23·17. What Tariff Policy Is Desirable? As soon as we start to ask what is desirable, just, ethical, or good, we get out of the field of science and into the field of ethics.[2] An argument as to what sort of tariff is desirable is not one that can be settled beyond dispute by processes of reason. Yet every citizen finds it necessary to make up his own mind on such matters, and we should not entirely dodge the issue here. The student should realize that what is said in this section is a matter of opinion and not subject to proof; yet the attempt has been made to state in rather general and tentative terms some of the points on which there seems to be reasonable agreement among economists.

In the first place, we must distinguish carefully the problems of revenue and protective tariffs. No one seems to object seriously to revenue tariffs except as they dislike taxes in general. If governments are to operate, they must obtain revenues somehow, and the import tariff is a method rather easy to operate and one with long historical precedent. We shall turn our major attention, then, to the more disputed question of the policy of tariff protection.

No matter how avid a free trader one may be, he has to realize that a policy of free trade could not easily or safely be instituted in one fell swoop in a community that had for some time been following a policy of protection. If the protective policy has worked at all, it will have stimulated investment in the protected fields, and workers will have developed skills in these fields. A sudden removal of protection might easily result in the loss of the particular industry as far as domestic production is concerned and, although the long-run effect would be to transfer domestic labor and capital into more efficient industries, this would be done only at the expense of a short-run loss on the part of those in the industry. Society would make a long-run gain, and individuals would incur a short-run loss. If the transition could

[1] The disinclination of Congress to reduce the tariffs of the Smoot-Hawley bill even after 12 or 15 years of undisputed Democratic control is but one of several indications that the former party disagreement on tariff matters has in large measure disappeared. Free trade seems still to be popular in the deep South, and among many of the commercial sections that depend on international trade. In contrast with a former sitiuation, many large industrialists now advocate free trade because of the importance of their foreign business. On the contrary, the agricultural Middle West (largely Republican) and the industrial cities of the Northeast (largely Democratic) seem on the whole to favor protection.

[2] See Secs. 2·1 to 2·5.

take place gradually, over a period of, say, 15 or 20 years, the individual losses would be much less serious. Workers already specialized in the field would have time either to complete their productive life or to acquire new skills. Men whose capital was invested would have time to recoup on their investment while the capital was wearing out. Young workers, just acquiring their skills, would see that they should look to some other industry, and capitalists making new investments would realize that they might better make them in other fields. Presumably, if the United States were to adopt a free-trade policy, it should do it gradually, although it is possible that the disruptions of industry during the World War were such that the change could have been made then in one jump, since industries had already been shifted from their ordinary lines of output and had to go through a process of reconversion at any rate. They might have been able, after the war, to convert to efficient industries almost as easily as they could reconvert to their old protected ones.

In considering the tariff question it is also important for the people of the United States to realize that, partly as a result of the two great World Wars, this country has become a great creditor nation.[1] Other nations all over the world are our debtors. They cannot pay their debts in gold, since we have almost all the gold and since the gold would be of no particular advantage to us anyway except as we could buy things with it. Exchange is always ultimately an exchange of goods for goods, and only proximately an exchange of goods for money. If other nations are to pay us what they owe, they can do it only by sending us their goods; and they cannot send us the goods if we penalize imports by high tariffs. If Mr. Kinsey makes a loan to Mr. Waggoner and then sets up conditions that make it impossible for Waggoner to repay the debt, Kinsey will find that he has actually made a gift rather than a loan. Similarly a creditor nation either has to accept imports or allow repudiation of the debts owed it. In this regard the position of the United States has been changed radically in the past generation or two. Policies, which might have had some superficial semblance of sense when we were a debtor nation, have neither the appearance nor the actuality of sense under present conditions.

It seems reasonably certain that the nations of the world would gain through a uniform generalized policy of free trade if its continuation could be relied on. Its advantages are so great, that there would appear to be advantage even for single nations in pursuing unilateral free trade if necessary. Yet in real life nothing is certain and, even if the great and awesome forces released during the second World War compel the nations to establish some sort of agency to ensure world peace, realistic men will reluctantly agree that their hopes may be ill founded and that the time

[1] This fact is discussed in detail in Chap. 24.

may come again when distant sources of supply are cut off. The less important the commodities involved, and the less likely the issue is to arise, the less we need worry over the eventuality. But, in a world of practical men, one may well feel that complete unilateral free trade is unreasonable and that, although we would gain from it as long as it continued, it is necessary in a very few and unusual cases to continue a policy of protection on the grounds of military necessity. If this is to be done, it should be done openly and knowingly and with its purpose clear. It should be understood that the policy is one of protecting the nation, and that any protection of the industry or the people in it is but an incidental and accidental by-product. As was suggested earlier, in such cases it might well be preferable to grant the few industries involved a direct subsidy rather than the protection of a tariff; if the subsidy policy were followed, there seems to be good ground for advocating the gradual and ultimately the complete elimination of protection as a policy, leaving only low revenue tariffs on goods produced far more cheaply in other lands, or revenue tariffs offset by internal revenue taxes of equal amount.

Questions

1. "Natural resources have almost no mobility, and labor is only very imperfectly mobile." Do natural resources have any mobility at all? Why is labor imperfectly mobile? What sorts of policies might increase the mobility of labor?
2. "Imports, not exports, are the goal of foreign trade." Explain what is meant. Do you agree that it is true? Defend or controvert the statement.
3. What is a favorable balance of trade? How did it get its name?
4. Name some visible and some invisible items of trade.
5. List any barriers to internal trade in the United States which are at least roughly comparable to the trade barriers of international commerce.
6. Suppose that all impediments to foreign trade were removed at once and never restored, and suppose that political boundaries remain just where they are today. A century from now, after all countries had become adjusted to the new situation, Brazil and Uruguay would each have some amount of domestic and some amount of foreign trade. If we computed for each country the percentage which its foreign trade was of its domestic trade, would we get the larger percentage for Brazil or for Uruguay? On what do you base your conclusion?
7. Suppose you wanted to find whether the foreign trade of England is greater or less than that of the United States. What sorts of problems would you meet in trying to ascertain the facts?
8. List a number of recent trends in the foreign trade of the world as a whole. Of the United States in particular.
9. See if you can discover, from the "World Almanac" or elsewhere, what were the four or five leading commodities among the exports of the United States last year. Among the imports.

10. Imagine that a wheat farmer says to his wife, "If I buy shoes for the youngsters at the store, we'll have the shoes and the store will have the money; but if I make the shoes myself, we'll have both the shoes and the money." What should his wife tell him?

11. If you wanted to raise revenue from a tariff, which of the following commodities would you tax? Why do you pick out these particular ones?

Wheat	Cotton cloth	Raw cotton
Fresh milk	Citrus fruit	Bananas
Coffee	Tin	Farm machinery
Rubber	Tobacco	Ivory

12. On which of the commodities listed in question 11 would a protective tariff be likely to work? Why?

13. Suppose that someone tells you that "the tariffs of the United States must be rather low, after all, since 98 per cent of all the goods imported came in duty free, and the average ad valorem rate on the other 2 per cent was less than 1 per cent." Assuming that his figures are correct, why might the facts lead you to think that the tariffs were high rather than low? Why is it that the percentage of goods brought in free of duty, often used as a test of the effect of protection, is a poor test?

14. Explain how specific and ad valorem duties differ in their effect at times of changes in the general price level.

15. Why do tariff bills always originate in the House of Representatives, rather than sometimes in the Senate?

16. The accompanying table gives hypothetical data, similar to those used in Sec. 23 • 12, showing the demand schedule for a given good, and the domestic and foreign supply schedules in the absence of any tariff.

Hypothetical Demand and Supply Data to Illustrate the Effect of a Tariff

Price per unit	Quantity demanded	Quantity offered for sale	
		Domestic	Foreign
$1	5280	0	100
2	4640	0	500
3	4000	150	1000
4	3500	300	2000
5	3000	600	3000
6	2600	1000	4000
7	2300	1500	5500
8	2050	2100	7000
9	1850	2900	9000
10	1700	3500	12,000
11	1600	4500	15,000
12	1550	7000	20,000

What will be the price of the commodity in the domestic market in the absence of a tax? Under such circumstances, what amount will be produced by domestic and what part imported from foreign producers? If, now, an import tariff of $3 per unit is established, what will be the price in the domestic market, what amount will be produced by domestic producers, and what quantity will be imported from abroad? How much revenue will the government get from the tariff? (In this problem the student should plot the necessary curves and make approximations as close to the answers as he can from his diagram.)

17. Answer all the parts of question 16 except for the assumption that the import tax is $2 instead of $3 per unit.

18. Name any commodities that you think might be protected under the military-necessity argument.

19. Look up the Republican and Democratic platforms of the most recent presidential campaign, and find what they say on the tariff. What are the most noticeable differences, if any?

20. England was noted for many years prior to the first World War as a great commercial nation, which followed in practice the policies of free trade. Yet even at that time it was pointed out that England had some import tariffs. How can a nation have import tariffs and still be a free-trading nation? What additional information would you need to have before you could decide whether England was really following the policy of protection or of free trade?

24. FOREIGN EXCHANGE AND THE BALANCE OF PAYMENTS

24 · 1. Par of Exchange. If an American buys goods in England, he finds that he has no English money with which to pay, and the American money that he does have is not acceptable to the Englishman. Similarly, when an American sells goods in England, he is unwilling to accept the English money that the buyer has, and the English buyer has no American dollars to proffer. But, in the days when the most important commercial nations were on the gold standard, it was possible to find an acceptable way to pay such a debt. The American who bought goods abroad could get his American money changed into gold, since all of our money was convertible into gold on demand. He could ship the gold to the Englishman, who could, in turn, have it coined freely into English money, since when countries are on the gold standard gold enjoys free coinage. Thus, one who bought goods abroad could export gold, which was acceptable in payment, or if he sold goods abroad he could import gold in payment.

We would still have the problem of determining, of course, how much gold to ship. If, for example, an American importer decides to buy 1000 yards of cotton cloth in England and finds that it will cost him £100 in English money, how much American money will it take to settle the bill? The problem is simple if we know the weights of the two basic coins: the dollar and the pound sterling. When these countries were both on the gold standard, the American dollar contained 23.22 grains of pure gold, while the pound sterling contained 113 grains of pure gold.[1] Simple division will demonstrate that the pound contained as much gold as there was in 113/23.22 dollars, or as much as there was in $4.8665+. If, for our purposes, we round this off to $4.87, we can say that the "par of exchange" of the pound was $4.87. When we give the par of exchange of any money, we tell how many units of our own money it takes to give the same gold content that

[1] Actually the British sovereign (their gold coin with a value of £1) weighed 123.274 grains and was 22-carat gold, or 11/12 gold, or 0.91666 fine. It took exactly 1869 sovereigns to weigh 40 Troy pounds. Computation will show that the weight of the gold in the coin was 113.001166 . . . grains, but for purposes of illustration we shall take the weight as 113 grains.

there is in one unit of the foreign money. Thus, in the old days when many countries were on the gold standard, the par of exchange of the French franc was \$0.0392; of the German reichsmark, \$0.2382; of the Russian ruble, \$0.5146. These figures indicate that, at that time, the gold content of the franc was about 4 per cent as great as that of the dollar, the gold content of the reichsmark about a quarter that of the dollar, and the gold content of the ruble about half that of the dollar. While we were quoting the pars of exchange in dollars and cents, an Englishman would have stated the par for the dollar in terms of his own money, in pounds, shillings, and pence. The par for the dollar in those days would have been roughly 4 shillings 1-1/3 pence in England, and approximately 25 francs in France, 4 reichsmarks in Germany, and 2 rubles in Russia. The par of exchange of any kind of foreign money is always the amount of domestic money that it takes to contain the amount of pure gold that there is in one unit of the foreign money. It is found by dividing the weight of gold in one foreign unit by the weight of gold in one domestic unit.

If, now, we return to the case of the American importer who has bought 1000 yards of cloth valued at £100 in English money, we can discover easily how much American money he would have to convert into gold to settle the bill. If each pound is the equivalent of \$4.87, he needs \$487 of American money. This much American money will contain 487 × 23.22 grains, or 11,308 grains of gold. If the American ships this much gold to England, the English exporter will find that he has just enough[1] to be coined into £100, which is just what he wanted. The debt has been settled, not by shipping the usual sorts of currency, but by shipping gold coin or bullion. In gold-standard countries, such coin or bullion can always be obtained and must be coined into money on demand.

24·2. Bill of Exchange. In practice, however, a merchant buying goods abroad would probably not ship gold in payment. He has another method available which is easier, faster, and less expensive. If I owe a sum of money to someone in England and have no English money with which to pay the debt, perhaps I can find someone who does have English money and will be willing to pay it for me. For example, I might find some English tourist, temporarily in this country, who had a balance in his English bank at home. He could use American money while traveling here, and I might be able to arrange with him that I would give him American money if he would draw a check ordering his British bank to pay my debt in English money. Thus, by paying American dollars, I would come into possession

[1] It appears that he has 8 grains more than necessary, but only because we have rounded off the weight of the pound at 113 grains and the par of exchange at \$4.87, instead of using more precise figures.

of a draft (a bank check) ordering someone in England to pay my debt. The check would be easy to mail as compared with sending the gold. All that would be necessary would be to reach some agreement with the English tourist. How many American dollars must I give him to get him to let me use his English balance? Perhaps he will be willing to carry out the transaction at par, charging me $4.87 for every pound that he puts at my disposal. Possibly he will charge me something more for the convenience involved, say $4.90 for every pound. If he does not charge too much premium, I may find that the ease and convenience make the transaction worth while. At least I shall save the cost of shipping the gold, which I would otherwise have to incur.

Although I could hunt around in the highways and hedges for someone who had a balance in a British bank, the natural thing for me to do is to go at once to someone who is in the business and is certain to have English funds available. Some large international bank, with branches in the principal countries of the world, would be such an agency. Thus the National City Bank of New York has a branch in New York and one in London, in addition to many others. If I go to their New York office, I can arrange to have them draw an order on their London branch, ordering that London branch to pay my debt in English funds. When we were discussing domestic commerce we called such a document, in which one party orders a second party to pay money to a third party, a "draft."[1] We found them very useful in financing business transactions. Similarly, most foreign trade is financed by drafts, although those which represent foreign obligations are usually called "bills of exchange." To be sure, usage varies, and to some extent the names are used interchangeably; but, if we follow the most common usage, we shall say that drafts and bills of exchange are the same kinds of documents, the sole difference being that the document is a draft if it arises from a domestic transaction and a bill of exchange if it arises from a foreign transaction. If, then, I want to have the National City Bank pay £100 to the cloth exporter in England, I go to the New York branch of the bank and buy from them a bill of exchange, in which they order their London branch to pay that amount. If they are selling bills of exchange at par, I shall have to pay them $487 for the bill; but they may, if they wish, charge me more or less than par. The actual amount of domestic money charged for one unit of foreign money payable in the foreign country is called the "rate of exchange." Thus, if the National City Bank charges me $4.90 for each dollar payable in London, the rate of exchange is $4.90, even though the par of exchange is still $4.87. Just as a butcher may, under ordinary circumstances, charge more or less for a pound of beefsteak, so a bank may charge more or less for bills of exchange. The rate of exchange is not, like the par

[1] See Secs. 19 • 3 and 19 • 5.

of exchange, fixed by law. It varies from day to day, like the price of beefsteak. And, just as various butchers may be charging different prices for steak at the same time, so different banks may be selling bills of exchange at different prices. If we read in the financial columns of our newspaper that the rate of exchange yesterday was $4.88 to the pound, this is merely a market quotation giving the average price, just as the same newspaper may show the market quotation for wheat or copper.

Men, such as importers, who have obligations to pay abroad or who wish to remit funds to a foreign country for any reason, need to buy bills of exchange from the banks. They make up the demand side of the bill market. If we turn to the exporting business, we find just the opposite sort of thing taking place. If an American exporter sells 10 automobiles in England and the English importer promises to pay £1600 for them, how is the American exporter to change the £1600 into dollars? How does he finance the transaction? He does it exactly as the seller of goods in domestic commerce does. He puts the automobiles on board the freight car or the boat and gets a bill of lading. He then makes out a bill of exchange, ordering the English importer to pay £1600 to the order of the American's bank. He takes the bill of exchange and the bill of lading to his bank and has them discounted just as in domestic commerce. Of course, if the bill of exchange is payable in 60 or 90 days, the bank will discount it for that period. But, we have the added fact this time, which did not complicate our situation with the domestic transaction, that the bill of exchange is drawn in pounds sterling, requiring the foreign importer to pay £1600. The American exporter sells the bill of exchange to his bank at the rate of exchange. If the rate of exchange, for example, is $4.50, the bank credits the American exporter with $4.50 × 1600, or with $7200 in American money,[1] which is the only kind of money that the exporter can use. The bank sends the bill of exchange and the bill of lading to its foreign branch.[2] When the papers arrive at the foreign branch, the officials there notify the English importer that they hold a bill of exchange and a bill of lading for a shipment of automobiles and, if he comes to the bank and accepts the bill of exchange, he may have the bill of lading which is necessary if he is to get delivery of the automobiles. The importer accepts the bill and, at the end of the credit period, if any, he pays the English branch in pounds sterling, the only kind of money that he has.

[1] To be sure, if the bill of exchange has 60 days to run, the bank will discount the $7200 at its regular discount rate; but, since we have explained discount before, it does not seem wise to complicate the present transaction by covering both foreign exchange and discount at the same time.

[2] There are, to be sure, relatively few banks that have foreign branches. It is likely that the exporter will take the bill to his own local bank, and this local bank will, in turn, pass it on to one of the great international banking companies. This would merely add a step to our program, without altering any of its essentials.

The American exporter received dollars, and the English importer pays in pounds.

The student may be puzzled to find, in our two illustrations, that once the importer bought a bill of exchange and the other time the exporter sold a bill of exchange. On every foreign transaction there are both an importer and an exporter, and of course only one of them actually deals in foreign exchange. What determines which one has to do it? This depends in part on the way in which the debt is incurred, and in part on custom. If the importer of cotton cloth finds that it is quoted in pounds sterling and if he promises to pay £100, he is responsible for having the dollars converted to pounds. He has agreed to pay in pounds, and he must get them to pay with. If the agreement had called for a payment in dollars, the Englishman, who wants pounds, would have had to take care of the conversion by buying a bill of exchange. In our other imaginary transaction, an American sold automobiles abroad. The price might have been agreed on in either dollars or pounds. If the Englishman had promised to pay $7200, he would have been responsible for getting the dollars with which to pay. He would have bought a bill of exchange. If, on the other hand, the contract had called for a payment of £1600, the American who wants dollars would have been responsible for converting the pounds to dollars. He would have sold a bill of exchange. Whether the exporter of the goods sells a bill of exchange or the importer of the goods buys a bill of exchange depends on the way in which the contract reads—whether it calls for payment in the kind of money used by the importer or by the exporter. The contract will have to call for payment in some form, either in dollars or in pounds. If it calls for payment in pounds, the American will be responsible for the conversion and will buy a bill if he is an importer or sell a bill if he is an exporter. If the contract calls for payment in dollars, the Englishman will be responsible for the conversion and will buy a bill if he is an importer or sell a bill if he is an exporter. According to actual custom and trade practice, it has been true for many years that a considerable proportion of international trade is carried on in terms of pounds sterling. If the contract does call for payment in pounds, as will usually be the case, then the American will be responsible for the conversion.

24 · 3. Effect of Foreign Exchange on Gold Balances. The buying and selling of bills of exchange is called "foreign exchange." If we note the effects of the two foreign-exchange transactions used as illustrations in the preceding section, we note that they have opposite effects as far as the banks are concerned. In the case of a typical import of goods, the importer goes to his domestic bank and buys a bill of exchange and pays for it. This increases the amount of money in the domestic bank. But the importer sends the bill abroad to the foreign exporter from whom he bought the

goods, and this foreign exporter collects the bill at his bank, thus reducing the money in the foreign branch. Every importation of goods

1. Increases the demand for bills of exchange in the importing country.
2. Increases the stock of money in the domestic banks.
3. Reduces the stock of money in the foreign branches of the banks.

On the other hand, in the case of a typical export of goods, the exporter goes to his domestic bank and sells a bill of exchange and collects the money for it. This reduces the amount of money in the domestic bank. The bank then sends the bills of exchange abroad, where its foreign branch collects the money from the foreign importer, thus increasing the money in the foreign branch. Every exportation of goods

1. Increases the supply of bills of exchange in the exporting country.
2. Reduces the stock of money in the domestic banks.
3. Increases the stock of money in the foreign branches of the banks.

If imports and exports balance each other, it is evident that the banks will be gaining cash on some transactions and losing on others, but keeping the totals roughly constant. If, on the other hand, we have a continued "favorable" balance of trade, with exports of goods exceeding imports, the stock of money in the domestic banks will slowly be drawn down lower and lower, while balances in the foreign branches will be built up larger and larger. Ultimately, the domestic banks would run out of funds and have to go out of business if they could not bring some of their foreign excess cash into the country. This they will tend to do. The exportation of goods, if continued without balancing imports of goods, will give rise to the importation of cash. If, however, we have a continued "unfavorable" balance of trade, with imports of goods exceeding exports, the stock of money in the domestic banks will grow slowly larger and larger, while the balances in their foreign branches will be reduced more and more. Ultimately, the foreign branches would run out of cash and have to go out of business if the domestic banks did not ship some of their surplus cash abroad. This they will tend to do. The importation of goods, if continued without balancing exports of goods, will give rise to the exportation of cash. Thus, we see that a continued balance of commodity trade in one direction tends to give rise to a movement of money in the other direction: importation of goods brings exportation of cash; exportation of goods brings importation of cash. This is what one would expect. If we import goods, we have to send cash out to pay for them. If we export goods, we have to bring cash in to pay for them. If we both import and export, and in equal quantities, then the exports and the imports cancel each other. One of them builds up domestic balances and reduces those abroad; but the other reduces domestic balances and builds up those abroad. No monetary movement is necessary, and the debts tend to cancel each other without the necessity of using cash. Here we find a

system of "clearing" foreign transactions similar to the systems of clearing that we discovered in domestic trade.[1]

24 · 4. Forces Affecting the Exchange Rate. Let us suppose, for the moment, that the United States is having a continued export balance of goods. Under such circumstances, as we have seen, stocks of cash in domestic banks will be reduced and stocks abroad will be increased. Ultimately, the bank will have to ship gold from its foreign branch to its home office.[2] But such a shipment will cost money, and the bank will prefer not to do it. It will try first to discourage the type of transaction that tends to make the gold shipment necessary. But these transactions are sales of bills of exchange to the banks by exporters. How can we discourage exporters from selling us bills of exchange? Obviously, by reducing the price that we will pay for them. We will reduce the price below par, perhaps; and, instead of paying $4.87, we will offer but $4.86 or $4.855. Continued exportation of goods without balancing imports tends to force the rate of exchange down. This we might well have guessed anyway, since we have seen that exports of goods constitute the supply of bills of exchange, and increased exports will increase the supply of bills and force their price down. The price of bills of exchange is the rate of exchange, so we see that, just as we noted before, exports of goods tend to lower the exchange rate.

On the contrary, suppose that the United States is having a continued import balance of goods. Under such circumstances, as we have seen, stocks of cash in domestic banks will be increased and stocks abroad will be depleted. Ultimately, the bank will have to ship gold from its home office to its foreign branch. But such a shipment will cost money, and the bank will prefer not to do it. It will try first to discourage the type of transaction that tends to make the gold shipment necessary. But these transactions are purchases of bills of exchange from the banks by importers. How can we discourage importers from buying our bills of exchange? Obviously, by raising the price that we charge for them. We will raise the price above par, perhaps; and, instead of offering bills at $4.87, we will charge $4.88 or $4.89. Continued importation of goods without balancing exports tends to force the rate of exchange up. This we might well have guessed anyway, since we have seen that imports of goods constitute the demand for bills of exchange, and increased imports will increase the demand for bills and force their price up. The price of bills, however, is the rate of exchange, so that we see again that we reach the same result when we consider cash balances that we would have reached had we confined ourselves to more common supply-and-demand analysis.

[1] See Sec. 19 · 9.

[2] We are still assuming that both countries are on the gold standard. If this is so, gold will be available abroad to ship and can be coined freely for use in this country after it is imported.

24 · 5. Gold Points. We have just seen that importation of goods tends to force the rate of exchange up, and exportation of goods tends to force it down. A continued "unfavorable" balance of trade might be expected to force the exchange rate above par, and a continued "favorable" balance to force it below par.[1] If the imports and exports balance, we might expect the exchange rate to remain in the neighborhood of par. Let us suppose that a country has a long-continued excess of imports and that the exchange rate as a result rises above par. When importers go to the bank to buy bills of exchange, they find that they must pay more than $4.87 for each pound sterling. How much of an excess will they pay over and above $4.87? The answer is that, if the premium gets large enough, it will pay them to ship gold rather than buy high-priced bills of exchange. How far above par the rate of exchange can go before it pays to ship gold depends on the cost of shipping the gold, and this cost varies according to the quantity of gold to be shipped and to various other circumstances. For example, ocean freight rates and insurance rates are certain to be high in times of war, while in times of peace they are much lower. In ordinary times and on reasonably large shipments, it costs roughly 2 cents per pound sterling to ship gold from New York to London or from London to New York. Let us adopt this figure of 2 cents a pound as the shipping cost for purposes of our illustration.

Suppose that an American importer buys goods in England and promises to pay one pound sterling for them. If he ships the gold, he will have to send $4.87 worth of gold to pay the bill; in addition, it will cost him 2 cents to ship the gold. His total outlay will be $4.89. If he can purchase a bill of exchange for less than $4.89, it is better to settle his debt by means of the bill of exchange than it is to ship the gold. Therefore, if he finds a rate of exchange of $4.88, he buys the bill of exchange and spends $4.88 rather than the $4.89 involved in shipping gold. Similarly, if the rate of exchange is $4.885, it still pays to buy the bill of exchange. But, as soon as the rate of exchange exceeds $4.89, it no longer pays to buy the bill of exchange. If the rate of exchange, for example, is $4.90, it will cost $4.90 to settle the debt by buying a bill of exchange, but he can settle the debt by shipping gold at an expense of $4.89. If the rate of exchange passes $4.89, importers stop buying bills of exchange and start shipping gold out to pay their debts. We say that the rate of exchange has reached the "gold export point"—the level above which it no longer pays to buy bills of exchange, but pays to export gold instead.

On the other hand, suppose that an American exporter ships goods to England and agrees to accept one pound sterling in return. To be sure, he

[1] This need not be true if we consider only the trade of one pair of countries, since a "favorable" balance with one country may be offset by an "unfavorable" balance with another. This is illustrated in Sec. 23 · 6.

has no use for English money, but he can always bring the actual gold into the United States and get it exchanged for $4.87 in American money. The difficulty is that it will cost him 2 cents to bring it in—2 cents of shipping charges, insurance, etc. Although he will get $4.87 worth of gold, he will have to spend 2 cents to get it, and he will have left a net return of only $4.85. If he can sell his bill of exchange for anything more than $4.85, it will be better to sell it than to import the gold. If, for example, he can sell his bill of exchange for $4.86, he has 1 cent more than he would have if he imported the gold. But, if the exporter takes his bill of exchange to the bank and finds that the rate of exchange is below $4.85—say, at $4.83—he will not sell them the bill. He will prefer to import the gold. When the rate of exchange falls to the point where it no longer pays exporters to sell their bills of exchange but where they prefer to import the gold, we say that the rate of exchange has reached the "gold import point." Imports of goods tend to force us toward the gold export point. Exports of goods tend to force us toward the gold import point. If we export goods, we import gold to pay for them; if we import goods, we export gold to pay for them. If we both import and export goods, the two tend to cancel each other, and we do not have to transport gold at all. We run into the process of cancellation of debts which we called "clearings" when we were dealing with domestic commerce.[1]

These two points, the gold export point and the gold import point, are together called the "gold points." They are two rates of exchange, one above and the other below the par of exchange, and distant from the par by an amount equal to the cost of shipping gold. The gold export point is a rate of exchange exceeding par by the amount that it costs to export one gold unit of the foreign currency. The gold import point is a rate of exchange lying below par by the amount that it costs to import one gold unit of the foreign currency. If the costs of transporting gold are the same in both directions, importing and exporting (this is by no means always true in practice), the gold points are equidistant from the par of exchange, each being distant from it by the cost of transporting one unit of the foreign currency.

When the countries concerned are both on the gold standard, the rate of exchange fluctuates between the gold points, moving up or down in any country depending on whether that country has an import or an export balance of trade in commodities. As long as the countries remain on the gold standard, the rate of exchange can not move outside the limits set by the gold points because, if banks set the price of bills of exchange outside

[1] In practice, gold is imported or exported, not by individuals, but by banks that hope to profit by the transaction.

these points, it pays to import or export the actual gold rather than to deal in bills of exchange.

24·6. Triangular Trade. So far we have considered the trade between only two countries, on the assumption that the trade between them could be isolated from the rest of their trade. This assumption is contrary to fact. The exchange rate between England and the United States depends not only on the trade between them but also on the trade of each with the rest of the world. This fact can be best illustrated, not by considering all the countries of the world at once with the necessary complexities incident to such a procedure, but by considering the trade of three countries at once as an intermediate example between the two-country situation and the situation of the real world. Let us consider, for example, the trade between the United States, England, and Brazil. Usually, the United States imports much more from Brazil (mainly coffee) than she exports to Brazil. Usually

Table 24·1. Hypothetical Example of Triangular Trade, All Exports and Imports Being in Millions of Dollars

(Exports from country at head of column to country at left side of the table)

	United States	United Kingdom	Brazil
United States.....	...	$200	$100
United Kingdom....	$350	...	10
Brazil.....	50	75	

the United States exports much more (cotton, etc.) to England than she imports from England. Usually, England exports much more (textiles, etc.) to Brazil than she imports from Brazil. Let us assume that in a given year the amounts of the trade are those given in Table 24·1, with all the trade measured in dollars to avoid confusion.

Thus, for example, the table tells us that the United States exported 50 million dollars' worth of goods to Brazil and imported 100 million dollars' worth from Brazil. If we read across a row in the table, we find the imports of the country at the left. If we read down a column of the table, we find the exports of the country at the top. Let us assume that there is no international trade except that between these three countries, and determine the effect of the trade shown in the table.

The data of Table 24·1 are shown graphically in Fig. 24·1, where the arrows show the direction and the figures beside the arrows show the amounts of the trade. We recall, from Sec. 24·3, that every export of goods tends to reduce cash balances in the banks of the exporting country, and every import of goods tends to increase cash balances in the banks of the importing

country. But inspection of Table 24 • 1 or of Fig. 24 • 1 shows that the United States is exporting 400 million dollars and importing 300 million dollars' worth of goods. She has an export balance of 100 million, which will tend to reduce cash balances in American banks by that amount. This is indictated by the figure—$100 at the top of the chart, showing a net cash deficit of 100 million dollars in American banks. Similarly, we see that the United Kingdom has an import balance of 85 million dollars and Brazil an import balance of 15 million, which tend to build up cash surpluses of these amounts in these two countries. The net result is that the banks of the United States need to import 100 million dollars of cash to restore their balances, and the 85 million surplus in the United Kingdom and the 15

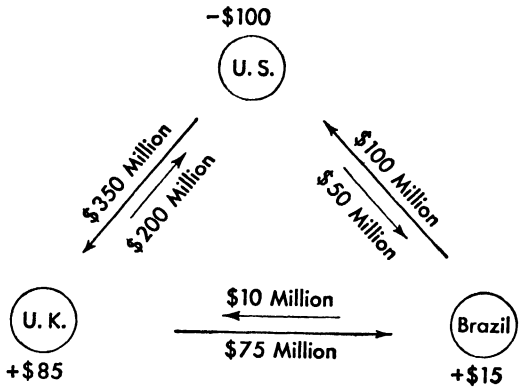


FIG. 24 • 1.—Graphic representation of hypothetical trade data of Table 24 • 1 showing triangular trade.

million surplus in Brazil are exactly enough to make up the deficit in the United States. The tendency will be to ship the excess funds from Brazil and the United Kingdom to the United States. This means that in the United States the rate of exchange on the pound sterling and on the Brazilian cruzeiro will tend to drop toward the gold import point; while in Brazil and in the United Kingdom the exchange rates on the dollar will tend to rise toward the gold export point.[1] This will bring about the transfer of gold that is necessary in our example if we are to restore the cash balances of the banks in the various countries. It is not the cash balances in the banks

[1] It should be obvious that there is a tendency for the English rate on the dollar and the American rate on the pound to move in opposite directions, since they are reciprocals of each other. If the rate in the United States shifts from $4 to $5 per pound sterling, the rate in England shifts from 1/4 to 1/5 pound sterling to the dollar. Thus, when the rate in the United States is dropping toward the import point, the rate in England is rising toward the export point. Ultimately, the exchange rates get to the point where it does pay to *export* the gold from England and to *import* it into the United States.

of any one country alone that are important. It is quite possible that our trade with Brazil is such as to increase domestic cash balances, while our trade with England is such as to deplete them. If the excess from Brazilian trade exactly offsets the deficit from English trade, no gold shipment is necessary, and the rates of exchange can remain at or near par. Figure 24·2 shows the actual balances of various countries and groups of countries in three different years between the two World Wars.

24·7. Exchange Rates in the Absence of a Gold Standard. So far we have assumed that there was a par of exchange, dependent on the weights of the basic gold coins of the countries involved and that businessmen could, as a last resort, obtain gold and ship it to settle their debts. If, as at the present moment, many of the leading commercial nations of the world are not on the gold standard, what becomes of our pars of exchange and our gold points? The answer is that they disappear; they are nonexistent. The rate of exchange is no longer confined between the gold points but many fluctuate without any such arbitrary restriction. Yet it would be incorrect to say that it can fluctuate without any restriction at all. Let us ask, for example, how high the rate of exchange can go in our dealings with England when neither country is on the gold standard. It can go to $4.87, to $4.89, to $5, or even higher. But, if it gets too high, it no longer pays American businessmen to buy goods in England. If 100 yards of English cloth are quoted at £10, this may make an attractive bargain to an American importer when the exchange rate stands at $4 to the pound sterling. He can buy the cloth at $40. If the exchange rate rises to $5 per pound, he finds that he has to pay $50 in American money for the same cloth, even though it is still quoted at £10 in English money, just as before. And, if the exchange rate were to rise to $7 to the pound, the American cannot use his previous method of exporting gold to pay, since we are not on a gold standard; but he can refuse to buy the cloth. The cloth will now cost him $70 in American money, and it is quite likely that he will prefer to buy American-made cloth rather than pay such a high price for foreign cloth. As the exchange rate gets higher and higher, instead of reaching a gold export point, we finally come to the place where it no longer becomes profitable to buy foreign goods and import them. The importation of goods forces the rate up and, when we are not on a gold standard, it keeps right on going up until it no longer pays to import.

On the other hand, suppose that we have an export balance, and the exchange rate starts down. If we were on the gold standard, we would soon reach the gold import point and start to bring gold in. But since we cannot do this, the exchange rate keeps right on going down. It passes $4.85 and goes down, let us say, to $4.50, to $4, or even lower. What stops it? The answer is that American exporters, who are selling their goods in England for, let us say, £100, find that they convert this money into but $400 in

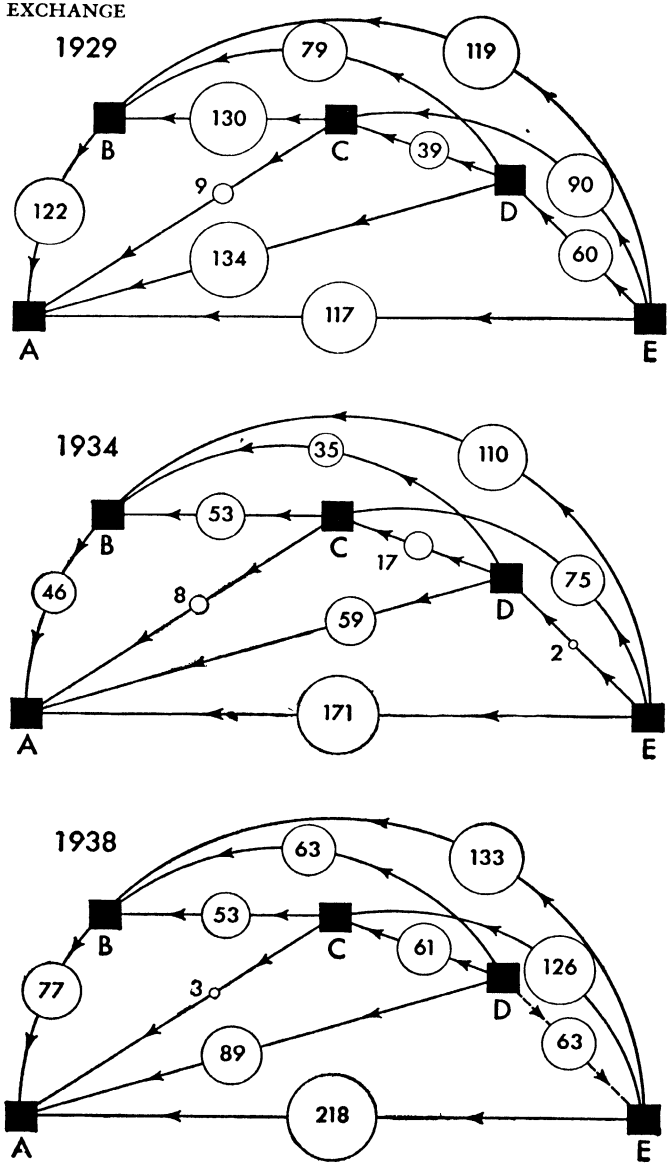


FIG. 24 • 2.—Import and export balances of various countries and groups of countries in 1929, 1934, and 1938. Each black square represents a country or group of countries as follows: *A* is the United Kingdom; *B* is 10 European countries (Austria, Belgium, Czechoslovakia, Denmark, Finland, France, Netherlands, Norway, Sweden, and Switzerland); *C* is Germany; *D* is the United States; and *E* is all other countries. Amounts of trade balances are shown (in millions of pounds sterling) by figures within the circles. Arrows show the directions of the balances, which, in all save one case, are from the right toward the left in the diagrams. The single exception is shown by the broken line for 1938. (*From Review of World Trade,* 1938, *League of Nations, Geneva,* 1939, *p.* 38.

place of the $485 that they got when the currencies were at par. And, if the exchange rate were to go down to $3 or to $2 per pound sterling, the American would get but $300 or $200 in exchange for goods that formerly brought him $485. The goods still sell for the same quotation in English money—£100—but, as the exchange rate drops lower and lower, the American exporter gets less and less American money from the sale. Finally, the return in American money becomes so small that it no longer pays to sell goods abroad. The exportation of goods has forced the rate of exchange down to the point where exportation is no longer profitable.

Thus we see that, in the absence of the artificial limitations of the gold points that exist when countries are on the gold standard, there are still limits to the fluctuation of the rate of exchange. Exports of goods tend to force the rate down, if they are continued, until exports are no longer profitable; continued imports of goods tend to force the rate up until imports are no longer profitable. Either type of transaction, if not offset by the other, tends to be self-limiting.

24 · 8. Exchange Regulation and Trade Restrictions. In Chap. 23, we turned our attention to the use of protective tariffs as means of regulating and influencing international trade. The tariff is the old and obvious well-tried mechanism for effecting such regulation. In more recent years, many countries have adopted the plan of establishing direct quotas on the importation of goods. Such quotas may merely limit the total amount of a commodity to be imported, or they may go further and allot the quota among various exporting countries. Import quotas have sometimes been linked with agreements under the provisions of which the exporting country agrees, in return for its quota, to import certain goods from the country establishing the quota.

Although quotas may be worked out directly in terms of the goods to be imported, a somewhat more common procedure has been to establish control over foreign exchange rather than directly over the goods themselves. A country following such practices ordinarily provides by law or by executive fiat that exporters must sell their bills of exchange to a state control bureau, which thus comes into possession of rights to demand foreign currencies. The importation of goods can then be controlled by restricting the sale of foreign exchange to those importers who buy the goods desired and who buy them from the countries desired. Similar methods have often been used by countries, not primarily to direct foreign trade into desired channels, but to regulate the foreign value of their money when they have left the gold standard. Under such circumstances, as we have seen, the international value of money is no longer confined to the narrow limits set by the par of exchange and the gold points, but may fluctuate very widely. The direction and extent of the fluctuation will depend on the char-

acter and direction of foreign trade, and exchange control may thus be used to keep the foreign exchange rate within bounds—to keep the domestic money from depreciating or appreciating too far in terms of the currencies of other countries.

24·9. Balance of International Payments. As was noted in Secs. 23·2 and 23·3, international trade in actual physical commodities is often given attention to the exclusion of trade in services. Actually, however, any business transaction that gives Americans a right to receive payments of money from abroad has the same effect on exchange rates as do exports of goods. When we export goods or perform services for foreigners or borrow funds from foreigners, we obtain rights to receive money abroad. When we import goods or buy services from foreigners or lend funds to foreigners, we obligate ourselves to make payments of money abroad. The balance of international payments of the United States is merely a list of all the payments to and receipts from foreign countries during a given period, usually a year, classified according to the type of transaction that gave rise to the transfer of funds. Although the war has recently disrupted foreign trade, we can get some understanding of the problem from Table 24·2 which shows the international balance of payments of the United States in 1940.

In Table 24·2, every export of goods, or every other transaction that gives Americans a right to demand money from abroad, is listed as a credit with a plus sign; and every import of goods, or every other transaction that obligates Americans to pay money abroad, is listed as a debit with a minus sign. The first items listed are the actual merchandise exports and imports of the United States in 1940. In that year, we exported just over 4 billion dollars' worth of physical commodities and imported about 2.6 billion. These are the so-called "visible" imports and exports. There was an export balance of 1,396 million dollars, which is listed in the last column of the table. This commodity trade gave us the right to demand a balance of 1,396 million dollars abroad.

In 1940, foreigners used American ships for some of their trade, and spent 336 million dollars for such services. Americans, on the other hand, spent 346 million dollars for freight in foreign bottoms, so that on balance Americans spent abroad 10 million more than they received for this kind of service. Similarly, American tourists spent abroad 86 million dollars more than foreign tourists spent in the United States. We can think, if we like, that Americans "imported" foreign scenery, hotels, etc., to this amount. People living in the United States sent money to their relatives or to others in foreign countries to the extent of 119 million dollars, which exceeded by 91 million the personal remittances sent in the other direction. Institutions in the United States (mainly, the Red Cross) sent 54 million dollars abroad. Americans received more interest on foreign investments than was sent

back by us to foreign investors in this country. The United States spent more abroad for government services than foreign governments spent here, and more silver was purchased abroad than was sold abroad. Thus, we find from the table that, when we consider the total merchandise and service items, the United States had "exported" roughly 1,433 million dollars more than she had "imported." The United States had had over all these items

Table 24·2. Balance of International Payments of the United States, 1940, in Millions of Dollars*

Item	Receipts from foreigners for "exports" (credits)	Payments to foreigners for "imports" (debits)	Net credits (+) or debits (−)
Trade and service:			
Merchandise	$4021	$2625	+$1396
Shipping and freight	336	346	−10
Travel expenditures	81	167	−86
Personal remittances	28	119	−91
Institutional contributions	.	54	−54
Interest and dividends	580	195	+385
Government transactions	36	123	−87
Silver	3	58	−55
Miscellaneous (net)	35	..	+35
Total	5120	3687	+1433
Gold movements (net):			
Imports and exports	....	....	−4744
Earmarking operations	....	....	+ 645
Total	....	....	−4099
Capital (net):			
Long-term capital	....	....	− 138
Short-term capital	....	....	+1657
Balance on capital transactions	....	....	+1519
Unexplained	....	....	+1147

* Data from "Statistical Abstract of the United States, 1942," Table 589, p. 538. The capital items are viewed as "exports" and imports" of evidences of indebtedness.

together a "favorable" balance of trade.[1] Foreigners owed Americans, roughly, 1,433 million dollars on balance for all these items of trade combined.

The figures throughout the remainder of the table do not show both the "export" and the "import" movements, but merely the net movements. They

[1] Perhaps it should be emphasized again that such a balance is no more desirable than an import balance. It might be better to speak of export and import balances instead of favorable and unfavorable balances. Some authorities use the terms "active" and "passive" balances for "export" and "import." Thus, the terms that are synonymous are, on the one hand, favorable, export, and active balance; and on the other hand, unfavorable, import, and passive balance.

show that the United States, on balance, received 4,744 million dollars of gold from abroad as foreigners and their governments shipped gold either in payment of bills or to get it out of the theater of war. Offsetting this in part was the fact that 645 million of gold already here was earmarked (set aside) for foreign governments, having the same effect as though this much gold had been sent to them. Therefore, the total effect of gold movements and earmarking was to increase the stock of gold in the United States by just over 4 billion dollars. This was far more than enough to offset the export balance on our merchandise trade and our services. If the transactions had stopped there, we would have received far more than enough to pay for our exports. But there are still the capital items to consider: foreign investment in this country and American investment abroad. The 1940 situation was not typical in this regard but was badly distorted by the war. Foreigners were buying war supplies in the United States. Prior to the passage of the Lend-Lease Act in 1940 they had been required, under the terms of our Neutrality Act, to pay for goods in dollars before the goods could be exported. Many of these goods were in fact paid for by shipments of gold to this country. But foreign shipments of gold were more than sufficient to pay for the exports, as we have seen, and the balance was largely held on deposit in American banks payable to the foreign country or firm. We had exported our promises to pay (bank deposits) and had received gold in return for them. When an American invests money abroad, we can think of it as an "import" of foreign stocks, bonds, or other securities. When foreigners invest money in the United States, we may consider it an "export" of our securities. The items are so listed in the balance of international payments, as noted in the footnote to Table 24·2.

The government has fairly accurate data on imports and exports of merchandise, since, barring smuggling, they pass through customhouses where records are kept of all such shipments. The other items, however, are not so easily nor so completely recorded. Tourists traveling abroad submit no records of expenditures. Capital transactions often do not get into the public records. We have emphasized before that, for the most part, international trade is between individuals rather than between nations, and there is no reason why records of private transactions should fall into the hands of government statisticians unless one thinks of statistics as an end in themselves. It is not surprising, therefore, to find that the data of Table 24·2 are incomplete and that some 1,147 million dollars' worth of transactions are unexplained. Naturally, if we had complete records on every transaction, the sum of the "plus" and the "minus" items would have to balance. As in any other account, the items are so recorded that they must balance if the accounts are correctly, completely, and accurately kept. Thus, we see that on the whole a country cannot have either a favorable or an unfavorable balance

of payments. The merchandise trade, to be sure, may show either an import or an export balance, as may any other single part of the account; but the account taken as a whole must balance. The receipts from other countries and the payments or debts to other countries must be equal. Every time the housewife buys a Thanksgiving turkey from her neighborhood market she "imports" a turkey; but she must either "export" the money to pay for it, or else she must "export" a "short-term capital item"; that is, she must go into debt for the turkey. The "import" of the turkey must be offset either by a payment of cash or by a debt. Similarly in international trade, every "import" item is offset by some sort of "export" item, and vice versa. If we sell goods abroad, the foreigner may repay his debt in any one of three ways: by sending us goods in return, by borrowing from us or going into debt to us, or by sending gold in payment. The idea of a unidirectional "favorable" or "unfavorable" balance is an illusion which misleads only those who have limited their thought to superficialities.

24 · 10. Price Adjustments to Foreign Trade. The theory underlying international trade is too complex to be covered in detail in an elementary book such as this one. We have already pointed out that international trade, like domestic trade, is based on the concept of comparative advantage and that, as long as there are differences between nations in soil, climate, natural resources, training and interests of the people, etc., just so long will it be profitable for nations to carry on international trade. We have also pointed out that, if one takes into consideration all the items of trade, including the invisible items, the debts, and the flow of money, the "imports" and the "exports" must balance.

Before we leave our elementary consideration of the subject, however, we should point out one tendency in trade which is simple and easy to understand on the basis of principles already discussed. Suppose that a country exports year after year larger amounts of merchandise and services than it imports. This must be offset either by foreign debts to the exporting country or by shipment of gold from the importing to the exporting country. If the countries are on the gold standard, the shipment of gold will be simple to arrange and will take place as soon as the excess of exports has forced the rate of exchange down to the gold import point. Gold will start to leave the importing country and enter the exporting country. As we noted in Secs. 21 · 7 to 21 · 9, the increased quantity of gold in the exporting country will tend to raise the general price level there, and the reduction in quantity of gold will tend to lower the general price level in the importing country. If this shift in prices continues, it will soon put an automatic stop to the shipment of goods; for it will no longer be possible to export goods after their prices have risen, especially to a country where prices have fallen. To be sure, this does not mean that all export will stop in the first country

and all importation stop in the second country. It does mean that, if we have for some time an export balance of commodity trade in country A and an import balance in country B, the price level will rise in A and fall in B until the exports of A are reduced and the exports of B are increased. Ultimately, the two will tend to balance. If we include all the visible and the invisible items of trade of a country but omit the international indebtedness and the movements of money, the country may have for some years a continued favorable or unfavorable balance of trade.[1] Even in this case, the balance is but a temporary one and will ultimately be stopped by movements of the general price level in the countries involved. Prices will tend to rise in countries with an export commodity balance and to fall in countries with an import commodity balance. This change in the price level will come about as a result of shipments of gold if the countries involved are on the gold standard; if the countries are not on the gold standard, the same sort of change will take place as a result of fluctuations in the price of foreign exchange, which is no longer confined between the gold points, but is free to fluctuate until continued import or continued export becomes unprofitable as explained in Sec. 24·6. Thus, we see not only that favorable and unfavorable balances of trade are not really desirable and undesirable, respectively, but also that they are temporary at best and mere figments of the imagination anyway, if we consider all items in the international balance of payments.

Questions

1. Before the United States left the gold standard in 1933, the dollar contained 23.22 grains of gold. The monetary units of certain other countries that were on the gold standard and their pars of exchange were as follows:

Austria	Schilling	$0 1407
Belgium	Belga	0 1390
Bolivia	Boliviano	0.3650
Brazil	Milreis	0.1196
Bulgaria	Lev	0.0072
Canada	Dollar	1.0000
Chile	Peso	0.1217
France	Franc	0.0392
Germany	Reichsmark	0.2382
Great Britain	Pound sterling	4.8665
Japan	Yen	0.4985
Russia	Ruble	0.5146
Spain	Peseta	0.1930

What was the weight of gold in each of the foreign units?

[1] Of course, even in any given year or week or hour, the trade must always balance if we include debt items and money payments.

2. How would a German have quoted the par of exchange of the French franc? How would a Frenchman have quoted the par of exchange of the German reichsmark?

3. Were the weights of the monetary units of most countries heavier or lighter than the gold dollar (assuming that they had a fineness equal to that of the dollar)?

4. On July 1, 1945, the U. S. Treasury Department quoted the value of the pound sterling at $8.2397 in American money. Under the laws of that date, a gold dollar, although not actually coined, was defined as containing 13.71 grains of pure gold. When, prior to devaluation, it had contained 23.22 grains of pure gold, the par of exchange for the pound sterling had been $4.8665. Was the devaluation sufficient to account for the shift to the new par, or were other factors involved?

5. On July 1, 1945, the U. S. Treasury Department quoted the value of the Liberian dollar at $1.6931. Recalling that the United States dollar was then defined as 13.71 grains of pure gold, compute the gold content of the Liberian dollar. What would the par of the Liberian dollar have been if the American dollar had not been devalued?

6. What is a bill of exchange? For what is it used? Where does it come from? That is, who draws it up?

7. "Exports of goods give rise to the supply of bills of exchange; imports of goods constitute the demand for bills of exchange." Explain.

8. What determines, on a given international transaction, whether the man who exports the goods sells a bill of exchange or the man who imports the goods buys a bill of exchange?

9. How does a bank that buys and sells bills of exchange make any profit on the transactions?

10. Suppose that an American exporting country ships a tanker full of fuel oil to England. Explain in detail what documents will be drawn up, how they will be handled, and especially what the effect will be on the cash balances in American and in English banks.

11. Explain briefly how and why exports and imports affect the exchange rate.

12. What is the gold export point? How is it determined?

13. Suppose that the American dollar contains 50 grains of gold and the French franc contains 10 grains of gold. Suppose that it costs 5 cents to ship a franc from the United States to France, or 8 cents to ship a franc from France to the United States. Considering the situation from the standpoint of an American, what is the par of exchange for the franc? What is the gold export point? What is the gold import point?

14. Suppose, under the conditions of question 13, that the actual rate of exchange is $0.13 to the franc. Will exporters of goods sell their bills of exchange at this low rate, or will they prefer to import gold? Explain.

15. Explain how it is possible to carry on great quantities of international trade without any shipment of gold if the imports and the exports balance. Describe how the clearings process takes place.

16. Suppose that all the figures of Table 24 • 1 are left exactly as they are with the exception of one, but the exports from the United Kingdom to Brazil are changed from 75 million to 50 million dollars. Work out a new diagram similar to Fig. 24 • 1, and compute the gains and losses to the cash balances in the banks of the various countries. What, in your opinion, will be true of the rate of exchange for the English pound in the United States? For the Brazilian milreis?

17. Describe what happens if, when neither of two countries is on the gold standard, one continues to export more to the other than it imports in return.

PART FIVE Special Problems

25. PROBLEMS OF MONOPOLY

25 · 1. Nature of Monopoly. We noted in Chap. 16 that monopoly differs from competition in that the monopolist is able, through complete or partial control of the supply of the product, to influence the price; while the producer who operates under competitive conditions is too small a part of the market to have any measurable effect on prices. Every businessman likes to establish a monopolistic position for himself if he can and, if he is unable to establish complete monopoly,[1] he prefers partial monopoly to complete competition. He tries to distinguish his product from those of his competitors, even if only in some inconsequential way, so that consumers will develop "good will" toward his product and insist on getting his product in preference to those of his competitors. This effort to make his product unlike others of the same general kind—to set it off and make it possible to identify it—is called "product differentiation." The automobile manufacturer tries, by various changes in the shape of headlights or by some such catchy slogan as "dynamic drive" or "the only car with a built-in-bar," to establish in the customer's mind the idea that his car is different and presumably better.

Monopoly is not merely a matter of size. The size of a business might be measured by the number of men employed, the amount of capital invested, or the value of its product. By any or by all of these criteria, a business firm might be very large and yet competitive. Thus, Sears, Roebuck and Company is a vast business enterprise, yet most of the things that it sells are available from many other sources, and Sears, Roebuck has to compete with the other vendors. On the other hand, a business firm might be rather small on the basis of the criteria listed above, yet it might at the same time have a very considerable degree of monopoly power. Let us imagine the situation, for example, of a small-town garage with an exclusive franchise empowering

[1] Complete monopoly never exists if we define it as ability to alter prices without affecting the quantity sold. The student will do well at this point to reread Secs. 16 • 1 to 16 • 5.

it to operate busses between its home town and a near-by city. No one else may operate in competition with them, yet perhaps they own only two or three cars or busses and hire only three or four men.

Monopolies are distinguished from other businesses, then, not by their size, but by their price-fixing power. In more advanced work in economics it is common to recognize one or more intermediate classes lying between complete competition and complete monopoly; but for our purposes we may well study together all those cases in which the seller has sufficient control over supply to enable him to influence the price.

25 • 2. Undesirable Features of Monopoly. In the United States the general tendency has been to frown upon monopolies—either to prohibit them entirely or to subject them to regulation beyond that usually given to competitive firms. Although it is evident that monopoly is necessary in some cases,[1] the tendency has been to regard it as a necessary evil—as something

Number of bananas	Total utility, utils	Marginal utility, utils
1	50	50
2	90	40
3	120	30
4	140	20
5	150	10
6	155	5
7	157	2

to be feared and, where possible, to be eradicated. There are several reasons for this attitude. In the first place, let us revert to the type of illustration used in Chap. 5 when we were considering consumer problems. Let us imagine that the figures in the accompanying table show the amounts of satisfaction or utility that a given consumer gets from various numbers of bananas. For convenience, let us assume that 1 cent has 1 util, so that we can shift back and forth from cents to utils in our illustration. We learned in Chap. 6 that, if these bananas are sold under competitive conditions by seven separate sellers, each with one banana that he wants to sell, the price cannot be above 2 cents apiece. At any higher price some of the bananas will remain unsold. Thus, under competitive conditions the consumer might buy 7 bananas at 2 cents each, spending a total of 14 cents. If, on the other hand, a single monopolist had owned all 7 bananas, he could have eaten 4 of them himself and sold me the other 3 at a price of 30 cents apiece. Thus, he would get 90 cents by selling 3 instead of 14 cents by selling 7. Control of supply often makes it profitable to limit the quantity sold, and to sell a smaller quantity at a higher price rather than a larger quantity at a lower price. Such

[1] See Sec. 16 • 5.

restriction of sales may be profitable to the seller, but it is manifestly disadvantageous to the consumer, who pays a larger sum of money for a smaller quantity of goods. Of course, it makes no difference from the consumer's standpoint whether one man owns all seven bananas, or whether seven separate owners agree to fix prices in collusion and to divide the income among them.

In our numerical illustration there is one additional thing that a monopolist might do. He could, as we have seen, set his price (say, at 30 cents per banana) and let the consumer buy as many or as few as he pleased (in which case our consumer would buy not more than 3 bananas). But he might not give the consumer even this much choice. He might force the consumer to buy 7 bananas at, say, 22 cents each, or take none at all. Under such circumstances, it would obviously be to the advantage of the consumer to buy, since he would give up $1.52 which would be worth 152 utils, and the 7 bananas would yield him 157 utils. Of course, if the price were set at 22 cents per banana and if the consumer were left free to buy as many or as few as he chose, he would buy only 3 bananas; but, under monopoly conditions, the seller can lay down restrictions on the sale, and in this case he might refuse to sell at all unless the buyer bought 7 units or more.

This simple numerical example shows us why consumers fear monoply. It is because they sense the fact that they may be put at a serious disadvantage as a result of the monopolist's power. They may be forced to pay more money and get less goods in return. But they are put at a disadvantage in another way which is not so evident and which, as a consequence, they less often comprehend. The tendency is, under conditions of competition, for the various factors of production to gravitate to the places where they will produce the greatest value of product. If man power or capital promises to yield more in the production of wheat than in the production of shoes, the producers of wheat are able to bid more for these factors than are the producers of shoes. Under a situation of competitive equilibrium, the marginal output of men or of capital would be the same in all industries. If, now, a monopoly in shoe production were to reduce output there, it would be necessary to shift some of the productive factors from shoe production to alternative industries, such as wheat growing. We would give up shoes, which we might have had, in order to increase our output of wheat. But we had already balanced the two products at the point where the values of the marginal outputs were equal. As we expand the output of wheat, the value of the marginal output will fall in accordance with the Law of Diminishing Productivity[1], but the value of the marginal product of shoes will rise as the output falls. Thus we shall give up a greater value of shoes than we gain in value of wheat. The shift of resources that comes about

[1] See Sec. 14 • 3.

as a consequence of monopoly tends to work to the disadvantage of consumers by giving them an increased output in some lines which is not worth so much as the loss that they suffer from decreased output in the monopolized lines.

25·3. Business Combination. Even a small business, as we have said, may be monopolistic, and even a large business may be competitive; yet it is true that monopoly power has often been obtained by combining business firms that had hitherto been independent and competing, forming one large firm to take the place of several small ones. The mechanism by which the monopolist obtains control over the several original competing businesses differs greatly from case to case, but the aim is always unity of action.

This unity of action is sometimes achieved without actually combining the original companies. Thus it is possible to have 10 separate corporations, each with its own officers and its own board of directors. If each board of directors consists of, say, seven men, and if we can elect the same four men to each board, we shall have a majority of each board which overlaps—which interlocks. These four men who hold common membership in all the boards are in a position to select the managers and determine the policies of the separate corporations. They can, if they wish, make certain that the corporations do not compete, that they maintain a common price policy, or that they divide the market and give each corporation a monopoly within some specified area. Any man who is a director of two or more corporations at the same time is an interlocking director; but, if one of the corporations is a bank, another a church, and a third a hospital, it is unlikely that the interlocking character of the directorate will cause any serious reduction of competition. The three corporations were not competing in the first place. *Interlocking directorates* become important when they occur in corporations that would normally compete with each other. In such cases, there is at least the possibility that they will reduce or eliminate competition. The interlocking directorate makes it possible to get some element of monopoly power while still retaining all the original corporations. No new business firm is established, but the policies of the old firms are unified by the common directors.

Roughly the same sort of result may be achieved by means of a *pool.* When two or more firms enter into a pooling agreement, they merely agree to handle some part of their business in common. Thus hundreds of independent florists pool their advertising appropriations to popularize the slogan "Say It with Flowers." This slogan is not the property of any particular florist, but the joint property of all members of the advertising pool. Such a pooling agreement is not illegal if it does not reduce competition; if a group of independent producers came together to form a price pool, agreeing to set uniform prices and not to undercut each other on the market, the

courts would hold that they were restraining trade in violation of law. Like an interlocking directorate, a pool is not illegal per se but is illegal if it is effective in reducing competition among its members.

Without the formality of a pooling agreement, several businessmen might in practice refrain from cutting prices or soliciting business in territory that they recognized as "belonging" to one of their competitors. Such an arrangement is often called a *gentleman's agreement* and continues only as long as the parties to it voluntarily curb their own activities. The interlocking directorate, the pool, and the gentleman's agreement are alike in that they operate without the formation of new business firms, and with the continuation of all the old firms which, at least on the surface, are "competing" with one another. They are simple mechanisms, and their legality or their illegality depends on their results—on whether or not they do in fact succeed in reducing competition.

The pool and the gentleman's agreement have the advantages that they are easy to form, and they retain the identity and independence of the original companies. Moreover, they can often be kept secret. If they do actually reduce competition, these agreements are not enforceable by the courts, and the arrangements continue only so long as the constituent concerns voluntarily cooperate. Sooner or later one or another of the firms is likely to find it advantageous to break away, and then the agreement is ended. Consequently pooling agreements and gentleman's agreements are often short-lived, and businessmen prefer some more formal and lasting type of combination.

25 · 4. Holding Companies. Perhaps the most common and most effective of the methods of combining separate businesses is the holding company. As we have seen,[1] a corporation is a peculiar type of business enterprise, established by law and given certain limited powers, which are enumerated in its charter. Thus a corporation might be empowered to operate a railroad, sell groceries, or produce moving pictures. In the past half century some corporations have been established, not for the purpose of manufacturing or selling commodities or services, but for the purpose of buying and holding the stock of other corporations, thus obtaining control over these other corporations. A *holding company* is itself a corporation, and its major function is that of controlling other corporations by means of owning their stock. Thus if, in a given neighborhood, we had 10 small electric-light and power companies, each serving its own small community, we might form a new corporation which would buy stock in each of the original 10 until it owned a controlling share of each. We would now have 11 corporations instead of 10, but the eleventh would engage in the business of buying and holding stock rather than in the business of producing and selling electricity.

[1] See Sec. 9 · 6.

As we have seen before,[1] one can always control a corporation by owning a majority of its stock, often by owning even less than a majority. In practice, the holding company might well control the 10 subsidiary companies by owning, say, 30 to 40 per cent of their stock. The holding company could then elect the directors and, through them, establish that unity of action on the part of the 10 power companies which is the essence of monopoly. The holding company offers an inexpensive means of securing such control, since often it is possible to procure the stock of the subsidiary companies without the use of cash. The holding company, being a corporation, issues stock of its own, and the owners of stock in the subsidiaries are often willing to exchange their stock for that of the holding company.

The holding company is a relative newcomer in the business field. New Jersey was the first state to empower corporations to own and to vote the stock of other corporations, taking this step about 1890.[2] Because the provisions of the law were more liberal in this regard than were those of other states, businesses began to incorporate in New Jersey in great numbers. Soon other states, notably Delaware, followed suit lest they lose businesses to New Jersey, and competition among the states led to progressively greater liberality in the laws. Only in recent years has the trend changed, as steps have been taken to curb the growth and power of holding companies.

The holding company raises certain new types of problems of concentration of business power which we have not found it necessary to discuss heretofore. Let us illustrate with a fictitious example. Suppose that each of the 10 electric-light and power companies that we have been discussing is capitalized at $20,000, consisting of $10,000 of bonds, $5,000 of nonvoting preferred stock, and $5,000 of voting common stock. The holding company can gain control of each subsidiary by buying a majority of the common stock. This means that the holding company must buy $2,500 of common stock in each of the subsidiaries—a total investment by the holding company of $25,000. But this $25,000 investment gives the holding company complete control over the $100,000 of assets of the subsidiaries. Suppose, now, that the holding company itself raises the $25,000 necessary to get this control by issuing $10,000 of bonds, $10,000 of nonvoting preferred stock, and $5,000 of voting common stock. One could control the holding company (and through it the 10 subsidiaries with their $100,000 of assets) by owning half the voting stock, or $2,500. Thus the control of vast undertakings can sometimes be concentrated into a very few hands by judicious investment in the right kind of securities. The control may become even more concentrated if the holding companies are "pyramided"—if we have a new holding company that controls several old holding companies by buying and holding their stock. In

[1] See Sec. 9 • 12.

[2] There were actually two acts involved, one in 1889 and the other in 1893.

this way the ownership of property may be effectively divorced from its actual control. Most of the owners under such conditions would be unable to exercise control if they cared to.[1] Although the illustration just given is fictitious, nevertheless there have been many actual cases where control of vast resources has been concentrated to quite as great a degree as that suggested here. For example, in the Associated Gas and Electric system there were 12 companies in one line of control from the top to the bottom of the "pyramid," each company holding a controlling interest in the next one. The Insull interests at one time controlled the West Florida Power Company by the ownership of five one-hundredths of 1 per cent of its total securities; and the same interests controlled the North State Beach Development Company by the ownership of but two one-hundredths of 1 per cent of its total securities. At one time an investment of $1 in the common stock of the Corporation Securities Company could exercise control over $2000 invested in properties at the bottom of the pyramid.[2] These are, to be sure, extreme cases; yet they illustrate the possibilities inherent in the holding company as a mechanism for corporate control. In these cases, as in many others, instead of having "majority control," we have control by a very small minority interest. A relatively small investment, judiciously made, enables the investor to control vast sums of money invested by other people.

As a form of business enterprise, the holding company has the advantages that it is relatively easy to form and that it preserves the identities of the constituent subsidiary firms. Thus, if these constituent firms have established their own trade-marks and brands and built up "good will," the holding company can continue to profit by continuing them. In some ways, the holding company is cumbersome. It is necessary to retain the separate corporate organizations of the constituent companies, each with its own officers and board of directors, its own annual meetings, and each paying its own corporation taxes.

25·5. Trusts. The word "trust" is used in several different senses, even if we think only of those trusts that are business organizations. When one speaks of the "trust-busting days" under President Theodore Roosevelt, he is using the term in a very different sense from that which he has in mind when he speaks of "banks and trust companies." In this chapter we shall confine our attention to the use of trusts as means of combining a number of separate

[1] The student should realize that most owners do not seem to want to exercise control over the corporations in which they have invested. Not only are they willing to invest in types of securities that have no voting power but, even when they do own common stock, they seldom exercise their right to attend meetings of the corporation and to vote for directors. They think of their stock holdings purely as an investment, and not as a source of voting power and corporate control.

[2] These illustrations are from Part 72-A of Senate Document 92, 70th Congress, 1st Session, pp. 159-161.

business firms; yet even here we can understand the problem most easily by starting with a description of the use of the trust for other purposes.

It is not uncommon for men to leave property "in trust" for other people. Thus, a man may leave his estate in trust for his minor children, or for his wife; or he may establish a trust fund the proceeds of which are to be used to assist worthy Presbyterians to go through Notre Dame. In such cases there are three parties involved (1) the *settlor,* who is the person putting the property in trust; (2) the *trustee,* who is the person to whom the management of the property is entrusted—the person who actually controls the property; and (3) the *beneficiary,*[1] who is the person for whose advantage the property is administered. Trusts of this nature are very old and very common, and unquestionably of social advantage. For centuries they have existed under the common law, and they continue to exist in large numbers today.

Not until 1879, was the trust, which we have just described, turned to business purposes. Then the Standard Oil Company was organized in the form of a trust, and for the next few decades this type of organization, originally established for nonbusiness reasons, was used increasingly to effect business combinations. The original form was retained, but the purpose was altered. We can illustrate the way in which the trust was adapted to business uses by assuming again that there are 10 independent small electric-light and power companies in a given neighborhood and that someone wishes to get control of them. This time, instead of forming a new corporation, the holding company, the holders of a majority of the stock in each of the constituent companies turn their stock over to a board of trustees, to be managed for the benefit of those people who originally owned the stock. The settlors and the beneficiaries are the same people. We turn over our stocks to a board of trustees and tell them, "Manage these businesses for our benefit." There is no new corporation—merely a board of trustees, which is usually self-perpetuating. The trustees usually issue "trust certificates" to the beneficiaries, to represent their interests in the undertaking; and, just as the holding company commonly exchanges its own stock for the stock of the subsidiaries, so the trust often exchanges trust certificates for the stockholdings of the settlors. In this way, little capital is required in forming a trust and, if sufficient stock of the subsidiary companies is turned over, the trustees can effectively control all the constituents.

In recent years, the courts have been increasingly skeptical of this perversion of the trust to business uses and, wherever such a trust organization

[1] We have here used settlor, trustee, and beneficiary in the singular, although there might, in practice, be several of each. Thus five men might combine, as settlors, to put their property, or part of it, in trust under a board of eight trustees, to be administered for the advantage of twenty beneficiaries.

serves in fact to reduce the competition among concerns that had formerly been competing, the trust is clearly illegal.

25·6. Merger or Consolidation. More and more in recent years, as the law has tended to discriminate against the holding company and the trust, the tendency has been to buy up competing companies, lock, stock, and barrel, and form one new corporation to replace the old ones. Such a combination of several companies is called a "merger" or a "consolidation." The old companies completely disappear. Their assets and liabilities are taken over by the new concern. Thus, if we return to our illustrative case, the formation of a holding company would add a new corporation, giving us 11 corporations in all. The trust would retain the original 10 corporations and would set up in addition a new board of trustees. But, if we merged the 10 companies, we would set up a new corporation which would buy out the old ones, and the original 10 corporations would disappear, leaving but one new large corporation in their place. This might have the disadvantage of losing the good will of the old concerns, but it would have the advantage of simplicity in organization; in addition, the status of the new concern under the law would be somewhat clearer. In the case of the holding company or of the trust it might be charged that the 10 constituent companies (all of which would still exist) were not competing with each other and that there was collusion in restraint of trade. But there is no problem of collusion between competing companies in the case of the merger, since there are no separate companies either to compete or to enter into collusion. There is but one company where there were ten.

25·7. Horizontal and Vertical Combination. In our discussion up to this point, we have been considering various methods of effecting the combination of separate business enterprises without considering the types of concerns to be combined. In practice, there are considerable differences in this regard. Sometimes a combination brings together a number of business concerns of the same general kind. Thus, in our illustrations, we have been assuming that several electric-light and power companies were to be combined by means of a holding company, a trust, or a merger. In such a case, when the companies combined are engaged in the same kind of business, we call the resulting combination a "horizontal combination." Thus, a chain store, such as the Atlantic and Pacific Tea Company, can be taken as an illustration of horizontal combination. But often the companies combined are doing different kinds of business. Thus, Henry Ford, describing the 88 plants of the Ford Motor Company in the United States and 28 foreign countries has said[1]:

> We are in the following lines of business, every one of which grows out of making motors: aeroplanes, coal mining, coke manufacture, by-products manu-

[1] Henry Ford in collaboration with Samuel Crowther, "Today and Tomorrow," 1926, p. 40.

facture, lead mining, iron mining, foundry, steel manufacture, tool making, machinery manufacture, car truck and tractor manufacture, glass manufacture, artificial leather, copper wire, Fordite, textiles, batteries and generators, paper, cement, automobile bodies, Johanssen gauges, electric power, filtered water, flour, motion pictures, hospital, farming and stock raising, radio, printing, photography, forging, flax growing, steam turbine, electric locomotives, logging, sawmills, body parts, dry kilns, wood distillation, products of hydro-electric power, grocery stores, show stores, clothing stores, butcher shop, railroads, educational, ocean transportation, lake transportation, tractors, and automobiles.

These constituent companies are engaged in varied businesses, but they all contribute directly or indirectly toward the major end of manufacturing automobiles. Such a combination, bringing together businesses of different kinds that contribute toward a common product, is called a "vertical combination," and the movement toward combining such diverse businesses is called "integration of industry." Either the horizontal or the vertical type of combination can be effected by any of the methods suggested earlier: by a holding company, a trust, consolidation, etc.

25 · 8. Indications of the Presence of Monopoly. Since, in practice, there is usually neither complete competition nor complete monopoly in any industry and since the degree of monopoly present differs by minute graduations from situations that are very nearly competitive to those which evidence a very great degree of monopoly, it is obviously impossible to draw any sharp line of demarcation distinguishing those industries which are monopolistic from those which are not. Especially in the middle ground there is always room for dispute and honest difference of opinion. Yet there are some things that any experienced investigator would look for as possible or probable indicators of monopolistic elements in an industry. In a competitive market, for example, the price of the product tends to vary with every change in supply or in demand. The more competitive the market, the more sensitive is the price to market conditions; and on the most highly organized competitive markets, such as the great organized exchanges dealing in produce or securities, it is not uncommon to have the price change almost constantly, with minute-by-minute variations in the quotations. The monopolist, on the other hand, can set the price where he wishes, and the tendency often is for those who have some element of monopoly control to keep the price of their product fixed for considerable periods. In such cases, a change in demand would be reflected in a change in the quantity taken by consumers without any change in the price. Therefore, when the economist finds an industry with fixed prices and fluctuating output, he becomes suspicious that there may be some element of monopoly control. This suspicion is likely to continue even if there are some changes in price as long as the price changes are minor and the changes in volume of sales are

major. If we classify the 747 commodities included in the wholesale price index of the U. S. Bureau of Labor Statistics according to the numbers of times that their prices changed in the 96 months from 1926 to 1933, we get the data of Table 25·1. The price of each commodity was listed once each month, and it was therefore possible for the price to change 95 times during the period. In fact there were 125 commodities that showed price changes 90 times or more out of the possible 95; there were also 95 commodities that showed 4 or less price changes during the 8 years.[1] Although one would cer-

Table 25·1. Numbers of Commodities with various Numbers of Price Changes in 96 Months

Number of Times Price Changed	Number of Commodities
0–4	95
5–9	97
10–14	79
15–19	58
20–24	45
25–29	23
30–34	29
35–39	34
40–44	16
45–49	12
50–54	9
55–59	18
60–64	15
65–69	15
70–74	12
75–79	9
80–84	19
85–89	37
90–94	125
Total	747

tainly want further information before making up his mind on the matter, he might well at least be suspicious in the cases of the 200 or more commodities that changed in price only ten or fifteen times or less in 8 years.

A second indicator of possible monopoly is the concentration of a large part of the production of a commodity in the hands of a single firm, or of only three or four concerns. Suspicion would be aroused even if there were many producing plants if a substantial number of these were operated by a few owners. The mere fact that a concern is large is not enough. A million-dollar concern in a million-dollar industry would have complete control;

[1] The data are from Gardner C. Means, Price Inflexibility and the Requirements of a Stabilizing Monetary Policy, *Journal of the American Statistical Association*, vol. 30, June, 1935, p. 402. The figures given here are read roughly from his chart.

while a ten-million-dollar concern in a ten-billion-dollar industry might well have many competitors. It is the relative rather than the absolute size of the firm that counts—the size of the firm as compared with the size of the industry. Some indication of the situation in the United States is given in a study based on the census of 1937. If the 1807 different products listed there are classified according to the degree of concentration of production in the hands of the four largest firms in the industry, we get the data of Table 25 • 2. Here we see that 840 of the 1807 products were produced in industries so highly concentrated that 75 per cent or more of the total output was produced by the four largest firms; while only 402 commodities came from industries in which the four largest firms produced less than half the total output. Over a third of the total value of output came from the highly concentrated businesses that had over three quarters of the output concentrated in the hands of

Table 25·2. Degrees of Concentration of Production for 1807 Products Listed in the Census of 1937

Percentage of total output supplied by the four largest firms	Number of products	Percentage of total value of all products
75 or more	840	34.4
50–74	526	23.1
49 or less	402	41.7
Data incomplete	39	0.8
Totals	1807	100

the four largest producers.[1] It is evident from these figures that monopoly elements were of considerable importance in the United States in 1937.

Although a high degree of concentration of production and an unusual stability of prices are indicators of possible monopoly, there are many other things that would be apt to arouse suspicion. Among these would be the rendering of identical bids on a contract by a number of supposedly independent and competing concerns, the charging of different prices to different classes of consumers for the same or substantially the same commodity or service, the dividing of the market into territories allotted to particular producers, or the existence of a "basing-point system" under which producers quote prices at some common basing point plus delivery charges from that basing point even though the producers' plants are at scattered locations.

25 • 9. Growth and Present Importance of Monopoly in the United States. The data of Tables 25 • 1 and 25 • 2 indicate that there is a considerable degree of monopoly in the United States today. To be sure, complete monopoly

[1] Data from *T.N.E.C. Monograph* No. 10, Industrial Concentration and Tariffs, p. 6, by Clifford L. James, Edward C. Welsh, and Gordon Arneson.

in the theoretical sense is unheard of; and it is almost certainly true that competitive forces play a more important role than do monopoly forces in shaping our peacetime economy. Yet, there are many indications that monopoly forces are important and their importance is increasing. In fact, many rather recent policies of our national government have tended to foster and encourage monopoly in fields where it was formerly uncommon.

Tables 25·3. Sizes of Firms Engaged in Manufacturing, United States, 1939

Size of firm, wage earners	Percentage of firms	Percentage of all employees
0	4.5	0
1–5	41.2	2.6
6–20	26.6	6.9
21–50	12.8	9.7
51–100	6.5	10.8
101–250	5.1	18.7
251–500	2	16.1
501–1000	0.8	13
1001–2500	0.3	11.9
2501 and over	0.1	10.5

Table 25·4. Sizes of Firms Engaged in Manufacturing, United States, 1939

Size of firm, value of products, thousands	Percentage of all firms	Percentage of all employees
$5–$19	33	2.4
20–49	22.8	4.1
50–99	13.9	4.9
100–249	13.4	9.7
250–499	7.1	10.6
500–999	4.7	13.2
1000–2499	3.3	18.6
2500–4999	1.1	11.5
5000 and over	0.8	24.9

Although mere bigness is not in itself sufficient to constitute monopoly, nevertheless the trend is in part reflected by the increasing size of business in the United States. This may be measured by the number of employees, the value of the product, the amount of the investment, and in other ways. Table 25·3 shows the percentage of manufacturing firms employing various numbers of wage earners in 1939, and the percentage of all wage earners employed by firms of each size.[1] It is evident that more than half of all

[1] Data are from the "Census of Manufactures, 1939," vol. I, p. 120.

persons employed in manufacturing were working for firms employing over 250 workers. Table 25·4 shows the percentage of manufacturing firms producing various values of product in 1939 and the percentage of all wage earners employed by firms of each size.[1] Here we see that more than half of all persons employed in manufacturing were working for firms that produced more than a million dollars' worth of product per year. If we were arbitrarily to call those firms "big" which were in the upper half and those businesses "small" which were in the lower half, then in manufactur-

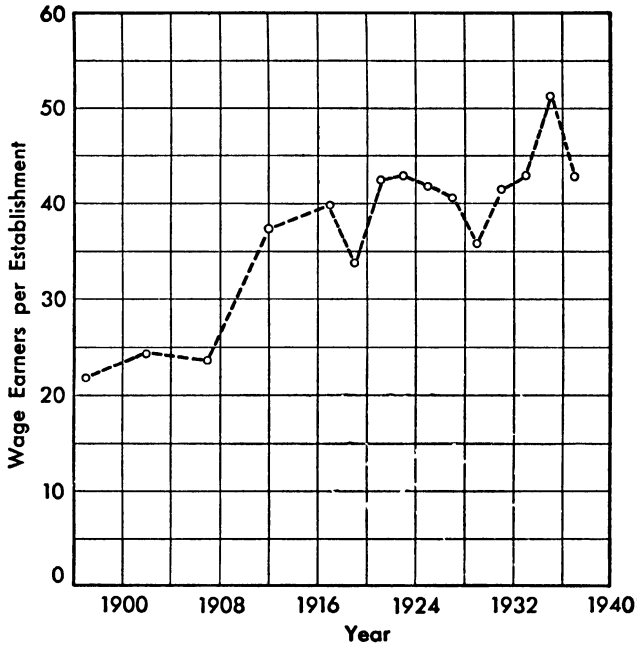


FIG. 25 · 1.—Average number of wage earners per manufacturing establishment, United States, 1899-1939.

ing we would say that a firm was a big firm if it employed more than 250 workers and produced more than a million dollars' worth of product. Of course in other lines (transportation, retailing, finance, etc.) it is quite likely that the dividing line should be put at some other point.

The tendency for firms to increase in size with the passage of time is indicated by Figs. 25 · 1 and 25 · 2, which show the increase in the average number of wage earners per establishment and in the value of output per establishment in manufacturing in the United States. Figure 25 · 1 indicates that the average number of wage earners in manufacturing plants grew

[1] *Ibid.*, p. 182.

from just over 20 to over 40 during the first four decades of the present century, while Fig. 25·2 indicates that the average value of output increased from about $75,000 to about $250,000 during the same period.[1] The number of employees has about doubled, while the value of output has about trebled. The typical manufacturing business of today is thus two to three times as large as that of 40 years ago.

As we saw in Chap. 9, it is difficult to carry on a large-scale business under other than the corporate form. Corporations may be small businesses; but large businesses must almost of necessity be corporations. For this

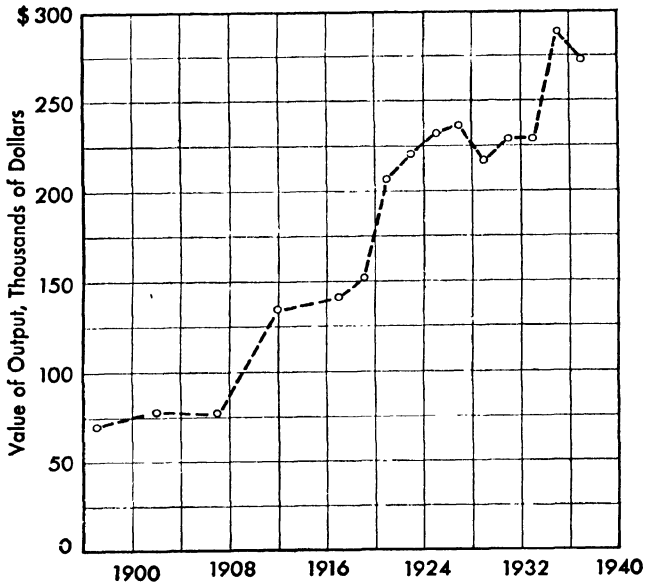


FIG. 25·2.—Average value of output at 1910-1914 prices of manufacturing establishments in the United States, 1899-1939.

reason, the growth of big business depends to a very significant extent on the development of the corporate form of enterprise. The present outstanding importance of the corporation is indicated by the figures of Table 25·5. Here we see that barely over half of the manufacturing concerns of the country in 1939 were organized as corporations, but that these firms employed 17 out of every 19 workers and produced thirteen-fourteenths of the value of manufactured products. The significance of the corporation in other branches of economic activity is portrayed in Fig. 25·3. Here we see that the corporation is still relatively unimportant in agriculture and that

[1] The charts are based on data from the "Census of Manufactures, 1939," vol. III, p. 19. The figures on value of output have been corrected for changes in the general price level, and thus reflect approximately changes in the volume of output at 1910-1914 prices.

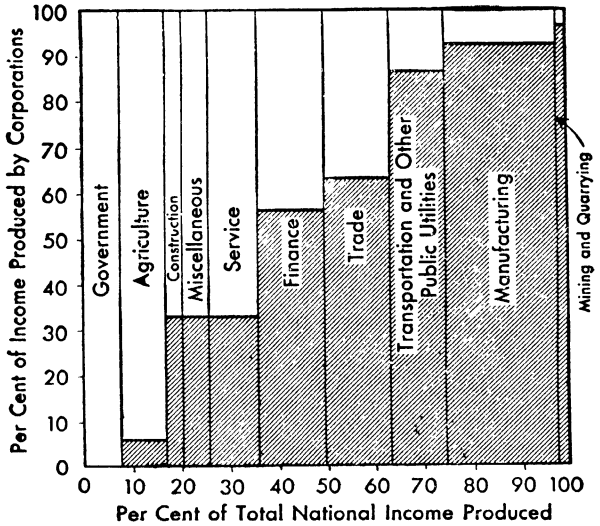


FIG. 25 • 3.—Relative importance of various branches of economic activity, 1929, and relative importance of corporations in each branch, 1933. The widths of the bars vary with the volume of production in the various lines of endeavor. The heights of the bars represent the importance of the corporate form of organization in each of the lines of endeavor. Thus, the final bar at the right is narrow because mining and quarrying account for but a small part of the total productivity; but this bar is very tall because almost all mining and quarrying is carried on by incorporated businesses. (*"Big Business: Its Growth and Its Place," Alfred L. Bernheim, editor; Twentieth Century Fund, Inc.; New York,* 1937; *pp.* 17-18.)

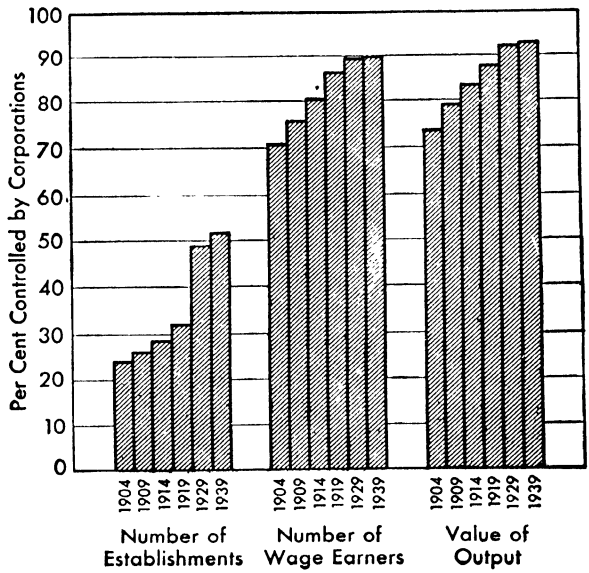


FIG. 25 • 4.—Growth in the importance of the corporate form of organization as evidenced by the percentage of all establishments operated by corporations, the percentage of all workers employed by corporations, and the percentage of the value of all output produced by corporations, 1904-1939. (*Based on Succeeding Censuses of Manufactures.*)

it has not yet become outstandingly important in the field of construction or in the service industries. In the fields of finance, trade, transportation, public utilities, manufacturing, mining, and quarrying, the corporation has come to be more important (if we measure importance by value of production) than all other types of organization combined and, if we lump together all production of all kinds, it is estimated that some 57 per cent of the total output of the United States is produced by corporations. The growth of the importance of the corporation in the field of manufacturing is depicted in Fig. 25 • 4. For further data on the subject, see Sec. 9 • 8.

Table 25 • 5. Relative Importance of the Corporation in Manufacture. United States, 1939*

Type of business organization	Percentage importance as measured by		
	Number of firms	Number of employees	Value of products
Corporations	51.6	89.4	92.8
Individual proprietors	31.9	5.6	3.5
Partnerships	15	4 7	3.3
All others	1.5	0.3	0.4
Total	100	100	100

* Based on data from "Census of Manufactures, 1939," vol. I, p 268

The two most pronounced periods of concentration in American business were, according to one of the outstanding authorities, the years 1897 to 1905 and 1925 to 1929. In both cases, the tendency to combine and concentrate was related largely to schemes for financial promotion rather than to the adoption of new techniques of production.[1] Although any attempt to classify businesses as big or small must be more or less arbitrary and although the criteria marking a firm in one industry as big may not be at all the criteria that should classify as big a firm in another industry, nevertheless one student has tried to separate the establishments in the various industries into size classes by criteria that seem to him to fit the industries in which the several establishments are located, and on this basis to determine the relative importance of big and small business in various fields of endeavor.[2] His results, which are approximate at best, are shown in Fig. 25 • 5.

[1] See statement by Willard Thorp, *T.N.E.C. Monograph,* No. 27, p. 231. On the two following pages Thorp gives detailed figures on the numbers of mergers each year from 1919 to 1939.

[2] WEISSMAN, RUDOLPH L., "Small Business and Venture Capital," p. 11, Harper & Brothers, New York, 1945.

25 · 10. Advantages of Monopoly Control. There are several reasons why businessmen desire to get as great a degree of monopoly control as they can. If we recall the character of monopoly,[1] we remember that the monopolist, unlike the competitor, can raise the price of his product without losing all of his business. There are either no substitutes for his product or, at least, none that consumers think are "just as good," and they continue to patronize the monopolist in some degree despite his higher price. Every businessman tries

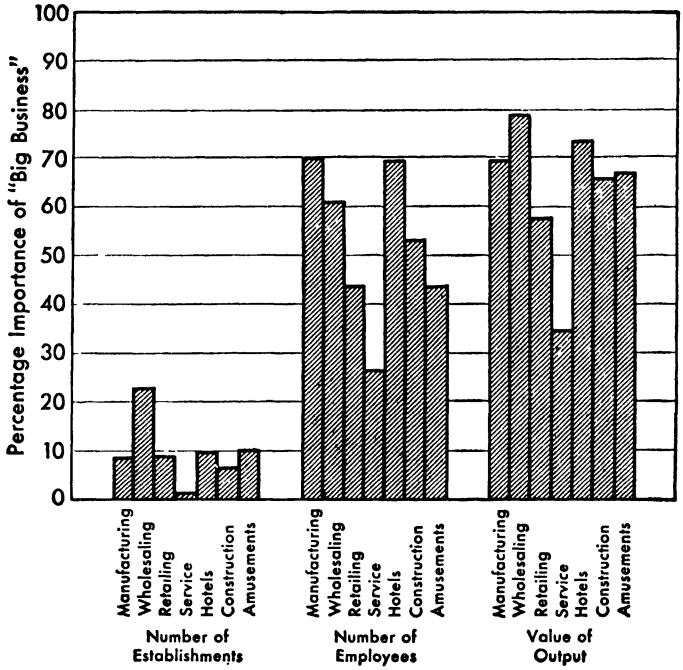


FIG. 25 · 5.—Relative importance of big business in various fields of endeavor, United States, 1939; as measured by numbers of establishments, numbers of employees, and value of output. (*Based on data of the U. S. Census and Rudolph Weissman's definitions of "bigness"; adapted from Rudolph L. Weissman, "Small Business and Venture Capital," Harper & Brothers, New York,* 1945, *p.* 11.)

by various methods to take his product out of competition—to establish a demand for his particular product as distinct from that for the products of other makers. The layman, in appraising this situation, is apt to think that the power to raise prices is the sole reason for monopoly—the sole motive actuating the businessman who tries to eliminate competition. Although the power to set prices is without question an important motive, it is by no means the only one.

[1] See Sec. 16 · 1.

A second motive that has stimulated men to form monopolistic combinations has certainly been the desire to make promoter's profits, those which come from the floating of the new issues of securities rather than from the operation of the businesses themselves. Thus, when the United States Steel Corporation was formed in 1901, the value of the concerns taken into the new combination is estimated to have been between 700 and 800 million dollars, yet the new concern was capitalized at over 1,321 million dollars. Especially in times of speculative optimism, it seems to be possible to convince investors that 10 companies combined are worth far more than the same 10 companies separately, and the promoters who bring them together may therefore be able to reap substantial profits.

These two motives for combination—the power to set monopoly prices and the desire for speculative promotional profits—are often frowned upon and decried as socially disadvantageous. Another possible incentive for monopoly control is often suggested, which may well be of advantage to both the monopolist and the general public. This is the ability of the monopolist, by controlling a larger volume of business, to reduce his costs of production. If, for example, we have 10 competing concerns, each with one-tenth of the total volume of business, and if they produce an article at an actual cost of $8 per unit, they will make no profit at all if the product sells for $8 per unit on the market. If, now, a single concern takes over the entire industry, merging the former concerns and building up under one management a volume ten times as large as any of the former concerns had, the new monopolistic concern may be able to make certain economies of operation as a result of the large scale of its business. Suppose that the new firm, as a result of these economies, can cut the cost of production to $6 per unit of product. If the new concern now reduces the selling price on the market from $8 to $7, it will make a profit of $1 on each unit sold. The public in such a case enjoys the product at reduced prices, and the monopolist makes a profit which the competitive firms were unable to make. It is obvious that such a situation could exist and that monopolistic control might result in savings to consumers and at the same time in profits for the monopolist. Yet the fact that this might happen (and probably sometimes does happen) is no reason for us to assume that it must happen or even that it usually happens.

The student sometimes has difficulty in understanding how one can make any reduction in cost merely by bringing several independent competing concerns under a single management. Yet there are a number of ways in which this may actually be done. For example, the new combination may buy its raw materials for all 10 plants at once, placing orders of much greater size than any of the individual plants could do and thus getting *quantity discounts*. These large orders may also make it possible to take advantage

of *lower freight rates* on carlot shipments, and these lower rates may also be available on shipments of finished products in larger quantities. The combination may be able to *eliminate cross-shipping* by assigning any order received to the plant nearest the customer. Thus, if the combination has one plant in Denver and another in St. Louis, it is quite possible for the Denver plant to receive an order from a St. Louis firm and for the St. Louis plant to receive an order from a Denver firm. If the plants are under separate management, the goods will be shipped over long distances with considerable expense; if the plants are under unified management, the St. Louis plant will transmit its order to the Denver plant, and the order will be filled there with little or no shipping cost. The large-scale operations of the combination may make it possible to *use better methods of production,* such as expensive machinery which would have been out of the question for a small firm but easily available for a large one. These large-scale operations also commonly make it possible to *utilize by-products* which could not be reclaimed profitably in a small business. Thus, for example, the small-town butcher could not possibly afford to extract the glandular secretions used in medicine and surgery, which the great meat packers do in fact reclaim and sell. The combination might easily make a *saving in selling expense* by sending a single salesman into a given territory where each of the former independent companies had to send a man. A large organization may get better *balance of equipment* than is possible in a small concern. For example, it may be that the productivity of two machines is such that they can best be combined in the proportion of two to seven—two machines of type *A* to seven machines of type *B*. In this combination the machines exactly keep step with each other. If a small concern has business enough to keep only one or two machines of type *B* operating, it will either have to forego the use of machine *A* entirely and substitute some inferior process, or it will have to use one machine of type *A* and use it inefficiently. A large organization may be able to use *more effective division of labor* with all the advantages described in Chap. 10. These and other *economies of large-scale operation* may well make it possible for the combination, like the hypothetical one in the preceding paragraph, to reduce prices to the public while increasing profits for itself.

25·11. Does Combination Actually Increase Efficiency? The preceding paragraph pointed out that increased size and unity of control may bring certain economies. In practice, it seems to have been true that combinations have more often been formed to take advantage of their monopoly power or to gain the speculative returns incident to the flotation of new corporate securities than because of actual economies in production or in distribution of the goods. When, in 1940, the Federal Trade Commission was investigating the relative efficiency of businesses of different size, they consulted

Dr. Myron W. Watkins of New York University, a recognized authority on mergers and consolidations in American business, and the Commission asked him, "Would you say that in general a prominent or even important factor in stimulating the mergers in American industry over the past half century has been the prospect of cheapening the cost of production or of improving the quality of the product or service sold?" Dr. Watkins' answer was "No."[1]

It must be remembered that, although it is possible to make certain economies when a business is operated on a large scale, as was pointed out in the preceding section, nevertheless there are limits to the advantage to be gained from mere increase in size. Ultimately, the continued growth of a concern is apt to bring *diseconomies of large-scale operation* which counterbalance the economies already enumerated. To quote one authority on the subject,

> It has been observed that after a point the increasing diseconomies of larger scale more than offset the economies in other directions. In large measure these diseconomies appear in the increasing inability of the management to maintain an effective control over the various parts of the organization, with the result that there is a decline in efficiency and a rise in unit costs. Furthermore the scale of operations at which these diseconomies begin to assume significant proportions is perhaps somewhat smaller than is commonly supposed. If it were not for the quasi-monopolistic advantages which tend to accompany increasing size it seems likely that the technically optimum size of firm in most industries would be considerably smaller than it actually seems to be.[2]

The "inability of the management to maintain an effective control over the various parts of the organization" which Buchanan mentions is easily understood when one envisions the tremendous size and complexity of some modern business combinations. The Federal Trade Commission gives two examples[3] that are telling enough to warrant quoting in detail. The Commission says,

> Students of the problem of business efficiency often wonder how a board of directors of even 15 or 20 extraordinary businessmen could be soundly acquainted with the labyrinthic production program of these conglomerates. Following are examples of the multiplicity of products and diverse businesses which have been brought under a common management in the American business structure:

[1] *T.N.E.C. Monograph,* No. 13, Relative Efficiency of Large, Medium-sized, and Small Business, Study made by the Federal Trade Commission, p. 98.

[2] BUCHANAN, NORMAN S., "The Economics of Corporate Enterprise," p. 153, Henry Holt and Company, New York, 1940.

[3] *T.N.E.C. Monograph,* No. 13, *op. cit.,* pp. 117-119.

Example 1.

1. Dyestuffs.
2. Dyestuff intermediates.
3. Organic chemicals.
4. Chemicals and dyes for the petroleum industry.
5. Sulphur dioxide.
6. Nitrated filter cloths.
7. Detergents.
8. Wetting-out agents.
9. Textile assistants and finishing agents.
10. Solvents.
11. Aromatics.
12. Perfume.
13. Photographic and pharmaceutical chemicals.
14. Antioxidants.
15. Vulcanization accelerators.
16. Rubber colors.
17. Neoprene.
18. "SDO2" corrosion-resistant coating.
19. Neoprene latex.
20. Rubber peptizing agents.
21. Accelerator retarders and activators.
22. Special chemicals for rubber.
23. All completely and specially denatured alcohol formulas such as antifreeze alcohol and industrial solvent
24. Motion picture film.
25. Portrait film.
26. X-ray safety film.
27. Lithopone.
28. Titanium dioxide.
29. Extended titanium pigments.
30. Leaded zinc oxide.
31. Aluminum hydrate, and
32. Dry colors.
33. Acetate rayon yarn and staple fiber.
34. Commercial cellulose acetate flake.
35. Viscose rayon yarn and staple fiber
36. Cellulose film.
37. Cellulose caps and bands for sealing containers.
38. Transparent seamless tubing.
39. Cellulose sponges.
40. Ammonia and allied products.
41. Methanol.
42. Higher alcohols.
43. Urea.
44. Fertilizer compound.
45. Hydrogenated products.
46. Methacrylate resins.
47. Aliphatic acids and esters.
48. Anhydrous ammonia in cylinders.
49. Aqua ammonia in drums and tank cars.
50. Ammonium carbonates.
51. Pyroxylin-coated and impregnated fabrics.
52. Sponge-rubber non-skid underlay.
53. Sponge-rubber rug cushioning
54. Pyroxylin-impregnated washable window-shade cloth.
55. Rubberized flexible ventilating duct
56. Rubberized fabric for heavy duty chair and automotive upholsteries.
57. Rubberized fabric for outerwear garments.
58. Latex-saturated fiber for midsole and innersole materials.
59. Hospital sheeting.
60. Neoprene-treated textiles.
61. Nitrocellulose.
62. Solvents.
63. Leather cements.
64. Pyroxylin solutions.
65. Paints.
66. Varnishes.
67. Synthetic-resin enamels.
68. Bronze powder.
69. Special finishes for automobile refinishing and all industrial purposes.
70. Polish.
71. Automobile wax.
72. Top dressings.
73. Protective cream and other automobile specialties.
74. Cellulose nitrate plastic.
75. Cellulose acetate plastic.
76. Methyl methacrylate resin, all in the form of sheets, rods, and tubes.
77. Molding powders.
78. Resin denature material.
79. Monofilaments and bristling filaments.
80. Toiletware, including brushes, combs, hair ornaments, and novelties.

81. Sodium.
82. Cyanides.
83. Peroxides.
84. Chlorine.
85. Heat-treating salts.
86. Formaldehyde.
87. Hexame thylenetetramine.
88. Gold, silver, and platinum preparations.
89. Ceramic materials of all types.
90. Methyl chloride and other refrigerants.
91. Noninflammable solvents and other chemicals for all industries.
92. Preparations to prevent and control many seedborne plant diseases.
93. Preparations to improve crop yields.
94. Shotguns.
95. Rifles.
96. Cartridges and shot shells.
97. Traps.
98. Targets.
99. Gun grease and oil.
100. Rust remover.
101. Powder solvent.
102. Complete line of pocket knives.
103. Professional and household cutlery.
104. Dynamite.
105. "Nitramon" blasting agent.
106. Black powder in granular and pellet form.
107. Blasting accessories.
108. Complete line of products for all industrial and agricultural uses.
109. Torpedoing oil and gas wells with liquid and solidified nitroglycerine to increase productivity.
110. Chemicals for leather, manufacturers.
111. Chemicals for textiles, manufacturers.
112. Chemicals for paper manufacturers.
113. Chemicals for paint manufacturers.
114. Chemicals for rubber manufacturers.
115. Chemicals for battery manufacturers.
116. Chemicals for ceramics manufacturers.
117. Chemicals for sheet, tin plate, and steel manufacturers.
118. Chemicals for petroleum manufacturers.
119. Chemically pure acids.
120. Electrolyte.
121. Filter alum.
122. Insecticides and
123. Fungicides for fruits, vegetables, lawns, and flowers.
124. Silicate of soda.
125. Soldering fluxes.
126. Zinc.
127. Wood preservatives.

Example 2.

This corporation runs at the same time businesses producing numerous breakfast cereals, animal feeds, gelatine, ice cream mixtures, a medley of desserts, a coffee and tea business, a coconut-meal business, a cake and bread business, a chocolate and cocoa business, a sirup business, a nut business, a salt business, a baking powder business, a business of manufacturing laundry aids, an oyster business, a business of producing frosted foods, a business of processing corn products, a business of canning fruits and vegetables, a business of manufacturing cottons and shipping cases and bags, a business of manufacturing tin cans, and a business of meat packing. This corporation was formed by promoters in the 1920's.

Critics of excessive size in business who view this kind of industrial phenomenon often ask, "Can one man, acting as president of either of these corporations, possess the capacity to know enough about such numerous and diverse businesses to effectively administer them from the standpoint of elim-

inating waste, promoting efficiency and technical progress in all of them at the same time?[1]

Although these two lists of activities prove nothing about the particular cases involved, they do help to illustrate how the increasing complexity of large-scale business may bring it to a point beyond which efficient management is difficult or impossible. If we wished to appraise the facts for a par-

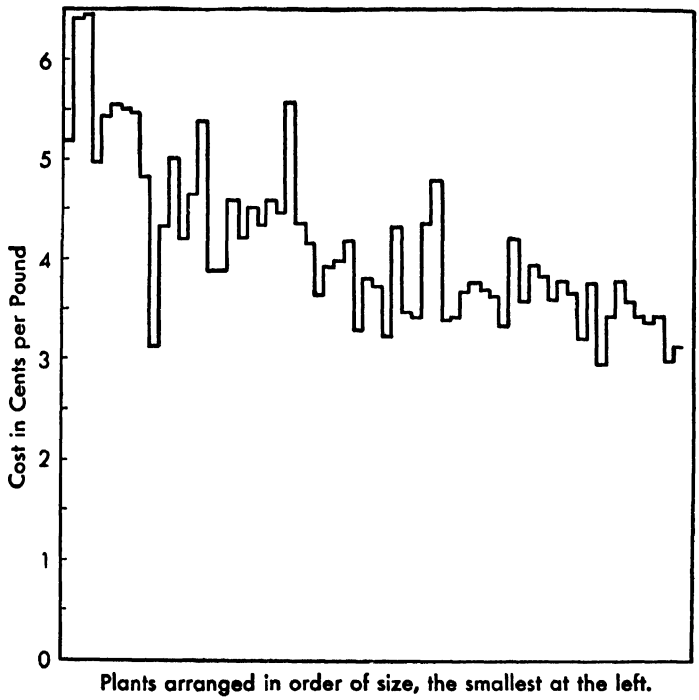


FIG. 25 · 6.—Cost of producing beet sugar in 64 plants. The plants are arranged in order of size, and in general larger plants seem to have lower costs.

ticular case, we would have to ask whether a large business combination actually manages to turn out products at lower costs than was possible when the constituent parts were separate. On this matter, we can find some fragmentary and conflicting evidence.

[1] The student should guard himself against accepting too easily the innuendo of the final paragraph quoted above. It would be easy to enumerate a long, long list of the activities and responsibilities that fall under the general supervision of the president of any large university—in fact, to make a much longer list than either of those given above—and to imply that no man can effectively supervise such varied types of activity. The student must remember that good administration consists largely in finding good subordinates who are able to make the minor decisions and leave only decisions on matters of major policy for the higher officials and for the directors.

The Federal Trade Commission summarized the results of a study of this problem[1] in these words,

> In the 233 combined tests, large size, whether represented by a corporation, a plant, a group of corporations, or a group of plants, showed the lowest cost or the highest rate of return on invested capital in only 25 tests. In these combined tests, medium size made the best showing in 128 tests and small size in 80 tests. Thus large size was most efficient, as efficiency is here measured, in approximately 11 percent of the total tests, medium size was most efficient in approximately 55 percent of the tests, and small size was most efficient in approximately 34 percent of the tests.

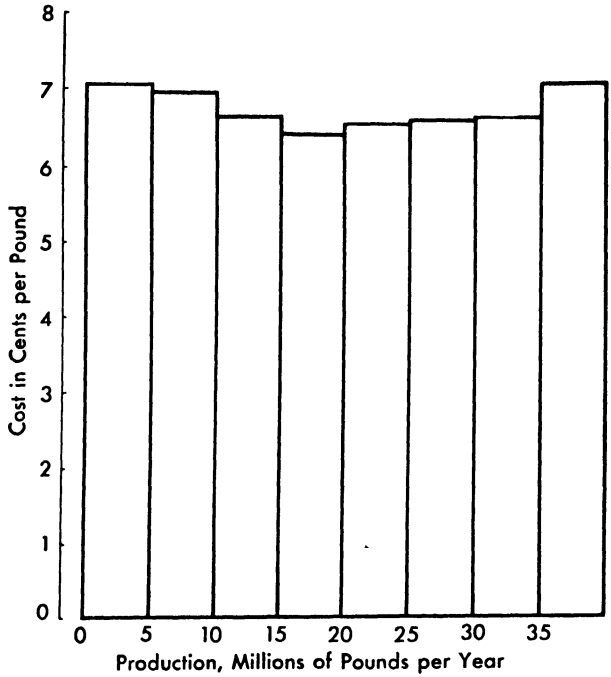


FIG. 25 • 7.—Variation in cost of producing bread in baking plants of various size, 1922-1925. The right-hand bar represents all plants producing over 35 million pounds per year. (*Willard L. Thorp, The Changing Structure of Industry, "Recent Economic Changes,"* p. 189.)

Figure 25 • 6 shows the cost of producing beet sugar in each of 64 plants, with the plants arranged in order of size.[2] One notes immediately that there are considerable differences in costs for plants of approximately the same size, yet these data seem to indicate a general tendency for costs to fall as size of plant increases, and the tendency seems to be continuing even at the

[1] *T.N.E.C. Monograph,* No. 13, *op. cit.,* p. 14.

[2] *T.N.E.C. Monograph,* No. 13, *op. cit.,* p. 45.

right-hand end of the diagram, where the largest plants are depicted. Figure 25·7, however, seems to indicate that a baking plant may be too large for efficient operation and that plants producing above 15 to 20 million pounds a year become less and less efficient as they grow in size.[1] A comprehensive study by William L. Crum [2] seems to indicate for industry after industry that the average rate of return increases with the size of the corporation, rapidly at first and then more and more slowly, until there is little or no additional gain after we reach the million-dollar class. Yet even here there seems to be no tendency for the rate of return to fall as the corporation grows even larger. Epstein concludes [3] that

> The larger manufacturing enterprises, in the main, do not earn profits at higher rates than the smaller ones. In 1928, for example, 1,421 small companies, with invested capital mostly under $250,000, earned an average net profit of 11.3 percent, while 1,970 larger companies, with capital of over $250,000, earned 10.6 percent. . . . The evidence seems fairly clear that, beyond a certain point, mere size is accompanied by no increased effectiveness in production, at least as reflected in terms of earning power.

Dewing made a study [4] of 35 large industrial combinations, selected at random from companies organized prior to 1904, and compared the records of the combinations with the prior records of their constituent companies. He found that these combinations earned only about half as much as their promoters had promised; perhaps even worse, that the earnings of the combinations were some 15 to 20 per cent less than the earnings that their constituent companies had made before combining. This decrease in earning was noticeable immediately, in the first year's operations after combination, and continued throughout the ensuing decade. For example, the companies brought together to form the United States Steel Corporation had earned, among them, about 108 million dollars annually before combination; the combination itself had average annual earnings of about 92 million in its first 10 years. The Twentieth Century Fund's study of the profitability of big business [5] contains the following conclusions:

> Large corporations that made profits made them at lower rates than small ones, while large corporations that lost money lost at lower rates than small

[1] THORP, WILLARD, "The Changing Structure of Industry, Recent Economic Changes," p. 189, McGraw-Hill Book Company, Inc., New York, 1929.

[2] CRUM, WM. L., "Corporate Size and Earning Power," Harvard University Press, 1939. Note especially the chart on p. 225.

[3] EPSTEIN, RALPH C., "Industrial Profits in the United States," pp. 45-46, National Bureau of Economic Research, New York, 1934.

[4] DEWING, A. S., A Statistical Test of the Success of Consolidations, *Quarterly Journal of Economics*, vol. XXXVI, pp. 84*ff*.

[5] BERNHEIM, ALFRED L., and others, "How Profitable Is Big Business?" Chap. 14, *passim*, Twentieth Century Fund, New York, 1937.

ones. Bigness seemed to act as a stabilization factor. . . . A study of banking profits shows clearly that the larger institutions had more favorable operating results than the smaller ones. . . . Larger corporations pay more liberal dividends. . . . Larger corporations have sounder financial set-ups.

As a final bit of evidence, we quote part of the conclusions of the National Industrial Conference Board's study of mergers in industry.[1] After an exhaustive statistical study the authors of this report conclude,

A study of the business history of a large number of consolidations, based upon published financial statements covering a period from ten to twenty-three years prior to the war, and upon the course of prices of their securities for a longer period, discloses that, by and large, these mergers did not prove exceptionally profitable. Some of them made high profits and grew in size and industrial importance correspondingly. But there were many that failed absolutely and passed out of existence. The majority did not achieve a conspicuous success as profit-makers. There is no means of knowing whether the financial record of these consolidations was better or worse than the financial record of business generally during the same period. . . . The long-run effects of industrial mergers upon technical efficiency have been much disputed, and there are *a priori* grounds for supposing that in this direction the influence of consolidation should not have been wholly bad or wholly good. Tested by the best available measure of actual operating efficiency, however, it appears that consolidations have in general achieved an appreciable superiority in production over independent producers in their respective fields. It should be emphasized, however, that the object of inquiry here is productive efficiency, solely, and not distributive, or marketing, efficiency. . . . But there remains the question whether fruits of this productive efficiency have been equitably shared with the consuming public. Society's paramount interest in industrial organization, *ultimately,* may be as stated above, but *immediately* its chief concern is with prices and with the steady operation of the industrial process. The effects of industrial consolidations upon these aspects of business operations are shown to have been, upon the whole, beneficial. . . . The presumption from the statistical evidence is strong, therefore, that the concentration of industry has been a factor tending to resist the general upward movement of prices. . . . Insofar as these conclusions are fair deductions from the statistical material here given, they justify the statement that in a very practical and substantial way industrial consolidations have been an economic benefit to the community.

In the light of the foregoing and much other conflicting testimony one must conclude that "circumstances alter cases"—that increasing the size of a business firm through combination may in some cases increase efficiency as a result of the forces listed in Sec. 25 • 10, but that it may on the other

[1] "Mergers in Industry," Chap. IX, *passim,* National Industrial Conference Board, New York, 1929.

hand reduce efficiency if the firm becomes cumbersome and unwieldy as the managers reach the limits of their supervisory ability. Although we must concede that monopolization of an industry *can* bring social savings as a result of increased efficiency, there is nevertheless no certainty that it *will* bring such savings.

25·12. Public Regulation of Monopoly. In the absence of regulation, a monopoly can be expected to charge whatever price promises to yield the greatest possible net income. As we saw in Chap. 16, this price will tend to be high if the demand for the monopolized product is inelastic[1] or if the industry is one of increasing cost; it will tend to be low, if the demand is very elastic or if the industry is one of decreasing cost. This means that we might expect high monopoly prices, in the absence of regulation, in the cases of necessities produced in the primary industries, and that even without regulation a monopolist would find it to his advantage to set low prices if he were producing luxury products in a secondary industry. Fortunately for consumers, it happens to be true that the industries most easily monopolized are those in which the monopolist finds it to his advantage to set low prices, and the industries in which monopoly has been hardest to achieve have been those in which monopoly price would tend to be high. Although this indicates that monopoly is to some extent self-regulating, nevertheless there has been a natural tendency for consumers to fear it and especially to distrust the monopolist's power to set prices. There has been a pretty general feeling, especially in a country where the spirit of individualism has run as high as it has in the United States, that we should not allow anyone to have monopoly power unless it can be reasonably well demonstrated that there will be some public advantage from it, and even then, that the public should protect itself against abuse of that power. This has led to the demand that monopolies be either publicly owned and operated, like the post office, or, if privately owned and operated, that they be subject to public regulation, as in the case of the railroads.

There has been a long and inconclusive dispute between the advocates of public ownership and operation and the advocates of private ownership and public regulation. Presumably the goal would be about the same in either case, and one would choose between them according as he thought the one or the other promised the greater efficiency of operation and the less incidental public inconvenience and disadvantage. It is generally admitted that monopoly control may have some advantages (such as those suggested in Sec. 25·10), and that as a result it is desirable to allow monopoly to exist under some conditions; but, where it exists, its powers must be so controlled that they will not be turned to public disadvantage.

[1] In fact, we learned in Chap. 16 that a monopolist always contracts his output until the demand is elastic, or until he is selling the smallest possible output.

25·13. Situation Preceding 1890. From very early times, the common law has frowned upon "restraint of trade." Agreements among businessmen not to compete with each other have been held to be against public policy and "in restraint of trade." Although such agreements were not enforceable at law and although the common law favored the preservation of competition, nevertheless there was no machinery set up to ensure that competition would be preserved. If two men agreed in restraint of trade, no one would stop them, although, if one of them refused to carry out his part of the bargain, the other had no standing in court when he attempted to sue for breach of contract. The court under such circumstances would declare the contract void, but it had no authority to punish the parties to the contract. Thus, the public received some negative protection of a kind, but there was no positive effort to bring monopoly under control until the start of the great combination movement after the Civil War. The fear of monopoly, coupled with the hardships of a great depression, brought about agitation for positive legislation. The railroads bore the brunt of the initial attack, and several of the states passed so-called "granger" legislation restricting them. This movement culminated with the passage of the Interstate Commerce Act of 1887. But the fear of monopoly was far more general than this, and many of the states, at first mainly in the Middle West, passed antitrust legislation. A dozen or more of them even inserted provisions in their state constitutions forbidding monopoly. But these scattered efforts of individual states met with relatively small success. In several of the states no real effort was ever made to enforce the laws and, even where there were vigorous attempts at enforcement, the individual states often felt that there was little that they could do effectively in the cases of great corporations whose activities reached to every part of the country. As a consequence, there was early support for the enactment of nation-wide legislation on the subject.

25·14. Sherman Act. In 1890, Congress passed almost unanimously[1] what has since come to be called the "Sherman Antitrust Act." This short act declares illegal "every contract, combination in the form of trust or otherwise, or conspiracy, in restraint of trade or commerce" in interstate or foreign trade.[2] It states further that anyone who shall "monopolize, or attempt to monopolize, or combine or conspire with any other person or persons, to monopolize" any part of interstate or foreign commerce is guilty of a criminal offense. Although the common law had held certain activities to be illegal but had set up no machinery to ferret out lawbreakers or to stop

[1] There was one dissenting vote in the Senate, and none in the House.

[2] This limitation is added, of course, since the Federal government is limited by the Constitution to jurisdiction over interstate and foreign commerce, and intrastate commerce is under the jurisdiction of the several states.

illegal acts, the Sherman Act took a positive course by enabling private parties to bring suit for damages or to secure relief through injunction. In addition, the Federal district attorneys were given responsibility, under direction of the Attorney General, for bringing action against violators. Persons violating the law are guilty of misdemeanors and may be fined not more than $5000 or imprisoned not more than a year, or both. This includes corporate officers or directors who are responsible for illegal acts of their corporations. In addition to such criminal suits, the Attorney General may bring civil suits leading to injunctions or to dissolution decrees. In the third place, prosecution may lead to a consent decree. In this latter case, the defendants, without having been tried and found guilty of violation of the law, agree voluntarily[1] to accept a court order governing their future activities. The consent decree has become increasingly more important as a method of handling cases under the Sherman Act.

The Sherman Act did little or nothing to define monopoly or to describe specific acts that were illegal. For these purposes it fell back on the long-established common law. The major advance legally under the Sherman Act was not, then, in making acts illegal that had been legal before, but in setting up machinery with respect to procedures and remedies. The responsibility for detecting and stopping violations was given to specific officials, and penalties for violation were established.

25 · 15. Results of the Sherman Act. Since the Sherman Act did not define the activities that were prohibited, merely stating that restraint of trade and monopoly were illegal, it became necessary for the courts to interpret the meanings of these terms. Such judicial interpretation naturally determined how the act would work in practice. During the first 10 years, little use was made of the new law, but President Theodore Roosevelt began an active and energetic campaign of "trust busting" soon after he assumed office in 1901. No attempt can be made here to summarize all the cases brought to trial, but one or two of the outstanding ones should be mentioned in passing.

Mr. J. P. Morgan dominated the Northern Pacific Railroad, and Mr. James J. Hill controlled the Great Northern. The two railroads served the same territory and were competing. Under the leadership of Hill and Morgan, the two roads engaged in a gigantic and dramatic financial struggle, culminating in the formation, in 1901, of the Northern Securities Company which controlled most of the stock of both roads. Competition between the two roads was eliminated, although the new system continued to compete with other systems. The government instituted a suit in equity to force the dissolution of the Northern Securities Company, and the case was carried

[1] Perhaps the word "voluntarily" should be put in quotation marks, since the firm often accepts the consent decree because it cannot afford to fight the case in court.

to the Supreme Court, where a dissolution decree of a lower court was upheld by a 5 to 4 decision. This was the first major victory in the Supreme Court for the government and encouraged the prosecution of other cases. It made it clear that the holding company could not be used as a device for the creation of monopoly.

The Standard Oil Company was incorporated in Ohio in 1870. Under the leadership of the elder Rockefeller, it soon established a loose federation with various of its competitors until over half of the refining industry was included in the Standard Alliance. In 1882, the relationship was strengthened by the formation of a trust, and in 1899 the Standard Oil Company of New Jersey was formed with a capitalization of 110 million dollars. This company came to own outright 65 other corporations and a controlling interest in 49 others. It completely dominated the field through the ownership of pipe lines, refining plants, etc. The brief submitted by the government in the case indicates that the original investment in the holding company had been in the neighborhood of 70 million dollars, and that the concern paid dividends on this investment over a period of 8 years in the amount of 326 million dollars. The company followed a practice of price discrimination, favoring markets in which it had to meet competition at the expense of other markets where it had the field to itself. It entered into agreements with the three principal railroads of the oil region under the terms of which the company received rebates on its own freight and on that of its competitors to the amount of 25 to 50 per cent. Finally, suit was brought against the company under the Sherman Act, and the case was carried to the Supreme Court. This court, in 1911, ordered the company to dissolve.

In 1890, the year of the passage of the Sherman Antitrust Act, Mr. J. B. Duke and others organized the American Tobacco Company, which at the start controlled roughly 90 per cent of the then small cigarette industry. The company quickly gained control of cigarette-making machinery and bought out a number of competing concerns. Weaker competitors were driven out by the use of discriminatory price policies or by their inability to use the most efficient cigarette-making machines. Dealers were given special discounts if they would agree to carry exclusively the brands of the monopoly. The company expanded into the field of plug tobacco, and then into the handling of snuff. In 1901, the monopoly was reorganized in the form of a holding company under the erroneous assumption that holding companies were immune from prosecution under the Sherman Act. The dissolution of the Northern Securities Company in 1904 dispelled this idea, and the companies involved were immediately merged into a new American Tobacco Company. In 1908, proceedings were started against it under the Sherman Act and, in 1911, the Supreme Court ordered its dissolution. Thus it was clear that neither the holding company nor outright merger protected a

monopolist or gave him a right to carry on business in ways that were otherwise illegal.

The Standard Oil case and the tobacco case were of additional importance because, in connection with them, the Supreme Court enunciated what has since come to be called the "rule of reason." Almost every economic situation involves a combination of competitive and monopolistic elements. It becomes important to determine just how far a business can go in the direction of trade restraint before it is guilty of monopolistic practices under the law. The notion of reasonableness played an important part from the first. As early as 1897, in the Trans-Missouri Freight Association case, it had been urged that Congress intended to prohibit only such restraints on trade as were unreasonable. The Supreme Court pointed out in this case, however, that the Sherman Act forbids *every* contract in restraint of trade, and not merely *unreasonable* contracts. In this case, the court refused to adopt the rule of reason. But almost a decade and a half later, the court restated its position in the oil and tobacco cases just described. Chief Justice White, writing the opinion, urged that one must go back to the common law to discover the meaning of the terms "restraint of trade" and "monopoly" and that only those agreements which restrained trade unreasonably were unlawful. Those agreements were unreasonable, in turn, which tended to reduce output, raise prices, or interfere with the freedom of businessmen. Under this interpretation monopoly and restraint of trade were synonymous.

Many years later Mr. Thurman Arnold, head of the government's Antitrust Division, summarized these early decisions and their later interpretations,[1] stating that the economic substance of the rule of reason is,

> You may grow as big as you can provided that you can justify the extent of your organized power by showing that it contributes to the efficiency of mass production and distribution. We will protect you against organized groups of small business which attempt to prevent you from giving cheaper goods to the public. We will, on the other hand, attack you if you seek to maintain your system of distribution by using your organized power to prevent experimental developments by others either in production or distribution or in price policies. Size itself is not an evil, but it does give power to those who control it. That power must be constantly watched by an adequate enforcement organization to see that it does not destroy a free market.

Commenting on this rule of reason, Mr. Arnold observes, "Without that rule of reason the Sherman Act would be unworkable in a machine age, because every combination between two men in business is in some measure a restraint of trade."

[1] ARNOLD, THURMAN W., "The Bottlenecks of Business," p. 125, Reynal & Hitchcock, Inc., New York, 1940.

25 · 16. Legislation of 1914. The Sherman Antitrust Act was criticized strongly from two contradictory points of view. On the one hand, some businessmen objected that the law was too vague and that they could not tell what sorts of activities were illegal. They asked for the establishment of some authority that might advise them as to the legality of proposed ventures. Some of them obviously hoped for more lenient interpretations of the act in favor of big business. On the other hand, many critics objected that there were not teeth enough in the Antitrust Act as it stood; that the interpretations by the courts, and especially the rule of reason, had made it possible for business firms to combine with impunity; that the Sherman Antitrust Act, as interpreted by the courts, gave little or no more control over monopoly than that offered by the common law. It seems to be generally conceded today that the language of the original act was so broad and general that judicial interpretation was demanded; yet at the time many opponents of monopoly felt that the courts were stepping out of their proper sphere and, under the guise of determining reasonableness, were actually exercising legislative functions.

Woodrow Wilson was elected to the Presidency in 1912 on a platform calling for greater control over monopoly. During his campaign[1] he had said:

> But what we are afraid of, as we have said in the Sherman Act, is such use of corporations as will be in restraint of trade; that is, such use as will establish monopoly. Very well, we have got to know what particular things do establish monopoly; because if we merely say "this concern is so big that it looks to us like monopoly" that isn't satisfactory. I am not afraid of any corporation, no matter how big. I am afraid of any corporation, however small, that is bad, that is rotten to the core, whose practices and actions are in restraint of trade. So that the thing we are after is not reckoning size in measuring capacity for damage, but measuring and comprehending the exact damage done. . . . Item by item we can put into our statutes what constitutes restraint of trade, not leaving it to courts for generalizations which may fit some cases and not others.

Soon after the election, the program of clarifying the provisions of the Sherman Act was begun.

In 1914, Congress passed two important acts; the Federal Trade Commission Act and the Clayton Act. The former set up the Federal Trade Commission, giving it the power to investigate business practices and to report its findings. In addition to its work in investigating and suppressing monoply, the Commission was to prevent "unfair competition." The phrase was not defined, although it was obviously intended to cover not only such patently unfair methods as fraud and misrepresentation, but also any prac-

[1] In an address at Nashville, Tenn. on Feb. 24, 1912.

tices that might serve to destroy competition or to encourage the development of monopoly. Finally, the Commission was given broad advisory powers, advising both the courts and Congress. The Clayton Act forbade price discrimination between purchasers of a commodity, if the result were to lessen competition substantially or to create monopoly, and forbade tying contracts and exclusive dealing contracts under which wholesalers and other dealers were forced to handle the goods of one seller exclusively or to carry other less desirable goods furnished by a given seller in return for the right to carry a particularly desirable item. The Clayton Act also forbade the formation of holding companies if their effect was to reduce competition and set limitations on interlocking directorates. Any concern that believed itself injured by "unfair" practices of competitors might complain to the Federal Trade Commission. If the Commission agreed that the practice was indeed unfair and fell within its jurisdiction (that is, that it involved interstate commerce and the public interest), it served a complaint on the offender, who appeared at a hearing. If, after hearing, the Commission still felt that the practice was unfair, it issued a "cease and desist order," which, as the name implies, was an order directing the offender to cease and desist from the practice in question. Such orders are enforceable by the courts, to which the offender may also appeal if the order appears to him onerous. The courts thus have the final jurisdiction in all such cases. Under an amendment of 1938, complaints may also be lodged by consumers who feel that they are injured by unfair practices, and the powers of the Commission over false advertising were extended.

The provisions of the Clayton Act prohibiting price discrimination under certain circumstances were strengthened by the passage of the Robinson-Patman Act in 1936. The original act had prohibited price discrimination that substantially lessened competition or tended to create a monopoly in any line of commerce. The 1936 act added, "or where the effect is to injure, destroy, or prevent competition with any person who either grants or knowingly receives the benefit of such discrimination, or with the customers of either of them." The intent seems to have been to make it unnecessary to demonstrate that a given discrimination in prices lessened competition throughout an entire industry or line of commerce, but merely that it reduced competition among people directly involved. Price discrimination may continue, but it must reflect actual differences in cost of manufacture or sale to the different customers.

The provisions of the Clayton Act are enforced by the Board of Governors of the Federal Reserve System insofar as they apply to banks and trust companies, and by the Interstate Commerce Commission insofar as they apply to common carriers. In all other cases they are enforced by the Federal Trade Commission.

25 · 17. Webb-Pomerene Act. The Sherman Act and its successors forbade monopolistic practices in interstate and foreign commerce. American exporters complained that, under the provisions of these laws, they were forbidden to use in foreign commerce many practices common to their foreign competitors. It seemed unfair to forbid to an American exporter selling in Brazil practices that were legal for a German or an Englishman selling in Brazil in competition with him. As a result, in 1918, Congress passed the Webb-Pomerene Act, which exempted from the provisions of the antitrust acts combinations which were engaged solely in export trade as long as the combination did not lessen competition within the United States or use unfair practices against American competitors. General supervision over such export combinations was given to the Federal Trade Commission, and such combinations are forced to register with the Commission. In practice, little use has been made of the provisions of this law, and the number of combinations registered has never been greater than 50 or 60.

25 · 18. Results of Antitrust Legislation. For some years after the original enactment of the Sherman Act little effort was made to enforce it. What efforts there have been have been sporadic and disconnected. Yet the government has won a number of very important cases, and there can be little question but that the threats embodied in the legislation have done a good deal to shape the course of the development of American business. On the other hand, it was pointed out in the early sections of this chapter that monopoly has not disappeared and even seems in some respects stronger as time goes on. The development may have been retarded but it certainly has not been stopped. It appears that there are forces leading toward monopoly that cannot be offset merely by the passage of legislation. The corporate form of enterprise is, of course, a mechanism well adapted to the organization of gigantic business undertakings, yet its advantages are so great that no one suggests scrapping it *in toto*. The protective tariff has been thought by many people to foster monopoly.[1] The great power inherent in monopoly itself is doubtless an attraction to energetic and ambitious businessmen. But, paramount among the forces leading to the continued growth of monopoly is the economy of large-scale production in certain lines of endeavor. If the big concern can operate more effectively than the small one, businessmen will find some way to take advantage of the greater efficiency; and it will be to society's advantage to let them do it if the powers incident to the monopoly can be so controlled and directed that one gets the efficiency without social harm. The patent laws of the United States have been stimulants to the development of monopoly in ways never intended when they were originally passed. For these and many other reasons, the antitrust legislation has been at best partly successful. Moreover it has often been offset by

[1] See Chap. 23.

contradictory laws that actually encourage monopoly, such as much of the recent farm and labor legislation, and many of the provisions of the National Industrial Recovery Act.

25·19. The Present Situation. We find ourselves in a society which attempts to prohibit monopoly, yet which contains a considerable degree of it. There seems to be little practical likelihood that monopoly could be entirely eradicated even if such a result were considered desirable. It is probably fair to say that some solution to this problem must be found if we are to maintain the type of economic system that we have known heretofore. In the words of three thoughtful American economists,[1] "To assert that the survival of the system of private capitalism—perhaps even of political democracy itself—requires the control of monopoly and such regulation of trade and industry as will maintain competitive conditions in a reasonably large portion of industry is not an overstatement of the case." Such a result cannot be easily achieved. Many people have given up hope of achieving it at all and have turned their energies to building up among farmers and laborers and others the same sort of monopoly power that seems to them inescapable in other segments of the economy. Others have preferred governmental ownership and operation of industry. Those who continue to favor a system of private competitive enterprise must understand that the continuance of such a system will require much skill and originality in improvising new methods of organization and control which will combine the efficiency of large business with the dignity, integrity, and power of individuals, as workers, stockholders, and consumers. It will require fundamental changes in the powers of the corporation and in intercorporate relations, and in the patent laws. Above all, it will require patience and understanding on the part of legislative and regulatory bodies, and a compelling desire on their part to reach a workable solution.

Questions

1. Give examples of product differentiation.
2. Mention any actual cases in which sellers insist that buyers purchase several units at a time if they are to buy at all. For what reasons do you think that it is done?
3. If a man is at one time a director in two or more companies, we call him an interlocking director. Would it be possible to have other sorts of interlocking, such as interlocking presidents, interlocking treasurers, etc.? Would such interlocking be open to the kinds of objections that are raised against interlocking directors?
4. What is the difference between a pool and a gentleman's agreement?

[1] FAIRCHILD, FRED ROGERS, EDGAR S. FURNISS, and NORMAN S. BUCK, "Elementary Economics," vol II, p. 282, The Macmillan Company, New York, 1939.

5. Banks and insurance companies often invest at least part of their capital in the securities of various corporations. If an insurance company owns shares of the stock of 100 different corporations, does this make the insurance company a holding company? Discuss.

6. In what ways is a trust like, and in what ways unlike, a holding company?

7. Explain how pyramiding works.

8. "The development of the corporation, and especially the use of the devices of the trust and the holding company in conjunction with the corporation, have brought about a separation between ownership and control in industry." Explain.

9. Give several actual examples of horizontal and of vertical combination.

10. The economist often argues that, if there is complete competition in the market, all producers will be forced to sell at the same price, since none will be able to exact a higher price and none will be forced to accept a lower price than his competitors. Yet we say that there is evidence of monopoly if a number of firms submit identical bids on a given contract. Are the two ideas contradictory? If so, which is correct?

11. Give two or three examples of cases in which a large part of the production of a commodity is concentrated in the hands of one or two firms. Give two or three other examples of cases in which no firm in an industry controls the production of any significantly large proportion of the output.

12. List some commodities that change in price often. Some that change in price seldom. Do you think that there is a greater degree of monopoly in the latter cases than in the former?

13. How big would a dairy farm have to be before you would call it big? (Number of men employed, value of output, amount of investment.) How big would a department store have to be before you would call it big? How big would an automobile manufacturing establishment have to be before you would call it a big one? Are there any common measures of size that we can sensibly apply to businesses of all types?

14. Suggest reasons for the fact that the corporation has remained relatively unimportant in agriculture (see Fig. 25 • 3).

15. Figure 25 • 4 indicates that corporations control a much larger proportion of the employees and of the value of output than they do of the number of establishments. What does this mean as to the relative size of the incorporated and the unincorporated establishments? Explain.

16. Why can a businessman afford to give a quantity discount? How can a railroad afford to give lower freight rates on carlot shipments than on smaller ones?

17. "Combination makes it possible to eliminate cross-shipping, but if there were complete competition there would be no cross-shipping anyway." Comment.

18. Give examples of by-products which are reclaimed by large businesses but which small businesses cannot afford to reclaim. Recalling what you learned of joint costs in Chap. 17, name the advantages that the public gets from the utilization of by-products.

19. Interpret in a paragraph or two your conclusions from Fig. 25 • 6. How would the appearance of this chart have been altered if the largest concerns were markedly inefficient; if, after reaching an optimum middle size, further increases in size had been really disadvantageous?

20. In Fig. 25 • 7 the difference between the most efficient and the least efficient groups of bakeries was only a fraction of a cent. Do you think that such a small difference is worth bothering with? Why or why not?

21. Explain what is meant by the statement that, in the light of the effects of elasticity of demand and trends of expense, "monopoly is to some extent self-regulating."

22. What is a consent decree?

23. Write out in two or three sentences the essential parts of the rule of reason.

26. PUBLIC UTILITIES AND TRANSPORTATION

26·1. What Is a Public Utility? In the last analysis, whether a given business is a public utility or not is determined by the courts; it is a legal question. Although we are all familiar with the phrase and understand in a general way that the *public utilities* are such businesses as street railways, gas companies, electric-light and power companies, bus lines, and steamboat lines, nevertheless there is no single clear-cut and accepted definition by which we can recognize a public utility when we meet one—no rule that we can apply with confidence that it will separate all those businesses which are public utilities from those which are not. In case there is a question about a particular business, the matter may be taken to court, and the court will decide whether the business is or is not a public utility. Different legal jurisdictions seem to apply somewhat different tests, and the concept changes from time to time within any given jurisdiction. In the absence of any reliable test that we can apply a priori, let us examine the businesses that have been held to be public utilities to see what common characteristics they have. Possibly in this way we can discover at least a rough empirical test.

26·2. The "Public Interest." The lawyers define public utilities as those which are "vested with a public interest"; those in which the general public has a peculiar and direct interest. Yet a definition in such terms does more to obscure than to elucidate. Every business is more or less "vested with a public interest," and in many cases the ones in which the public is most vitally interested are not classed as public utilities. For example, the public is obviously tremendously interested in the industries that produce food: wheat raising, potato growing, animal husbandry, etc. Yet these basic agricultural industries are not classed as public utilities despite their fundamental importance. If the public interest is the test, it must be some special and peculiar kind of public interest, and not mere public dependence on the product of the industry.

26·3. Monopolistic Character of Public Utilities. When we examine the businesses actually classed as public utilities, we find that they usually exhibit

a considerable degree of monopoly power. This character depends, in turn, on a number of things, most of which we have already studied. For example, most of these businesses are very heavily capitalized and have very heavy fixed costs. A railroad line, a telephone company, a street railway, or a gas company, has heavy investments in fixed capital in the form of roadbeds, pipe lines, poles and lines, rights of way, etc., and in each case the variable expense is small as compared with the fixed expense. For the railroads of the United States, for example, it has been estimated that "approximately one-half of the expenses are constant and one-half variable."[1] We have already learned that businessmen tend, when hard-pressed by competition, to neglect their fixed costs and that those businesses with large fixed costs tend, as a result, to become monopolistic.

Even if there were not a substantial element of natural monopoly in the public-utility enterprises, we would probably be forced in many cases to insist that they operate monopolistically because such operation would be to the public advantage. Competition between a large number of small concerns would be inconvenient for the public. Imagine, for example, the problems that would arise if 20 different street railway companies were to run their rails down Willowbrook Road, or if 40 different gas companies were to dig up Hillside Circle to lay their mains. Electric power lines and telephone wires present problems enough when there is but one company involved but, if Main Street were to carry the poles and wires of 30 competing concerns, there would be no room left for other traffic. It must be understood by the student that one of the reasons for the existence of some degree of monopoly in the public-utility industries is that monopoly in these industries is socially convenient and that competition would be inconvenient.

One additional peculiarity of the competitive situation in the public-utilities field is sometimes mentioned. In most businesses, if one's competitors cut prices and take away trade in one market, one can ship his goods to another market to escape the problem. If he finds that the price of potatoes is ruinously low in the New York market, he can ship his crop to Boston or Philadelphia. But the typical public utility is irrevocably tied down to its own territory. The gas company can sell its product only in the district through which its gas mains extend; the street railway company must get all its business in the territory reached by its rails. If competition were to spring up within this district, the manager of the public utility would

[1] JONES, ELIOT, "Principles of Railway Transportation," p. 77, The Macmillan Company, New York, 1925. The student will realize, if he understands the meanings of the terms "fixed" and "variable" costs as they were explained in Chaps. 14 and 17, that we cannot say that the fixed costs make up any particular percentage of the total costs, since by definition, as the output of the industry increases, the percentage of the total cost that falls in the fixed categories becomes smaller and smaller. Jones evidently means that about half the costs are fixed when the volume of business is at its usual level.

find that he could not rely on other territories. He could not, as other businessmen do, turn to alternative markets.

Partly because of the nature of the public-utility businesses themselves and partly because competition in these fields would be inconvenient for the general public, it is common to give businesses in these fields *exclusive franchises,* permitting them, and them alone, to carry on the given type of business within a given territory. Thus, one company is often given the exclusive right to operate streetcars within a city, and another company the exclusive right to sell illuminating gas. Bus companies are given the exclusive right to operate busses over particular streets, and power companies to string wires and sell current in a given city or other district. In this way we recognize, formalize, and legalize the monopoly that already existed for the reasons mentioned.

26 · 4. Right of Eminent Domain. The right of eminent domain is the right that the government has to take private property for the public benefit in return for a just compensation. It is often necessary for the state to exercise this right in the behalf of business enterprises. For example, if a corporation were laying out transmission lines for electric power and if someone along the route refused to sell land to the corporation, the entire project would fail. It might well be that a vast project involving hundreds of miles of line and millions of dollars of investment would fail because someone refused to sell a few acres of land strategically located. Since, under such circumstances, the public would be unable to get the electric service, the public advantage of forcing the reluctant landowner to sell his land for a reasonable sum is obvious. The same sort of problem arises in building railroad lines, laying gas mains, setting up telephone or telegraph lines, etc. Most public utilities are of such a nature that the business could not be established in the first place without this unusual form of public assistance which we call "eminent domain." Most other businesses do not require it. If a corporation wishes to open a garage, a meat market, or a factory and is unable to buy the site that it prefers, it can still buy reasonably good substitutes. The entire business does not fail because of the inability to get a particular site. On the other hand, when a railroad right of way is laid out, refusal to sell any piece of land along it makes the enterprise impossible.

26 · 5. Nature of the Demand for the Products of Public Utilities. In Sec. 26 · 2 we discussed the lawyers' statement that public utilities are "vested with a public interest." In one sense this may be true, even though in this sense it is not peculiar to public utilities. Perhaps the thing that the lawyers have in mind in making the statement is that the demand for the services sold by public utilities tends to be rather inelastic, so that the public would be harmed in an unusual degree by the raising of prices. To be sure, inelastic

demand is not peculiar to the products of public utilities; but we have learned that the combination of inelastic demand and monopoly power is particularly dangerous.[1] It is a combination that naturally leads to very great restriction of output and to very high prices. Many industries whose products are characterized by inelastic demand are intensely competitive industries, as, for example, is true in the production of the great staple agricultural foodstuffs. Modern society depends in considerable degree on the products (electricity, gas, transportation, etc.) of the public utilities, and the combination of inelastic demand with a decided element of monopoly gives to these enterprises a special sort of public interest—an interest based on fear of restricted output and high prices in the absence of regulation. In this sense, at least, it is reasonable to say that the public utilities are, to an unusual degree, "vested with a public interest." For this reason, among others, these industries are subjected to unusual regulation by the government.

26 · 6. Changes in the Public-utility Concept. From very early times, men have segregated certain industries and subjected them to special regulation and supervision. We are told[1] that the early English common law provided for such regulation of barbers, surgeons, innkeepers, gristmillers, wharfingers, and carriers. With the passage of time and the changes in the economic situation, various ones of these businesses lost their peculiar status of public interest and were replaced by others, until today, as we have said, the phrase "public utility" is ordinarily used to include such businesses as those supplying water, gas, electricity, telephone and telegraph service, and transportation on streetcars, busses, railroads, and airplanes. Although no one of these businesses incorporates all the characteristics thought of as distinguishing public utilities (some element of natural monopoly, large fixed costs, public inconvenience of competition, use of eminent domain, inelastic demand for products, etc.), nevertheless they do have a good deal in common.

Just as the concept of public utility has changed in the past, so it is undergoing slow changes today. Because such utilities are subject to an unusual amount of public control, there is a tendency for those whose political and social philosophy favors public regulation to argue that more and more businesses are "vested with a public interest"; while those who look with disfavor on public "interference" with industry tend to lean toward a narrow interpretation of the phrase. Businessmen themselves take curiously contradictory views of the subject at different periods in the business cycle. When business is good and profits are large, businessmen generally tend to object to government regulation and to favor a policy of

[1] See Chap. 16.

[1] THOMPSON, C. W. and W. R. SMITH, "Public Utility Economics," p. 3, McGraw-Hill Book Company, Inc., New York, 1941.

laissez faire. When, on the other hand, we are in the depths of depression and profit margins dwindle and disappear, businessmen are filled with apprehension and, like a child in the dark, yearn for the mother's protecting hand. At this stage of the business cycle, we find them clamoring for public regulation in their industries, arguing that the milk industry, the ice industry, or whatever industry they represent is "vested with a public interest" and that it should be regulated. Strangely enough, the regulation desired in such cases is usually to reduce competition, since these industries do not in most cases share the peculiarities common to public utilities as listed in the earlier sections of this chapter.

The student who is interested in learning more of the present status of the public-utility concept will have to go to the original legal decisions, where, in both majority and dissenting opinions, he will find the arguments on both sides given in detail. Starting, perhaps, with the case of *Munn v. Illinois,*[1] he will continue with the German Alliance Case,[2] the Wolff Packing Company Case,[3] the New York Ticket Agency Case,[4] the Oklahoma Ice Case,[5] and the case of *Nebbia v. New York.*[6] Since we are not here interested, except incidentally, in the purely legal aspects of the problem, we can content ourselves with the observation that the courts of the United States have long held that the two major characteristics of public utilities are economic necessity (inelastic demand) and monopoly, but that more recent decisions have made it difficult to discover any consistent pattern on which one can rely. In the last analysis, the decision as to whether a given business is or is not a public utility, like the decision as to whether an automobilist was or was not driving recklessly, is based on human judgment and depends on the decision of the courts.

26·7. Regulation of Public Utilities. The businesses that we have been describing as public utilities cannot operate unless they are given such unusual privileges and benefits as monopoly power, exclusive franchises, and the right of eminent domain. They are given rights that are withheld from most businesses. As a result, they are required to submit to unusual regulations. It is common observation that great power brings with it great duties and responsibilities, and it is surely reasonable, if society is to give businesses unusual powers and privileges, for society to insist that these powers be not used to the social disadvantage. As a consequence, the government has

[1] 94 U. S. 113 (1877).
[2] 233 U. S. 389 (1914).
[3] 262 U. S. 522 (1923).
[4] 273 U. S. 418 (1927).
[5] 285 U. S. 262 (1932).
[6] 291 U. S. 502, 2 P. U. R. (N. S.) 337, 352 (1934).

Brief summaries of these and of a few other oustanding cases are found in Thompson and Smith, *op. cit.,* pp. 58-64.

always been allowed to go further in regulating public utilities than in regulating other businesses.

The regulation of public utilities takes many forms, but by far the most important is the regulation of prices charged for the goods or services produced. Price regulation is natural whenever society grants monopoly power; especially is it necessary if the demand for the monopolized product is inelastic. We find that, whenever the government sets up such regulatory bodies as public-utilities commissions or commerce commissions, one of their most important powers and duties is that of regulating the prices charged by the concerns subject to their jurisdiction. These prices, for obvious reasons, are seldom or never fixed by statute. Instead, the statutes provide for the establishment of a regulatory body which is given price-setting powers. Yet, the regulatory body is not given, and cannot constitutionally be given, unlimited powers in the field of price setting. It is important for the student to understand that public-utilities commissions do not set prices at random, but in accordance with fairly well-established rules.

In the first place, lawyers would say that the prices must be "reasonable"; yet by itself this restriction tells us little unless we know what prices are reasonable and what are unreasonable. The lawyers help us by going another step and stating that such a price is one that yields a reasonable return. Thus, if the price is set so low that the public utility is forced to operate at a loss, the price is unreasonable. Sensible public regulation does not consist of setting prices as low as possible, since if the price is too low no one can afford to operate the business, and society will not get the desired electricity, gas, or transportation service. The superficial student is apt to overlook the fact that, even when judged from the social standpoint, prices may be set too low as well as too high.

If a reasonable price is one that yields a reasonable return, we are justified in asking how large a return is reasonable? Should one aim for a return of 6, 10, or 2 per cent? Evidently, the answer is to some extent arbitrary, although to some extent it must depend on the returns being currently earned in other lines of business. If businessmen in general are making 12 per cent on their investments and if we regulate the prices of public utilities in such a way that funds invested in these enterprises earn but 8 per cent, investors will be loath to put their funds into public utilities, and as the capital in these industries wears out it will not be replaced. The services will slowly deteriorate and the public will suffer. If, on the other hand, businessmen in general are making but 3 per cent on their investments, there is no justification in allowing public utilities to charge prices that will yield a return of 6 per cent. A reasonable return for a public utility is, thus, a return that is substantially equivalent to that received by investors in similar but unregulated businesses. In practice, it is impossible to make any precise

adjustment, and actual rate setting is approximate and experimental. As one authority[1] has expressed it,

> There is no divine revelation to the commission and courts as to what fair and reasonable is. There is no abstract standard of fairness and reasonableness, and I submit that what the practice has been is in an experimental way to fix those rates at about what they would be if there were normal competition, when in fact there is not normal competition. . . . The public looks to the Commission to fix rates that are commensurate with ordinary competitive conditions instead of with monopolistic conditions.

26·8. Bases for Rate Making. If we could dodge all the difficulties involved in the foregoing, and if we could decide that a fair price is one that will yield a return of 6 per cent, we would still have to ask, "Six per cent of what?" What are we to use as the basis against which we are to compute the size of the return? To the uninitiated, the question sounds easy; in practice, it turns out to be a very difficult one, to which no entirely satisfactory answer has been (or seems likely to be) found. A "fair price" is one that yields a "fair return" on a "fair value" of the property of the utility. But what is a "fair" value of a power company, railroad, or a bus line?

Some people have suggested that we take as the value the actual *original investment* in the company, but this gives an unfair advantage to the concern which had high original investment but which has not kept its property in a good state of repair. We would allow it to charge as much as a company with the same original investment which had kept its facilities in good condition and which, as a consequence, was rendering better public service. Moreover, if the original construction were inefficient and costly or if the venture were foolishly undertaken at a time in the business cycle when costs were high, we would allow the concern to charge high rates to cover these high costs, while the concern that used inexpensive and efficient methods and was judicious in selecting the best time to build would get no benefit from its savings. A method that bases charges on the actual costs makes it possible for inefficient concerns to escape the penalty of their inefficiency and makes it impossible for the efficient concern to reap the reward of its efficiency.

To get around this difficulty, it has been suggested that prices be set on the basis of the *cost of reproduction* rather than on the actual costs of production. We would ask, not how much did the utility invest, but how much would it have to invest today to build the same sort of facilities at present prices. This would put all concerns on the same price basis. One can still object to the cost of reproduction as a basis for price setting, however, since it takes no account of changes in tastes and desires of consumers. When the automobile replaces the trolley car as a means of short-haul

[1] FETTER, FRANK A., *Testimony in the Matter of . . . Rates . . . of the Wisconsin Telephone Company*, 2-U-35, pp. 1167, 1168, 1932.

passenger transportation and when prudent investors would no longer build a trolley system at all if the choice could still be made, there seems little justification for allowing an existing trolley system to charge rates that would earn a fair return on even the present cost of reproducing its plant. If sensible men would not reproduce the plant, sensible men should not set rates that would make the reproduction profitable. Therefore, it is often argued that the proper basis for price setting is neither actual original investment nor the cost of reproduction, but the actual *present value of the property*. Unfortunately, as any accountant or appraiser can testify from his own experience, it is extremely difficult to ascertain the value of any piece of property, and the results are always tentative, approximate, and open to question.

26·9. Capitalization of Net Income. To be sure, the economist does have a very useful method for estimating the approximate actual present value of any income-producing property. This method, called the "capitalization of net income," is one that we shall have occasion to use again and, since it is a method for estimating present values, it is perhaps natural that misguided persons should have suggested that it be used in ascertaining values for price-fixing purposes. It will be necessary to describe the method before we can appraise its adaptability for this particular use.

Suppose that you own a share of stock in a corporation, that the dividends on this share of stock amount to $6 per year every year, and that you expect this income to continue. How much is the share of stock worth? The answer is that its value depends in part on its earning power (here taken as $6 per year) and in part on the earnings that can be made in other investments—on the typical or going rate of interest on investments. If investments in general are expected to yield 6 per cent returns, then this share of stock is worth about $100; if investments in general are expected to yield but 3 per cent returns, this share of stock is worth about $200. We find these values by the process of capitalization, using the following formula:[1]

$$\text{Value} = \frac{100\ (\text{annual net income})}{\text{interest rate}}$$

The results obtained by using the formula are, to be sure, only approximate; but the surprising thing is that in actual practice the estimates made by using it turn out to be fairly good approximations to actual values, for reasons

[1] The formula is usually given in the form of the equation. The figure 100 in the numerator is inserted because the interest rate is taken as a whole number rather than a decimal. To be sure, an interest rate of 6 per cent is a rate of 6 one-hundredths, or 0.06. If we were to use the interest rate in this correct decimal form, we would omit the figure 100 in the numerator, but if we wish to use the figure 6 to represent 6 per cent in the denominator we must offset it by using the multiplier as given,

that will become apparent shortly. In the meantime, let us illustrate the use of the formula. Assuming that the rate of interest expected on investments is 6 per cent and that our share of stock is earning $6 per year, we substitute in our formula to find

$$\text{Value} = \frac{100\ (\$6)}{6} = \$100$$

But, if the "going rate of interest" were 3 per cent, our formula would yield the following results:

$$\text{Value} = \frac{100\ (\$6)}{3} = \$200$$

This capitalization formula is used wherever we wish to estimate the value of income-producing property, and it is important to understand that the income used is the *annual net income.* Suppose, for example, that I own a house in which someone else lives and that they pay me a rent of $60 per month. What is the approximate value of the house? My first step is to convert the income to an annual basis. If I receive $60 per month, I receive $720 per year. But this is annual gross income, and I want annual net income. If my expenses (insurance, taxes, etc.) amount to $300 per year, my annual net income is $420 per year. Now, if the general rate of interest on investments is 5 per cent, I can find the approximate value of the house by capitalizing the annual net income at 5 per cent, thus,

$$\text{Value} = \frac{100\ (\$420)}{5} = \$8400$$

I have now estimated that the house is worth approximately $8400.

How am I justified in assuming that a house that yields an annual net income of $420 will be worth about $8400 when investments in general are paying 5 per cent? The answer is that $420 is 5 per cent of $8400, so that an investment of $8400 can be expected to yield about $420 per year income. As the house yields $420 per year income, it is bringing in just about what an investment of $8400 would bring. If the house could be bought for less than $8400—for example, for $7000—it would be an unusually good investment, since an investor could make as much ($420 per year) by investing $7000 in the house as he could make by investing $8400 elsewhere. Investors would compete for such an attractive investment opportunity and force its price up. If, on the other hand, I tried to get substantially more than $8400 for the house—for example, $10,000—investors would hesitate to buy, because they would be getting only as much from their $10,000 investment as they get by investing $8400 in other ways. Therefore, I would find it difficult to get much above $8400 for the house, and I would probably not have to accept much less than that amount for it.

All economic laws are, as we have emphasized repeatedly, statements of general tendency. They are at best approximations. This is true of the state-

ment that *the value of income-producing property depends on the amount of income that it earns as compared with the amount that could be earned in alternative investments.* Although this statement is approximate, it is considerably more reliable than most statements in the field of economics and has much more the character of a verifiable quantitative rule. In practice, values of property seldom drift far from those which would be estimated by our formula, and we shall find the formula a useful one at later stages of our work.

26 · 10. Capitalization for Price-setting Purposes. We must not lose sight of our present goal. We have been trying to ascertain some method for finding a reasonable basis for setting prices of products sold by public utilities. We have been forced to acknowledge that several suggested methods are open to serious criticism, and we find that it has been suggested that we allow utilities to set rates that will yield a fair return on the actual present value, this value being found, at least approximately, by capitalizing the net income of the public-utility corporation.

Let us overlook for the moment the fact that net income is itself difficult to ascertain. Any accountant, save the one who is filled with the assurance of his first and only course in accounting, realizes that measures of net income are approximate and are based on more or less questionable and arbitrary methods of appraising and assigning costs. Although this book has not gone into matters of accounting except briefly and superficially, we have seen something of the difficulties involved in situations of joint costs, or in those cases where fixed costs play an important part. We have also noted the difference between marginal costs and average costs, and we would have to decide what kinds of costs we were to deduct from total income in order to find our annual net income. Fortunately, we need not argue these questions here, important as they would be in practice; for the method of capitalizing net income to determine value is useless for our present purposes on quite other grounds. Useful though the method is for other purposes, we should not use it here. For, as soon as we stop to reflect, we realize that the amount of the company's net earnings will depend on what prices they charge. If they charge too much or too little, their net income will presumably be small, while if they set their price at some "happy medium," they may make a large net income. How can we use the value as a basis for setting the price when the price itself will determine the value? We shall have fallen into a fatal circuity. The error of our thinking, when we apply this method to rate making cases, can perhaps best be understood by writing out explicitly the steps in the "reasoning" involved. We might say to ourselves:

1. If we let this company charge a little more, they will get a higher net income.

2. If they get a higher net income, the value of their property will be greater.

3. If the value of their property is greater, it is reasonable for them to earn a larger income.

4. Therefore, if we let the company charge a little more, it will be reasonable to let them charge a little more!

And we might just as logically have "proved" that, if we reduced their charges and made them earn less, it would be reasonable for them to earn less.

26·11. "Proper" Basis for Rate Making. It is quite reasonable for the student, having learned that this method or that method of computing prices and rates is not workable, to ask, "What, then, is the right method? What is the correct dope?" Unfortunately, the answer is that no one knows. Like so many problems of real life, this is a problem on which "there is much to be said on both sides." Courses in economics would be easier for students, just as life would be simpler for everyone, if all problems had easy and obvious answers. Here the important thing for the student to realize is that we do not dare to give to any corporation those unusual powers which we give to public utilities without somehow protecting ourselves against the misuse of those powers; yet we cannot agree that there is any one particular way to regulate public utilities that is demonstrably the "right" way. We are forced to fall back on trial and error, and to admit with Professor Fetter that there is no divine revelation to tell us what is fair and reasonable. Behind all the theoretical hocus-pocus that regulatory bodies use to justify their results we should see the practical experiment of trial and error as public-utility commissioners strive to fix rates "at about what they would be if there were normal competition, when in fact there is not normal competition."

26·12. Nature of Transportation. Let us turn from a general and abstract consideration of the problems of public utilities and illustrate them from one of the major public-utility fields, that of transportation. We recall[1] that production consists in adding utility to things. Most productive activity is concerned with form and time utilities, but the transportation industries are concerned with adding place utilities. We might usefully make a distinction between two major classes of transportation: of people and of goods. Early transportation was largely the former. Resources were not evenly distributed, and men moved from place to place to reach the things that they wanted. Since the mountain did not come to Mohammed, Mohammed went to the mountain. People were nomadic, living primarily by hunting and fishing, and moving themselves to the places where the desired goods were to be found.

Since, however, men want many goods and since there is no one place where they are all found, the method of moving men to their goods can never be a satisfactory one. Means of transportation were slowly developed by which goods from many places could be brought together at the points

[1] See Sec. 10·2.

where men wanted them. There was a tendency for men to settle down, and to transport goods rather than people. Through long centuries men developed various kinds of animal power for transportation: the dog, the horse, and the llama. The force of the winds was harnessed behind the sails of vessels in the development of water transportation, taking advantage of the cheapness of a roadbed that was a free gift of nature. The development of the wheel and the slow transformation of footpaths into cart tracks, of cart tracks into highways, and of highways into the modern paved arteries of motor traffic marked ever new chapters in progress. Thus we condense the achievements of centuries into a paragraph or even a sentence.

Both within our own country and abroad there was tremendous interest and investment in the building of canals in the early part of the nineteenth century, and with the rapid development of the steamboat after about 1840 both canal and river traffic multiplied again and again. Also, about 1840, the steam railroads began to spread their net across the land, first as feeders to the canals, and later as independent enterprises. They reached Chicago in 1855, and in 1869 the first transcontinental line was completed. In the latter half of this same century the street railway came into its own, reaching a peak just prior to the first World War. The great growth of motor-car and truck transportation began about 1910, and continues to grow. Pipe lines, first built about 1878, began their rapid period of growth in the early twentieth century, and air transportation, begun in a serious way in the 1920's, is now just beginning to make its mark. Thus we get some idea of the revolution that has taken place in the field of transportation roughly within the past 100 years. There has been a greater change in the commonly used methods of transportation within the past century, or century and a half, than there had been in 1000 or 1500 years before. Still the purpose is to get men to desirable goods and to get desirable goods to men. The method may become more complicated, swifter, cheaper, or more spectacular, but the purpose stays the same.

It is easy to distinguish transportation from the other sorts of economic activity. There is seldom any question whether a given activity should or should not be classified as transportation. But transportation activities are themselves multiform and varied. The different kinds compete with each other as users seek the cheapest, the quickest, the safest, the pleasantest, or the most romantic means of conveyance. Water transportation competes with rail transportation; rail transportation competes with motor transportation; motor transportation competes with air transportation; while still over vast regions of the earth men carry heavy burdens on their own backs (or more often on the backs of their wives) or make use of the slow-plodding buffalo and ox. We have not the space here to describe in detail the problems of each of these kinds of transportation but reduce our field of

investigation again, this time to consider briefly the problems of railroad transportation as roughly illustrative of the broader problems of transportation in general, just as the problems of transportation, in turn, are roughly illustrative of the larger field of public utilities. Among the problems of railroad transportation, we shall naturally neglect the engineering problems, those of metallurgy and of physics, and the purely legal problems, and confine our attention to the economic characteristics of the railroad industry and to the problems that arise in connection with them.

26 · 13. Cost Problems in Railroad Transportation. The railroad industry offers a splendid illustration of several principles already studied in connection with costs. For example, the fixed costs that are sunk into rights of way, roadbed, stations, freight terminals and classification yards, warehouses, shops, etc., are tremendous and are, to a considerable extent, independent of the volume of freight or passenger traffic handled by the railroad. Any industry that has a large element of fixed costs is apt to operate under conditions of decreasing cost. Additional business may increase the total cost slightly, but the increase is so small that the average cost is reduced. Suppose, for example, that a given railroad has annual fixed costs of 10 million dollars. It is carrying 1 billion ton-miles[1] of freight, and has variable costs of 1 cent per ton-mile for fuel, etc. Thus the total variable expense is 10 million dollars and the total expense is 20 million, or 2 cents per ton-mile. If the railroad doubles the volume of freight carried and if it can do so without increasing the fixed expense, the total cost will now be 30 million dollars and the cost per ton-mile on the 2 billion ton-miles will drop to 1.5 cents. The student can easily see that this railroad would continue to operate under decreasing costs, with the cost per ton-mile dropping to 1⅓ cents when the traffic rose to 3 billion ton-miles, to 1¼ cents at 4 billion ton-miles, to 1⅕ cents at 5 billion ton-miles, etc., as long as the road could handle the increased traffic without being forced to increase its fixed costs. Ultimately, to be sure, it would become necessary to lay extra tracks or provide additional terminal facilities. Then, and only then, would the tendency for decreasing cost temporarily disappear.

The relative importance of fixed costs in railroad transportation has several obvious and necessary results. In the first place, it means that competition between railroads is apt to be suicidal and lead to cutthroat competition and the ultimate deterioration of service.[2] Regulation of railroad rates has been necessitated at least in part, surprisingly enough, by a tendency

[1] The ton-mile is the unit most commonly used for measuring the volume of freight traffic. If a railroad hauls 100 tons for 10 miles it has carried 1000 ton-miles of freight; the hauling of 1 ton for 1000 miles would likewise have been counted as 1000 ton-miles. The total ton-miles of traffic are found by multiplying the tonnage hauled by the distance over which it is hauled.

[2] The student who does not find this point immediately apparent should review Sec. 17 · 6.

for the roads to set rates that are too low when they are not regulated. Here we have another reason, in addition to those mentioned early in this chapter, for public utilities to be at least in considerable degree monopolistic.

The large element of fixed costs also makes for unusual fluctuations in the earnings of the railroads. We can illustrate this with the hypothetical situation of the second paragraph preceding. Assume that the railroad is receiving 1.2 cents per ton-mile as its income. When it hauls 6 billion ton-miles, it will have total income of 72 million dollars and total expenses of 70 million, or profits of 2 million. If we increase its traffic from 6 to 7 billion ton-miles (an increase of one-sixth, or 17 per cent) we increase the profits from 2 to 4 million dollars, or 100 per cent; if we reduce the traffic from 6 to 5 billion ton-miles, the profit will be wiped out entirely. Although these figures are arbitrary and hypothetical, they illustrate the kind of thing that does actually happen. *In any business that has large fixed costs, small percentage variations in the volume of business ordinarily bring large percentage variations in the amount of profit or loss.* Even minor year-to-year variations in the volume of freight have important effects on railroad profits, and a period of serious depression is apt to leave scars that can be effaced only by refinancing the enterprise.

26 · 14. Railroads and "Joint Cost." Textbooks on transportation often tell us that the railroads are examples par excellence of products produced under conditions of joint cost. They point out, for example, that an ordinary freight train carries many different products and that the cost of the roadbed, the fuel, and the wages of the train crew—in fact, practically all the costs—are borne jointly by these diversified kinds of freight. The car full of gravel, coal, furniture, or farm machinery, the tank car full of oil, the refrigerator car full of perishable fruit, and the car carrying miscellaneous job lots of freight cannot be separated in the assignment of costs. There is no way to distinguish the cost of hauling apples from the cost of hauling lumber.

Although this is what many textbooks say and although the facts are substantially true as described, the nomenclature is not that usually accepted by economists. The student will recall[1] that the situation just described is a case of *overhead costs* or of *unassignable costs,* but not of *joint cost.* We do not call it joint cost, because it is easily possible for the railroad to carry more lumber without carrying more apples. The quantities of the various kinds of freight are not irrevocably bound together, and their proportion can be and often is changed quite radically. There is, to be sure, some reason to talk of joint costs in rail traffic when we consider the relationship between freight hauled in different directions—between eastbound and westbound freight, for example. The railroad cannot send more cars, more engines, and more train crews in one direction than in the other. If some set of forces

[1] This matter was treated in Sec. 17 · 7.

increases the number of cars running eastward, the number of cars available to run westward is automatically increased, and presumably the charge for westbound freight will fall in the absence of regulation.

Even if we quarrel with the nomenclature commonly used, we would have to admit at once that the very large proportion of overhead or unassignable cost in the railroad business is a matter of importance. It means, obviously, that it is impossible to ascertain with any accuracy the actual cost of handling any particular kind of freight.[1] The railroad cannot tell, except on the basis of purely arbitrary accounting distributions, what it costs to ship a car of apples over its lines. It cannot separate, as we have already said, the apple cost from the lumber, the gravel, the milk, or the furniture cost.

If the cost cannot be ascertained, what is the railroad to charge for its services? How much should it charge to haul the carload of apples? Frankly, the railroad has no way of knowing. As a result, if it appears likely that the apple business will be lost to competitors, the railroad tends to reduce the rates lower and lower, not being held back by any definite cost figure. It is hard to prove that the company is losing money by carrying the apples. We just mentally assign the costs to the other products and let them "pay the freight." The trouble is that, if our freight agent does this successively for each of the products hauled, none of the freight will pay its way and the railroad company will lose money. Here we find still another reason why competition among railroads tends to be costly and cutthroat and to drive rates below the actual cost of doing the business.

We shall see shortly in addition that this inability to ascertain costs accurately, or at least the inability to apportion them logically among the various kinds of freight, leads to the charging of very high rates on some commodities and very low rates on others; "charging what the traffic will bear," as the saying goes, rather than charging the actual cost of the service. Some authorities have argued that rates should represent the cost of the service; others, that they should represent the value to the shipper. In a famous case the U. S. Supreme Court held[2] that, "What the company is entitled to ask is a fair return upon the value of that which it employs for the public convenience. On the other hand, what the public is entitled to demand is that no more be exacted from it . . . than the services rendered are reasonably worth." Here we see both ideas juxtaposed: that of rates based on cost and that of rates based on the value of the service or "what the traffic will bear." The Supreme Court seems to feel that the actual rate should lie between the two. If costs cannot be ascertained (in fact, they cannot) at least one-half of the court's rule falls to the ground. Probably we should add that it is quite as difficult to ascertain the value of the service

[1] The point could be made as well with passenger service, of course.

[2] *Smyth v. Ames*, 169 U. S. 466-550 (1898).

to the shipper as it is to discover the cost of the service to the railroad. Of the two bases suggested in the so-called *Smyth v. Ames* rule, the courts, being staffed with lawyers rather than with economists, have tended to lean rather heavily on the cost approach, assuming implicitly in most cases that the cost of the service is something that can be measured. It would probably be a good rule for every embryo student of economics to paste in his hat and remember that, *in practice, costs are usually very difficult to ascertain except on some arbitrary basis.*

26·15. Discrimination in Railroad Rates. We discovered in Sec. 16·14 that monopolists find it advantageous to charge different prices to different users of their products. Here again the railroad industry furnishes ready-made examples. In connection with their passenger traffic, it is common for railroads to make different charges for travel in compartments, in lower berths, in upper berths, in parlor cars, and in day coaches. The seller offers slightly different kinds of service to different customers and charges them radically different prices. Similarly, in connection with freight traffic, the railroad usually charges very different rates for hauling different kinds of freight. Some of these charges, to be sure, reflect differences in cost. Thus, the railroad charges more for less-than-carlot shipments than for carlot shipments, or more for perishable commodities that require icing and special handling than for nonperishable commodities; but where a higher freight rate merely reflects higher cost to the carrier, we do not class it as discrimination. Discrimination in rates is the charging of different rates that do not reflect differences in cost. Such discrimination is a corollary both of the monopoly position that rails usually enjoy, and of the large fixed and unassignable costs mentioned earlier in this chapter.

Discrimination in railroad rates is of three different kinds: *Commodity discrimination* is the charging of different rates on different commodities other than those which reflect differences in the cost of handling them. Figure 26·1 shows graphically the rates charged for hauling 100 pounds of various commodities a distance of 400 miles from Boston in 1946. In all cases the length of haul, the weight of the commodity, the point of shipment, and the destination were the same. It will be noted that the cost of shipping 100 pounds of wheel chairs was almost four times as great as that of shipping 100 pounds of nails. Yet these cases are by no means extreme ones. It is quite possible to find some rates ten or twenty times as high as others.

Even if there are no differences in costs of handling, it is still easy to understand why commodity discrimination takes place. Some can stand higher transportation charges than others. A 100-pound keg of nails might cost in the neighborhood of $5, while a 100-pound wheel chair would cost perhaps $50. If they were both charged the same freight rate, the freight cost would be a much larger *percentage* of the cost of the nails than of the wheel

chair. Thus, if the freight cost were $1 on each, it would amount to 20 per cent on the nails and 2 per cent on the wheel chair. The point is that, if the roads charged equal rates per pound on different kinds of freight, they would not be charging anything like equal proportions of the value of the thing shipped. A hundred pounds of radium would be worth so much that a freight rate of $100 would be negligible; but 100 pounds of sand are worth so little that the buyer could not afford to pay that amount to have it shipped. At such a freight rate those commodities, like sand, gravel, iron ore, and coal, which have only a small value in great bulk, could not be shipped. And the freight agent, seeing that they cannot be shipped and knowing that there will be little *added* cost to the railroad by carrying them along with

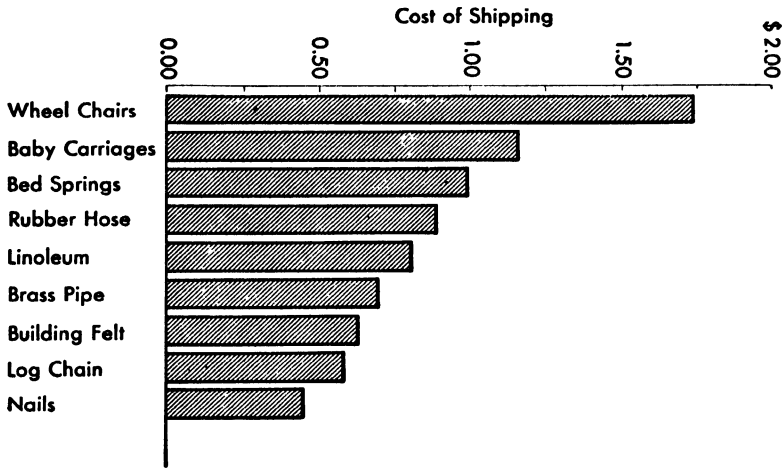


FIG. 26 · 1.—Cost of shipping 100 pounds of various kinds of freight a distance of 400 miles from Boston in 1946.

other goods (since most of the costs are fixed), offers special reduced rates to encourage their shipment. As we know, if he gets anything more than enough to cover the variable expense (and the variable expense is very small indeed), he will be better off to carry them than to pass them by. Some commodities pay the variable expense plus a share of the overhead cost, while others are carried at a rate that barely covers the variable expense. If the carrier is not regulated, it will tend to charge "what the traffic will bear" —high rates on some goods and low ones on others.

It is commonly, although erroneously, supposed that "charging what the traffic will bear" puts an unfair burden on freight that pays the high rates and that this high-pay freight somehow subsidizes the low-pay freight and helps pay its way. But it should be clear that no freight whatever will be carried unless it pays enough to cover the variable cost. If it pays anything

at all toward the fixed cost, there is just so much less fixed cost to be paid by the high-pay freight. The low-pay freight has not increased the burden on the high-pay freight. If it barely paid the variable expense, it left that burden the same; if it paid anything whatever above the variable expense, it decreased that burden.

26 · 16. Local Discrimination. A second type of discrimination in rates is that between localities or areas. Ordinarily the cost of hauling freight increases with the distance, and consequently higher rates for longer distances cannot be considered discrimination in themselves. But, if rates

FIG. 26 · 2.—Hypothetical case of local discrimination in freight rates.

differ for the same distance or if a higher rate is charged for a short than for a long haul, discrimination can be suspected. Such *local discrimination* may be the result of special conditions of competition. Thus, for example, it has been true that commodities shipped from San Francisco to Albany carried a higher rate than the same commodities shipped from San Francisco to Boston, even though the shipment to Boston passed through Albany en route. In fact, the rate from San Francisco to Albany was equal to the rate from San Francisco to Boston plus the rate from Boston back to Albany! The reason was that shipments from San Francisco to Boston had to meet the competition of cheap ocean traffic through the Panama Canal. Unless the transcontinental rail rate were low, the freight would go by

boat rather than by train. This reduced the San Francisco-Boston rate. But water-borne freight destined for Albany had to go to Boston and then finish the trip by rail. The Albany freight was not so directly competitive with water freight as was the Boston freight. Such illustrations of local discrimination have been common and where, as in this case, they merely recognize a natural advantage which one area or market (Boston) has over another (Albany), the advantage is not the result of the discrimination, but would exist without it.

We can illustrate the situation graphically as in Fig. 26·2. Suppose that we have a railroad running from St. Louis to Louisville and from Chicago to Louisville. The two distances are approximately equal. If firms in St. Louis and in Chicago are both manufacturing the same commodity and both selling in Louisville, the Chicago producers are sure to complain of discrimination if the railroad charges substantially less to haul from St. Louis than from Chicago; yet if water transportation from St. Louis to Louisville on the Mississippi and the Ohio Rivers is inexpensive, St. Louis already has an advantage over Chicago in serving the Louisville market. Differences in freight rates do not make the advantage and, if the railroad does not give the low rate from St. Louis to Louisville by rail, the commodity will go by boat and will still undersell the Chicago product at Louisville.

26·17. Personal Discrimination. The third type of discrimination in railroad rates is what we call "personal discrimination." We have already seen that an unregulated monopoly will tend to discriminate among its customers, charging higher prices to some than to others. As long as the differences in the charges represent differences in the cost of the service rendered, there is no reasonable objection, nor would we regard such differences as discriminatory. Yet we know that a monopoly finds it profitable to discriminate even where there are no differences in cost.

Prior to 1887, when legislation was passed to prohibit it, personal discrimination had been widespread. So common was it that it was sometimes said that no one except the innocent paid the announced rates. Friends of the management, large shippers, or influential politicians were given passes, which allowed them to ride on the roads without paying. Important companies were given special rates on the shipment of freight. The monopoly of the petroleum business built by Standard Oil relied primarily on a system of railroad rebates under which the oil company paid the same rates for shipping oil as anyone else but secretly received a rebate of over 70 per cent of the freight rate on all oil that it shipped and, in addition, received a rebate of the same amount on all oil shipped by its competitors! This was an unusual case, but personal discrimination was not unusual. It was the rule rather than the exception. One of the officials of the New York Central Railroad, in commenting on the situation, remarked that "the condition of giving the

special rate is making the application."[1] That is, special rates were given to anyone and to everyone who asked for them.

Personal discrimination is looked on by most people as decidedly objectionable, not only as being undemocratic but as tending to destroy that equality of opportunity in business which is essential for the maintenance of anything approaching a competitive system. Ever since 1887, when Congress passed an "Act to Regulate Commerce" which sought to eliminate unjust discrimination, one act after another has been passed in an effort to eradicate the evil. The Interstate Commerce Act (as the act of 1887 is commonly called) states in Sec. 3,

> That it shall be unlawful for any common carrier subject to the provisions of this act to make or give any undue or unreasonable preference or advantage to any particular person, company, firm, corporation, or locality, or any particular description of traffic, in any respect whatsoever, or to subject any particular person, company, firm, corporation, or locality, or any particular description of traffic, to any undue or unreasonable prejudice or disadvantage in any respect whatsoever.

Yet personal discrimination takes so many forms and can be so easily disguised that it is exceedingly difficult to wipe it out altogether. The student will note that Sec. 3 of the Interstate Commerce Act, quoted above, prohibits personal, local, and commodity discrimination—all three—but it prohibits them only as they are undue or unreasonable. Regulatory bodies and commissions and the courts have been more liberal in allowing commodity than local discrimination, and far more liberal in allowing either of them than in allowing personal discrimination. Each kind of discrimination is practiced less frequently than it was prior to 1887, and it is probably fair to say that personal discrimination by railroads has become rather uncommon.

26 · 18. Transportation and Price Geography. Goods produced in different localities have to sell at different prices because of differences in the local costs of production. Cotton grown in Maine would have to sell for much more than cotton grown in Texas if the production were to be profitable; Minnesota lemons would have to sell for much more than Florida lemons if they were to repay production costs. If each locality produced all its own products for consumption, prices in different places would differ widely.

In practice, however, it is often possible and economical to transport goods from the places where they are produced cheaply to the places where production is more costly, and we have seen that where such transportation is feasible, there is ordinarily a gain through trade to both the importing and the exporting area.[2] Under these conditions, *the prices of a commodity in two different markets cannot differ by more than the cost of transporting the com-*

[1] Report of the Hepburn Committee, p. 159, New York, 1879.

[2] See Sec. 10 · 9 *et seq.*, 23 · 1, and 23 · 9.

modity from the low-priced to the high-priced market. If, for example, eggs are worth 50 cents a dozen in Boston and 75 cents a dozen in New York, and if the cost of transporting eggs from Boston to New York is $1 a dozen, the price discrepancy can continue; but if, under the same circumstances, the cost of transporting eggs from Boston to New York is 10 cents a dozen, the discrepancy will in part disappear. Merchants will find it to their advantage to buy eggs in Boston for 50 cents, spend 10 cents to ship them to New York,

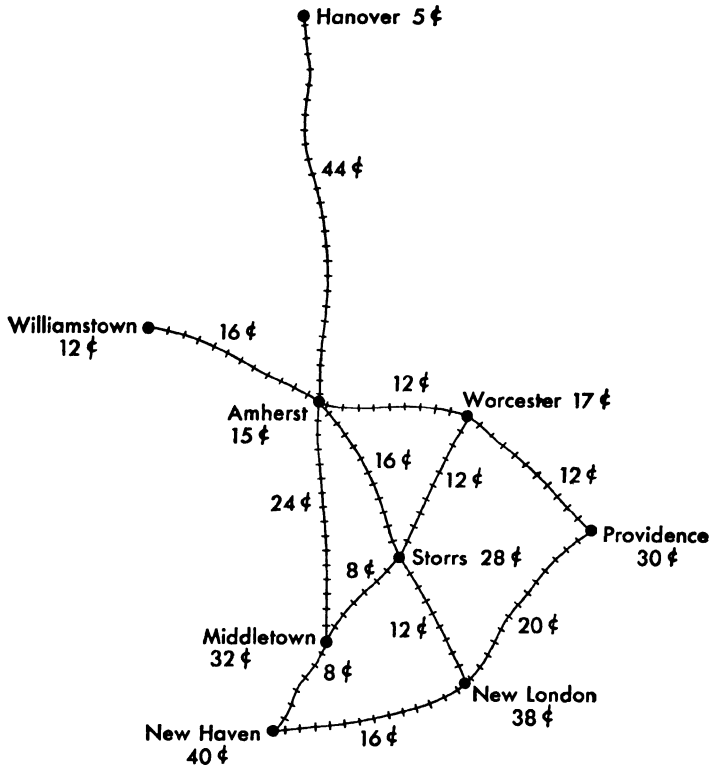


FIG. 26 • 3.—Cost of production within and costs of transportation between New England towns.

and sell them there at the prevailing New York price of 75 cents. As more and more merchants attempt to take advantage of this price situation, they bid against each other for eggs in Boston, raising the price there, and they compete with each other to sell eggs in New York, lowering the price there. This will continue as long as there is any profit in it—as long as the New York price exceeds the Boston price by more than 10 cents. The transportation of the goods from one market to the other will tend to lower the prices in the high-priced market and raise those in the low-priced market until they differ by an amount equal to the cost of transportation.

Perhaps the situation can be understood better in connection with a more complicated illustration. Figure 26·3 gives data for a fictitious case, showing the cost of producing a commodity in each of nine cities, and the costs of transporting the commodity between these cities. On this map, the figure that accompanies the name of each city represents the cost of producing a unit of the commodity in that city, while the figure adjacent to the railroad line between any pair of cities shows the cost of transporting a unit of the commodity between those cities. Thus the diagram indicates that it costs 30 cents to produce the commodity in Providence and would cost 12 cents to ship it to Worcester from Providence (or from Worcester to Providence).

Naturally, men are never going to find it advantageous to buy a commodity in a high-priced market and pay transportation costs to ship it to a market where the price is already lower. But it will pay to buy in any low-priced market and ship to any market where the local cost of production is higher *if* the difference in prices exceeds the cost of shipping. In this particular case, Hanover uses locally produced products, which cost 5 cents per unit. Amherst produces the commodity at 15 cents, but it would not pay to import from Hanover and pay 44 cents shipping cost, nor to import from Williamstown and pay 16 cents shipping cost. Inspection of the diagram will show the sources of supply and delivered costs of the commodity given in the accompanying table.

City	Source	Price, cents
Hanover	Hanover	5
Williamstown	Williamstown	12
Amherst	Amherst	15
Worcester	Worcester	17
Providence	Worcester	29
Storrs	Storrs	28
Middletown	Middletown	32
New London	New London	38
New Haven	New Haven or Middletown	40

Providence could produce the commodity for 30 cents, but it is cheaper to buy it for 17 cents in Worcester and pay 12 cents ship it to Providence, making the total cost 29 cents. New Haven can use either the homemade product at 40 cents or buy Middletown products at 32 cents and pay an 8-cent shipping charge. The final cost will be the same either way.

Now, let us suppose that the production costs in the different cities stay unchanged but that the cost of transportation is cut in half, to just half the

amounts shown in Fig. 26·3. The student will discover after a moment's trial that the sources and costs of the product in the different communities will now be as shown in the next table.

City	Source	Price, cents
Hanover	Hanover	5
Williamstown	Williamstown	12
Amherst	Amherst	15
Worcester	Worcester	17
Providence	Worcester	23
Storrs	Amherst or Worcester	23
Middletown	Amherst or Storrs	27
New London	Storrs	29
New Haven	Middletown	31

Under these new transportation conditions, Storrs, Middletown, New London, and New Haven all give up production and rely on other areas where costs are lower. In each case, the new cost to the consumer is lower than the former cost. The well-located areas of low-cost production, Amherst and Worcester, have now taken over all the market save that in Hanover and Williamstown. New Haven buys from Middletown, but Middletown buys from Amherst or Storrs, and any that is bought from Storrs came, in turn, from Amherst or Worcester. Thus New Haven's supply comes ultimately from Amherst or Worcester, even if it comes proximately from Middletown. The student will discover that, if the cost of transportation is again cut in half, to one-quarter the amounts shown in Fig. 26·3, the tendency will be to concentrate production even more in Amherst, and again to lower the price of the product in all importing areas. In general, it can be said that *every reduction in the cost of transportation tends to localize production more and more in centers of low-cost production, and to reduce the price of the product in those areas which cease to produce and start to import.* If transportation could be made entirely costless, all production would take place at the single point of most efficient production, and all other areas would get the full advantage of the low cost in that most efficient area.

Changes in transportation cost, then, have a good deal to do with determining the economic geography of a nation or of the world. The differences in local production costs (which are the result of differences in climate, richness of resources, etc.), coupled with the cost of transportation, determine the price geography of the area. We can see this in the hypothetical case of Fig. 26·3, or in actual practice. Thus Fig. 26·4 shows the 10-year average

price of potatoes from 1906 to 1915 in different parts of the United States. All points on any given line in this diagram had the same average price, and points between lines had intermediate prices. A quick glance at the diagram will indicate that the price was low in the two great producing areas of Minnesota and Maine and that, as potatoes were shipped farther and farther from these markets, the prices became higher and higher with the inclusion of transportation costs. Similarly for every other commodity, the prices are a composite of all production costs if we use the word "production" in its broadest sense to include the adding of place utility. Every reduction in transportation costs makes it possible for men to get the products that they

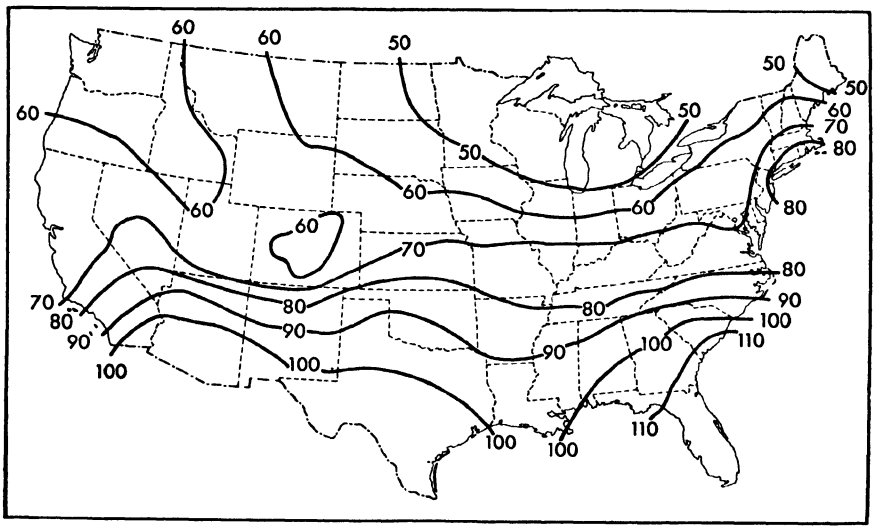


FIG. 26 • 4.—Average farm price of potatoes, 1906-1915, cents per bushel. (*From Holbrook Working, "Factors Determining the Price of Potatoes in St. Paul and Minneapolis, Univ. of Minn. Agric. Expt. Sta. Technical Bull. No.* 10, 1922, *p.* 6.)

want with less effort, by bringing them from areas of most efficient production.[1]

26 · 19. Cost of Transportation on American Railroads. Having discussed at some length in abstract terms the economic principles that underly the transportation of goods, let us turn to an appraisal of the transportation system of the United States in the concrete. At the close of the calendar year 1944, the railroads of this country were operating 240,000 miles of first main track[2] using 42,000 steam locomotives, 900 electric locomotives, and 3400

[1] See Note 14, Mathematical Appendix, p. 909, for a more technical treatment of some aspects of this matter.

[2] First main track excludes 41,000 miles of secondary track and 117,000 miles of siding and yard track. The figures in this paragraph are from the 59th Annual Report of the Interstate Commerce Commission, Nov. 1, 1945, Appendix C.

diesel-electric locomotives to pull 1,800,000 freight and 38,000 passenger cars. The total railway capital amounted to 19.5 billion dollars, and the net income of the roads totaled 733 million. Some idea of the tremendous expense involved in railroad building can be gathered from the fact that the investment per mile amounted to $118,000. Freight was almost four times as important as passenger traffic as a source of railroad income. Freight revenue amounted to 7.0 billion dollars and passenger revenue to 1.8 billion. The Class I railroads[1] employed an average of 1.4 million persons at an average wage of $0.965 per hour. The railroads carried 740 billion ton-miles of freight and 96 billion passenger-miles of passenger traffic. This was done with an income of 0.957 cent per ton-mile of freight and 1.875 cents per passenger-mile. These figures are roughly half the costs for the corresponding services on the railroads of European countries. At least part of the low American cost is from the long average hauls. The average freight shipment in this country traveled 473 miles in 1944, and the average journey per passenger was 104 miles. Both of these were unusually large because of war conditions in 1944.

The stupendous size of the task undertaken by the American railway system, and the surprising efficiency with which the work is accomplished, can be best visualized by means of a striking example suggested by a leading scholar in the field of American Economic history.[2] Let us suppose that a man could carry a load of 100 pounds a distance of 20 miles in one day.[3] If we put every man in the United States between the ages of twenty and sixty-four to work carrying freight and kept them at it 365 days per year without any holidays, it would take them 26½ years to carry the revenue-paying freight hauled by the American railroads in 1940. The railroads carried each ton-mile (which we have considered would be a hard day's work for a man) for just under 1 cent, while pure man power would doubtless have cost 200 to 300 times as much. This suggests what a laborsaving device modern society possesses in the railroads. They did this work in 1940 with about 1 million employees and about 25.6 billion dollars of investment, or an investment of about $25,000 per employee. This investment made it possible for men all over the country to concentrate their efforts on the production of the things that they could produce best, relying on the people of other areas to produce most of the things that they desired.

The total labor force of the United States in 1940 consisted of about 53 million people. Of these, a few over a million were employed by the rail-

[1] Class I railroads are those having operating revenues in excess of 1 million dollars per year.

[2] The example is brought up to date from Chester W. Wright, "Economic History of the United States," p. 597, McGraw-Hill Book Company, Inc., New York, 1941.

[3] This is 1 ton-mile. The student might as well imagine carrying 50 pounds for 40 miles, or any equivalent task.

roads, or about 2 in every 100. If we include all transportation agencies, the number of employed was about 2.2 million, or about 4 in every 100. If we include all other public utilities along with transportation, we find some 3.1 million employed persons, or about 6 in every 100. Very evidently a large part of the productive energy of the American people is being spent in transporting goods from place to place, probably a larger proportion than at any other time in history.

26·20. Position of the Railroads among the Transportation Agencies. Figures 26·5 to 26·7 show graphically the relative importance of the railroads as compared with other means of transportation in the United States.

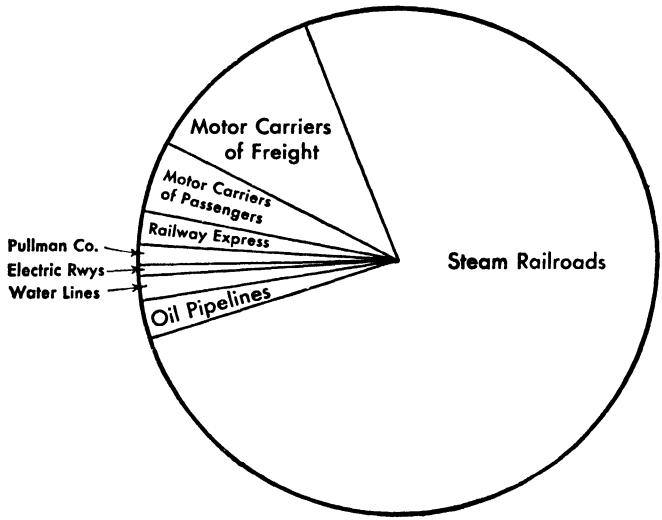


FIG. 26 · 5—Distribution of operating revenues among various sorts of carriers, year ending June 30, 1945. (*Data from 59th Annual Report, Interstate Commerce Commission, November* 1, 1945, *p.* 12.)

Figure 26·5 throws together both passenger and freight service, making the comparison on the basis of total income. Here the railroads account for just over three-quarters of the total, being three times as important as all other carriers combined. Figure 26·6 separates out the freight service, and here the railroads account for 70 per cent of the total. When we shift to the passenger traffic, shown in Fig. 26·7, we see that the privately owned automobile has become the chief carrier, accounting for over 50 per cent of the total. The railroads fall in second place with just over a third. This growing competition of private automobiles and passenger busses, making up between them almost two-thirds of the total intercity passenger traffic, has been one of the major transportation developments of the past two and a half decades, and the accompanying problems have been difficult ones for the railroads.

In Sec. 26 • 13 we noted that even small fluctuations in gross income are apt to bring major fluctuations in net income for businesses that have, as the railroads do, large fixed costs. And the reduction in income that has accompanied the loss of passenger traffic has, accordingly, had a disproportionately large effect on the prosperity of the railroads.

26 • 21. Regulation of the Railroads. During the early years of their expansion, the railroads were met with wild public enthusiasm and popularity. Every town sought to entice one or more roads to pass through it. States and the Federal government offered various kinds of inducements to encourage the building of railroads. The Federal government gave over 130

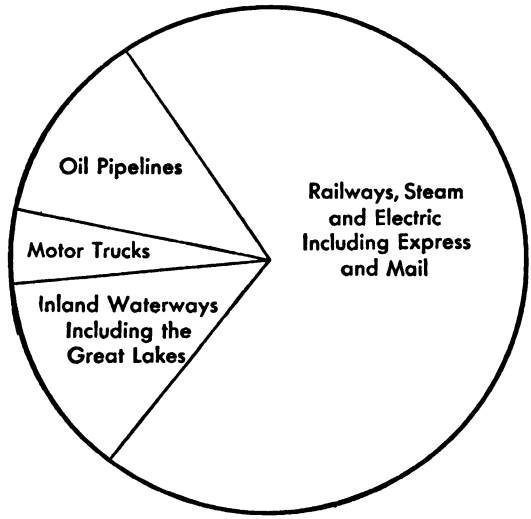


FIG. 26 • 6.—Distribution of ton-miles of intercity freight carried by various types of carriers, 1944. (*Data from 59th Annual Report, Interstate Commerce Commission, November* 1, 1945, *p.* 13.)

million acres of land (an area approximately as large as that of the combined states of Ohio, Indiana, and Illinois, and a third again as great as New England and New York combined) for this purpose alone. It is only fair to say, of course, that this 9.5 per cent of the public domain given away for this purpose served to make the remaining 90.5 per cent far more valuable —to open up more quickly the great open spaces of the West. Railroads could hardly afford to build into unsettled territory without some sort of subsidy, yet railroads were necessary for swift settlement. Moreover, the railroads receiving the grants gave a valuable *quid pro quo* by granting reduced rates for carrying mail, government freight, and government passengers.

The early popularity of the railroads tended to disappear, however, as the abuses of rebates and discrimination became widespread, and as the

general public, suffering keenly from the effects of the severe depression of 1873, looked for a scapegoat to carry the blame. The depression hit hardest in the agricultural regions of the Middle West, and here it was that distrust of the railroads led to the first repressive and regulatory legislation, the so-called "granger" laws. Various states set up commissions with power to set rates and otherwise regulate the railroad operations. These measures were bitterly contested by the railroads, but the right of the states to regulate was finally upheld by a divided Supreme Court,[1] which held that the government might regulate any business that was "clothed with a public interest," and included among the permitted types of regulation the right of price control.

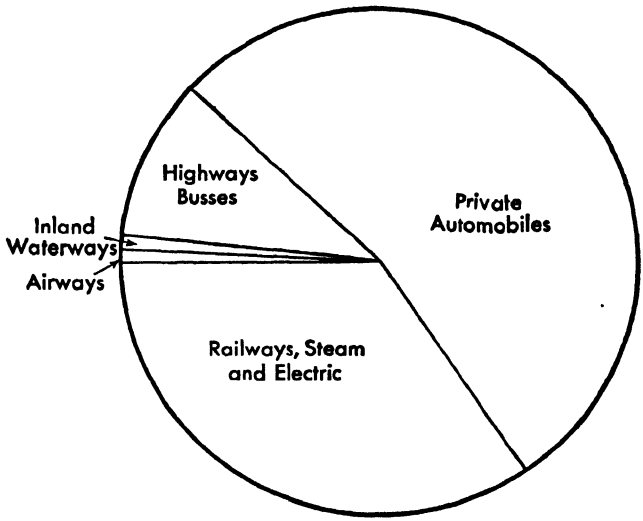


FIG. 26 • 7.—Distribution of intercity passenger mileage among various types of carriers, 1944. (*Data from 59th Annual Report, Interstate Commerce Commission, November* 1, 1945, *p.* 13.)

But the granger laws were for the most part of little effect, even though they were held to be constitutional. Their weaknesses were of two kinds: (1), they had been written in a spirit of hostility and of ignorance of underlying economic forces and (2) they were perforce confined to intrastate commerce, whereas the business of transportation was already becoming primarily interstate in character. The major contribution of these laws was not the detail of their content, but rather the fact that they broke the ground, established a legal foundation for regulation, and pointed out some of the pitfalls to be avoided.

In 1887, Congress passed the first Federal regulatory act, An Act to Regulate Commerce. This act, since known as the Interstate Commerce Act,

[1] The basic case is 94 U. S. 113, *Munn v. Illinois,* 1876.

subjected the interstate activities of the railroads to Federal regulation, prohibiting discrimination of various kinds in the wording already quoted in Sec. 26·17 and attempting by various means to preserve competition among the railroads. The public viewpoint had been changed to the extent of considering the railroads as semipublic in character, somewhere between pure private business on the one hand and government-owned businesses on the other. But the public was not yet ready to acknowledge their fundamentally monopolistic character and for several decades continued to add to the difficulties of the industry by attempting to force competition in a field that was unsuited for it.

The Interstate Commerce Act established the Interstate Commerce Commission, commonly called the "ICC." This was the first great public regulatory agency established by the government, and in the following 60 years it was to set the example for many others. Like all regulatory bodies, its most important policies are those which have to do with the regulation of rates, but it also is required to eliminate unjust discrimination, to approve the issuance of securities, the construction or abandonment of lines, and the consolidations of carriers. It has jurisdiction over the use of safety equipment by the railroads and may prescribe the accounting systems to be used by them. It has jurisdiction today over all interstate transportation except that by air, including railroads, waterways, pipe lines, and motor vehicles. The Commission at present consists of 11 members, appointed by the President "by and with the advice and consent of the Senate."

The ICC did not get undisputed powers immediately. Amendment after amendment was passed to the Interstate Commerce Act by succeeding Congresses. Not until the passage of the Hepburn Amendment in 1906 was the Commission given final power in rate making (subject, of course, to review by the courts), and since that time the courts have been restricted to decision of points of law in connection with transportation cases, leaving the points of fact to the Commission. So things continued until the outbreak of the first World War. During this conflict, the United States government took over the railroads and operated them for 2 years and 2 months as a single unified system. The owners were recompensed by an annual rental equal for each road to its average earnings during the preceding 3 years. From the standpoint of military operations, the period of government operation was a success; goods and men were moved in unprecedented quantities and with dispatch. Financially, the experiment did not turn out so well. The deficit for the twenty-six months of operation was some 1.2 billion dollars. It is but fair to state, of course, that during this period the primary objective was not that of earning a profit, but rather that of prosecuting the war to a successful conclusion.

With the return of the railroads to private hands on Mar. 1, 1920, it was felt wise to adopt a new public policy toward them. This was embodied in

the Esch-Cummings Act, or, as it is commonly known, the Transportation Act of 1920, which is in the nature of an amendment to the original Interstate Commerce Act of 1887. Now, for the first time, the public had come to look upon the railroads, not as separate competing private agencies, but as a coordinated quasi-public system—as a "public utility." It was at this time that the Commission was given jurisdiction over the issuance of railroad securities. The act also set a new standard for rate making which was somewhat more specific than the vague "fair return on a fair valuation." The Commission was instructed to ascertain the value of each railroad, to divide them all into groups, and to set the rates in such a way as to allow each group as a whole to earn between 5½ and 6 per cent on its value. There was no guarantee that any particular road would earn this much, or, indeed, that it would earn anything at all; but the attempt was to be made to set rates in such a way as to ensure this supposedly "reasonable" return on the average. The Transportation Act of 1920 also directed the ICC to prepare plans for the consolidation of the railroads into a number of systems, thus turning away from the old idea of enforced competition to that of encouragement of efficiency through regulated monopoly. The Commission, however, has no power to enforce such consolidation, and to date nothing of importance has been done under these provisions.

Congress could not know, however, that it had awakened too late. After having tried for decades to force competition in a field that was naturally monopolistic, Congress recognized the facts just as they were on the point of changing. For the period since 1920 has seen radical changes in the transportation field, and most of which have been in the direction of increased competition for the railroads rather than of increased monopoly. The development of the pipe lines for shipping petroleum played a part, but by far the most important element in the new situation was the motor car—the private passenger automobile, the passenger bus, and the motor truck. These new agencies were not tied down to any fixed routes but had a flexibility that the railroad could not meet. Then on top of the new competition came the depression of the thirties. Net operating income for the railroads of the United States, which had been running at a level of about 1.1 to 1.2 billion dollars per year, dropped in 1930 to 847 million, then in succeeding years to 528 million and 325 million dollars. Not until the outbreak of the second World War did it again reach a billion dollars. Railroad net incomes were cut in 4 years to a quarter of their predepression level. Under such circumstances it was manifestly impossible for the ICC to establish rates that would earn 5½ or 6 per cent on the value of the property, and in 1933 Congress, in an Emergency Railroad Transportation Act, repealed this provision. No final new policy of regulation had yet been adopted, however, at the time of the outbreak of the second World War. For the time being, at any rate, the depression problems disappeared. The railroads had more business than

they could handle. This time the roads were allowed to remain in private hands and acquitted themselves well under the pressure of emergency wartime traffic. Although it is still too early to tell what direction postwar regulation will take, it seems certain that the railroads will continue to be looked upon as semipublic agencies and that their natural tendency to consolidate into larger and larger units will be encouraged under some safeguards, which will see to it that the strongest roads do not band together and leave the less efficient ones on the outside.

Questions

1. Describe briefly the right of eminent domain.
2. Why is it more necessary to exercise the right of eminent domain for a railroad than it would be for a barber shop?
3. Why is the demand for the products and services of public utilities rather inelastic?
4. Why is the public especially interested in those commodities which, produced by monopolies, are characterized by inelastic demand?
5. Does it seem fair to you for the government to force public utilities to submit to various sorts of regulation from which other businesses are exempt? Is this discriminatory? Discuss.
6. Give an example of the disadvantage of choosing, for price-fixing purposes, a basis of actual original investment. A basis of cost of reproduction. A basis of actual present capitalization (amount of stocks and bonds outstanding).
7. A given business has gross annual income of $15,000 and gross annual expenses of $9,000. If it is anticipated that these figures will continue to describe the situation and if the going rate of interest on investments is 6 per cent, what is the business worth? What is it worth if the going rate of interest is 5 per cent?
8. What is the effect of a change in the going rate of interest on the value of property that yields a fixed income? Why?
9. If a government bond with face value of $1000 yields 2 per cent interest while other equally attractive investments yield 6 per cent interest, what is the market value of the bond? Does it make any difference how long the bond has to run before maturity? Discuss.
10. In Sec. 26 • 13 appears the statement, "Any industry that has a large element of fixed costs is apt to operate under conditions of decreasing cost." Would it have been correct to state that such businesses *always* operate under decreasing cost? Comment.
11. If the railroad mentioned in the third paragraph of Sec. 26 • 13 hauls only 4 billion ton-miles of freight, what will be the amount of its profit or loss?
12. Explain the relationship between fixed costs and discrimination in railroad rates. If none of the costs were fixed, how would the situation be different?

13. Give examples of price discrimination by other monopolies than the railroads.

14. What do you mean by charging what the traffic will bear? Do you take it that it means to charge the highest possible price for each commodity?

15. In the case illustrated in Fig. 26 • 2 the railroad would be apt to charge less for the haul from St. Louis to Louisville than for the haul from Chicago to Louisville. How far could they reduce their prices on the St. Louis to Louisville haul rather than go out of business on that branch and leave the traffic to the steamboat companies?

16. Using Fig. 26 • 3, with the costs of transportation just as they are given there, assume that there is a commodity that can be produced in the various cities at the costs given in the accompanying table.

City	Cost per unit, cents	City	Cost per unit, cents
Hanover	90	Providence	20
Williamstown	60	New London	8
Amherst	46	Middletown	9
Worcester	45	New Haven	5
Storrs	18		

What will be the price of the commodity in each city? What will be the source of supply for each city?

17. Work out the results for question 16 if the costs of transportation are just a quarter those shown in Fig. 26 • 3.

18. From the figures of Sec. 26 • 19, what percentage were the railroads earning on their total capital in 1944?

19. Note that in Fig. 26 • 5, covering both freight and passenger traffic, the railroads account for over 75 per cent of the total; yet, when freight and passenger traffic are separated in Figs. 26 • 6 and 26 • 7, the railroads account for less than 75 per cent in both cases? How is it possible to provide over 75 per cent of passenger and freight service together and still provide but 70 per cent of the freight and a third of the passenger service?

27. PROBLEMS OF ORGANIZED LABOR

27 · 1. Special Character of Organized Labor. Basic problems of the labor supply or human resources of the nation were discussed in Chap. 12 and, if we use the term "labor" in its broadest sense, we can say that we have already covered the major labor problems. In recent years there has been a tendency to restrict the meaning of this phrase and to use it loosely to refer to those problems which arise in connection with the organization of laborers into unions of various sorts. The student should guard himself against the notion that the problems of organized labor are the only labor problems, or even the most important ones, yet they are certainly some of the most difficult, troublesome, and urgent problems faced by society today—problems crying out for some solution that recognizes the interests and advantage of all parties involved.

It may seem peculiar to the student at first sight for us to discuss these problems of organized labor in conjunction with our discussion, in the two preceding chapters, of monopoly and of public utilities. Yet the problems are to a considerable extent analagous. The details of historical background are, of course, different; but the principles that lie beneath are for the most part identical. When workers combine in unions and bargain collectively, they are forming at least a partial monopoly. Their major purpose is that of eliminating competition among the individual workers. To be sure, there is no element of eminent domain; but, just as the public utility receives unusual privileges from the government in the form of eminent domain, forcing third parties to give up rights for the benefit of the public utility and of society as a whole, so the union receives from the government other sorts of unusual privileges which force third parties to give up rights for the benefit of the union and of society as a whole. Recent years have seen the development of a whole new body of law, putting labor organizations in a special position and allowing them certain activities that would be prohibited to other groups.

27·2. Historical Background. Although, as the preceding sentence suggests, the status of labor organizations has changed rapidly and radically in recent years, nevertheless they are by no means new. They have existed in some form or other ever since some men have worked for others; ever since the wage system came into being. As long as independent men worked for themselves, or as long as those men who worked for others held the status of slaves, there was no organization of labor as such, because the problems in connection with which labor organizes had not arisen. When men started to work for other men for hire, there came naturally to be some division of interest between employer and employee. This divergence of interest can, of course, be exaggerated. In most ways the interests of employer and employee coincide, as the two cooperate in the production of a common product in competition with other products on the market. Employers compete among themselves, and workers compete with each other for jobs. There is competition between employers and between employees, which is in many ways more significant and fundamental than any competition between the employer and the employee. Yet a number of men working for the same employer have certain common problems, and it is natural for them to unite or cooperate in facing them. Such cooperation is as old as the wage system itself.

What we might call "modern" labor organization goes back only to about the time of the Industrial Revolution a century and a half or two centuries ago, starting, as did the Industrial Revolution itself, in England. At first these early English unions were held to be illegal, and strong efforts were made to eradicate them. Since 1824, however, the English law has permitted their formation, they have become rapidly more and more important.

In the United States the first small local labor organizations arose in the skilled trades in the New England towns—organizations of tailors, of weavers, and of shoemakers. They were short-lived for the most part as individual societies, but as one disappeared others sprang up. Not until almost the time of the Civil War did more permanent and widespread groups begin to form. From that time on we have seen a slow and much-interrupted general movement of growth in the organization of labor. During the last half of the nineteenth and the first quarter of the twentieth century, the growth was closely allied with the changes of the business cycle. During periods of prosperity, when labor was in great demand, the laborers were able to strengthen their organizations and membership grew rapidly; in periods of depression, when workers were laid off and men competed desperately for any work at all, membership in labor unions tended to fall. Yet, despite these interruptions, the basic tendency was for continued growth of the movement; and since perhaps the time of the first World War it has seemed to be true that the rate of growth has depended more

on the attitude of the government than on the stage of the business cycle.

27 · 3. Union Structure. Unions are by no means all alike. Each is organized in accordance with its own individual plans, and each displays its own peculiarities and idiosyncracies. Yet there are similarities enough among them so that we have little difficulty in most cases in assigning any given union to one or another of a few major groups or classes, which differ primarily in the criteria by which they judge eligibility for membership in the union. First, we have the *craft union* or the *trade-union*. Here men are eligible for membership according to the trade or craft they follow. All members do the same general kind of work and are eligible to membership because they do that kind of work. Thus, for example, a carpenters' union would be a craft or trade-union, and all its members would be carpenters. The trade-unions are, by their nature, pretty much confined to those workers who have learned trades—to the skilled or at least the semiskilled. The totally unskilled worker has no particular trade or craft and is not eligible to membership in any craft union.

Second, we have the *industrial union*. Any worker employed in a given industry is eligible for membership in his industrial union regardless of the type of work he does. An industrial union can and usually does embrace workers with many different skills and trades. Thus, for example, a union of miners would admit to membership anyone who worked in the mining industry, whether he was a carpenter, electrician, secretary, or an unskilled laborer. In the United States the early unions ordinarily developed on the craft basis, which meant that unskilled workers were almost entirely unorganized. The past two decades have seen a relatively much faster growth of industrial unions.

Third, we have the *labor union*, which seeks to unite all laboring people regardless of trade or craft or industry. Such unions are usually very "class-conscious," and are relatively unimportant in the current American scene. The Knights of Labor, which dominated the American labor movement for roughly the decade from 1877 to 1887, was an organization of this kind, as was the International Workers of the World which had a sudden but short prominence at the time of the first World War.

Finally, there is the type of organization known as the *company union*. Only those people who work for a particular company are eligible for membership in a company union. Thus, for example, if membership in a union were limited to people who work for the Eastman Kodak Company, the union would be a company union. Company unions have been in rather bad odor with those people who are enthusiastic advocates of the organization of laborers, because many of them have been inspired by, organized by, and controlled by employers. The executives of a company, fearing that an outside union would come in and organize their workers, took the initiative

in forming a union of their own workers first, often offering the use of company facilities for the holding of meetings, etc. Members of other types of unions have felt that the company union was a weapon used by the employer to forestall the formation of "honest-to-goodness" unions and that the company unions have been dominated by the employers for their own interests rather than by the employees. Although many company unions have been of this type, there is no inherent reason that this need be true. It is quite possible that the workers of a given company should organize a union of their own, entirely independent of the management, and such a union would under our definition be a company union. It is unfortunate that many students of the subject have been blinded to the possible social advantages of unionism on a company basis by the unpopularity of those company unions which were stalking-horses for management. The phrase "company union" makes many a man "see red," not because the company is an improper unit for labor organization but because many company unions were not bona fide organizations of employees at all.

27 · 4. Federation. Both craft and industrial unions are organized for the most part as voluntary[1] unincorporated associations. The early ones were started on a local basis, with, say, separate carpenters' unions in various cities. These distinct local unions tended to come together to form national organizations; in recent years, the tendency has been to form the national organizations first and the locals afterward. Thus, a typical national (or, as they are often called because of membership in Canada, "international") union employs organizers whose business it is to organize local "chapters" of the national union. Thus, the typical modern union is a large organization, embracing members scattered all across the country, or at least in most parts of the country in which the craft or industry in question is of importance.

These national or international unions, in turn, find many common problems, especially in the field of legislation. As a result, they tend to band together into more or less loosely knit federations. The older of the two important federations now existing is the American Federation of Labor, or AFL. This is not a union, nor an organization of working men at all. It is, rather, an organization of unions. For the most part the member unions of the AFL are craft unions, although there are some important exceptions. Each of these constituent national unions retains its own autonomy and individuality, its own officers, and its own rules and regulations. The Federa-

[1] A voluntary association is one that one joins without any legal obligation, of one's own free will. There is serious question whether some unions, operating under closed-shop contracts and legally enforced agreements under the terms of which an agency bargains for all employees of a company whether as individuals they approve or not, can properly be called "voluntary" associations.

tion itself, formed in 1881, was reorganized on substantially its present basis in 1886, and today its member unions enroll perhaps 7 million members.[1] The Federation has no authority over its constituent unions, calls no strikes, and negotiates no collective bargaining agreements. It acts primarily as a lobbying organization in connection with proposed "labor legislation,"[2] as a publicity agency and a molder of public opinion.

For many years the AFL was the only important union or federation of labor groups. But changes in industry discussed elsewhere were slowly altering the situation of working people. Industrial establishments were growing larger, and with the change in size new methods of production were being developed which made the skilled craftsman relatively less and less important. On the one hand, were developing more professionalized skills such as those of the engineer; on the other hand, the technique of the assembly line and other mass production methods were tending to replace the skilled craftsman with semiskilled or even untrained workers. The AFL had been based from the start primarily on craft unions, and it continued blindly to insist that the craft union was the only proper form of labor organization. In some areas, especially in the building trades, these craft unions had been eminently successful, but even some of the most powerful of the craft unions (notably the Railway Brotherhoods) were not affiliated with the AFL, and such noncraft groups as the "white-collar workers" and the unskilled laborers were hardly touched by unionism at all.

Many stanch advocates of organized labor felt that the AFL was making a mistake to insist on the craft form of organization, and strong efforts were made within the Federation for a shift toward the industrial form. These efforts culminated in the convention of 1935, where the advocates of industrial unionism were defeated by roughly a 5 to 3 vote. Leaders

[1] Accurate figures on union membership are very difficult to get, and exaggerated claims are commonly made. The figures given here on membership must be thought of as only rough approximations.

[2] The phrase "labor legislation" is generally used today to mean that legislation which is sponsored by organized labor. Much legislation that has to do with the broader labor problems discussed in Chap. 12 is not commonly called "labor legislation," and there have been numerous examples of legislation sponsored by organized labor groups and called "labor legislation" which has been for the advantage of special organized groups of workers and to the disadvantage of working people as a whole. The advocate of unionism is apt, of course, to assume that he speaks for all labor, and to try to get other people to accept the point of view that his program is sponsored by and to the advantage of all working people. The student should be careful, however, to distinguish clearly in his own mind the idea of labor in general from that of organized labor in particular. Sometimes the two are in harmony, sometimes not; at any rate the two are by no means identical, and the student who confuses them is certain to be led into slipshod thinking.

of the movement formed an organization of their own, the Committee for Industrial Organization, which worked for a short time within the Federation. The Federation, however, would not countenance what seemed a challenge to its authority and leadership, and the unions that constituted the Committee were at first suspended and finally expelled. In 1938, these expelled unions formed a permanent organization, the Congress of Industrial Organizations, or CIO. The constituent unions of the CIO today probably enroll about 6 million members, or somewhat fewer than the AFL. The present picture of union strength in the United States is, then, roughly as follows:

	Millions
Total labor force	50
Organized wage earners	14
AFL	7
CIO	6
Members of unaffiliated unions	1
Unorganized persons	36

This picture is apt, however, to be misleading, since the total labor force on which it is based contains physicians, lawyers, business managers, corporation executives, and many others who are not usually thought of as potential union material. If we exclude professional people and the proprietors and managers of businesses, the picture is approximately as shown in the accompanying table. These latter figures can be most easily compared

	Millions	Percentage
Total wage earners	40	100
Organized wage earners	14	35
AFL	7	17½
CIO	6	15
Unaffiliated unions	1	2½
Unorganized wage earners	26	65

by reference to Fig. 27 · 1. Yet this figure and other general statements about the extent of unionism are apt to be misinterpreted unless one keeps in mind the fact that union organization is by no means uniformly spread throughout the economic system. Some industries are almost completely unionized, while others are practically untouched. The unions have also been far more successful up to date in some parts of the country than in others, especially in the industrialized Northeast and along the Great Lakes, and relatively unsuccessful in the deep South. Many states of the Middle West are so predominately agricultural that unionism has been slow to

gain a foothold, although some attempts have been made to organize farm workers in recent years.

The CIO is much more centralized in authority than is the AFL. In the Federation, the member unions hold the major power, while in the CIO the overhead organization is itself the powerful body. The CIO has tended more strongly than has the AFL to enter the political arena and has adopted new and aggressive and effective tactics for organizing industries that had long opposed unionization. It is perhaps too easy, however, to play up the differences between the two opposing groups. In practice, the differences are those of degree rather than of kind, and the two organizations have co-

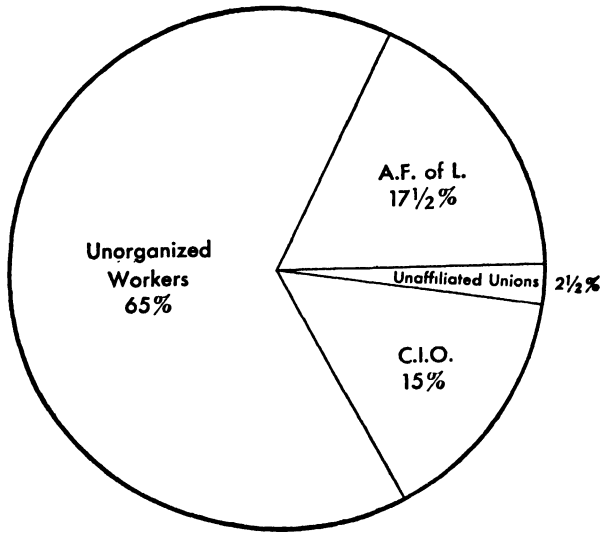


FIG. 27 • 1.—Approximate extent of unionization in the United States, 1946.

operated on many legislative programs. Yet the fact remains that, for the time being, the leadership of the labor movement in the United States is divided. Both proponents and opponents of organized labor differ in opinions as to whether the split has been advantageous or disadvantageous. Certain it is that the rivalry of the two groups has led to more energetic efforts to unionize workers, and that the rivalry has probably been a factor in the fact that the number of organized workers is at an all-time high.

27·5. Aims of Organized Labor. Just as it is often suspected that some politicians are in politics for what they can get out of it personally, and just as it is sometimes suggested that the militarists have provoked wars to show their own indispensability, so it is often argued that the leaders and officers of organized labor groups are motivated primarily by selfish considerations—that they are more interested in their jobs, their salaries, and

their own advantage than they are in any abstract ideas of human justice. Yet it would be unrealistic to suppose that the entire union movement is based on nothing more substantial than the exploitation of the great mass of workers by their self-styled "leaders," just as it would be unrealistic to suppose that all these leaders are actuated by high and unselfish motives. If we are to explain the tremendous growth of unions in modern times, we must look a great deal deeper, to the structure of the economy itself.

A contest ceases to be a contest when the participants differ materially in strength. A boxing match between an ordinary college man and a six-months-old baby is not a contest but a slaughter. The two cannot properly be said to compete at all. As industry waxed larger and larger and as corporations developed to the point where they represented millions or billions of dollars of investment and employed thousands of workers, these workers felt with some justice that they were not on an equal footing with their employers when it came to bargaining for wages, hours of work, or working conditions. If we consider the individual worker employed by, say, the United States Steel Corporation, he feels that he is insignificant in the plans of his employer, while his employer's plans are vital to him. If he asks for an increase in wages, the employer can dismiss him and never notice the difference; yet the job, which is of such minor importance to the employer, is absolutely essential to the worker—a life and death matter. Thus, the unorganized lone worker finds, in practice, that he is forced to "take it or leave it"—at least, so the argument runs.

To be sure, this analysis may be to some extent superficial. If there is a reasonable degree of competition among employers and if one employer is paying a worker substantially less than he is worth, another employer will find it to his advantage to hire the worker away at higher wages. If the worker is really worth $2 a day and is receiving $1 some competing employer will find it advantageous to offer him $1.50. Under such circumstances, the new employer will receive $2 worth of work for $1.50. The worker thus has some protection against his particular employer if there is competition among employers. In the first place, such competition is never perfect; in the second place, even where such competition exists in a considerable degree, the worker is not aware of it—he does not see it—he still feels impotent and powerless before the employer. As we have noted elsewhere, men's behavior is dictated by their feelings and beliefs rather than by the facts and, even if they had complete protection against their employers by means of competition, they would still desire group action if they thought that they were weak and the employer was strong. In fact it must be admitted that competition among employers is by no means perfect and that the protection that workers get from this source is only partial. Yet workers often overlook the fact that there is any protection from this source at all and

talk as though the unorganized worker were entirely at the mercy of the employer, who might, in turn, set wages anywhere he liked.

Because the individual worker, trying to strike a bargain for wages with his employer, feels that he is too small to have any importance, it is natural that workers should band together to bargain collectively. *Collective bargaining,* as the words themselves imply, is merely bargaining carried on by a group acting together rather than by unorganized individuals. If each employee makes his own arrangements with the employer as to wages, hours of work, etc. (usually in such a case accepting without question what is offered him or else refusing the job entirely), we would have a case of individual bargaining; but, if the workers band together and form an organization and if a representative of that organization bargains with the employer for all the workers, then we have a case of collective bargaining. The United States Steel Corporation, bargaining with an individual worker, might not be at all upset if he refused the company's offer and decided to work elsewhere. After all, what is one worker among so many? If the workers bargain as a group and if, dissatisfied with the company's offer, they decide as a group to withdraw their labor *in toto,* the company is bound to show an active interest. The company management, representing all the investors and bargaining with the union agents, representing all the workers, finds that it is now matched on something like equal bargaining terms. The two participants in the process are of roughly equal power, rather than being matched like the college boxer and the six-months-old infant.

In the early days of unionism, collective action by workers was held to be combination in restraint of trade on the same basis as collective action by manufacturers in the sale of goods. Court decisions were by no means uniform, and the exact status of collective bargaining was not clear. The tendency was, by both court interpretation and amendment of the statutes, to bring collective bargaining more and more clearly within the pale, exempting it from the provisions of the antitrust acts. The ill-starred National Industrial Recovery Act tried by what turned out to be unconstitutional means to safeguard the right of collective bargaining; but the Wagner-Connely National Labor Relations Act of 1935 reaffirmed the right and strengthened it in pretty much its present form. Section 7 of this act states:

> Sec. 7. Employees shall have the right to self-organization, to form, join, or assist labor organizations, to bargain collectively through representatives of their own choosing, and to engage in concerted activities, for the purpose of collective bargaining or other mutual aid or protection.

Other sections of the act state that those who are chosen as bargaining agents by a majority of the employees shall represent all the employees in

collective bargaining, that employers must not refuse to bargain collectively with their employees, that employers must neither encourage nor discourage union membership on the part of their employees, that employers must not interfere with employees in the exercise of the rights of collective bargaining, and that employers may not dominate or interfere with the formation or administration of unions. The act established a National Labor Relations Board to supervise its operation.

Collective bargaining is by no means the sole function of labor unions, even if it is the most important one. Unions have offered life insurance to their members, have furnished protection against old age, sickness, and unemployment, and have at times taken an active interest in politics. But their major function has been that of organizing workers in order to present a unified front in collective bargaining.

Of course, it takes two to make a bargain, and unions used to be met by the refusal of the employer even to discuss proposals. This problem of obtaining *union recognition* has historically been a difficult one, and in many cases could not be solved at all prior to the legislation just mentioned, which requires employers to bargain collectively with their employees through representatives that the employees themselves select. Present legislation allows either individual or collective bargaining but gives the preference to the latter. Whereas union members used to find that their union could not obtain recognition and that they were forced against their will to deal individually with their employers, now nonunion members often discover that they cannot bargain individually even if they wish but must be represented by a union.

Having organized and having been recognized the union next negotiates an agreement or contract stipulating the wages that will be paid for various kinds of work, the hours of work, the conditions of work (safety appliances, methods of hiring and firing, etc.), and possibly other things. These other provisions include many things that are incidental—policies that seem important to the union as guaranteeing its own existence or the effectiveness of its bargaining power. These provisions are not really ends in themselves but are means to ends, and must be so considered if they are to be justified at all. We turn now to a consideration of some of them.

27·6. The Closed Shop. Most unions have attempted to obtain agreements under which employers bind themselves not to employ men who are not members of the union. A company that refuses to hire nonunion men is said to operate a *closed shop*—an establishment closed to nonunion men—as contrasted with the *open shop* which is open to all workers regardless of their union affiliation or lack of it. The open shop has been called by those employers who advocate it "the American plan."[1] Actually there is

[1] See footnote, p. 653.

no such simple distinction as that implied by the division of businesses into those which operate open and those which operate closed shops. Other classifications are necessary. There are "preferential shops" which hire both union and nonunion labor but give a preference to union labor. And there are shops, which their operators insist on calling "open shops," that really refuse to employ union members at all. They are closed to union men just as the usual closed shop is closed to nonunion men. This latter type of shop is illegal in industries engaged in interstate and foreign commerce, and also in intrastate business in many states, as a result of legislation that insists that it is an unfair labor practice for employers to encourage or discourage membership in any union by discrimination in regard to hire or tenure of employment. Such a provision appears to prohibit the ordinary closed shop just as it prohibits the shop closed to union members; actually, special provision is made for the former in the statute, as can be seen by referring to the actual wording in Sec. 27 • 17. We shall use the term "closed shop" in the sense that it has come to have from long usage—a shop in which union membership is a prerequisite to employment.

No one attempts to justify the closed shop on any ground save with the argument that unions are desirable and that they could not exist without it. Those who object to unions, or those who believe that unions can operate efficiently without the closed shop, object to the interference with the rights of nonunion men who are forced either to join the union against their will or to give up work in the closed industry. Union men, on the other hand, argue that if nonunion men can get all the advantages of higher wages, better hours, etc., that are won by the union, without having to pay the dues required of members, there will be no incentive to join the union. On the contrary, there will be an incentive to stay out of it, and the union will be weakened. The union man argues that the workers should all share in the costs and the responsibilities since they all share in the benefits. It is easy to understand why union men would like to force others to join if they cannot persuade them to do so voluntarily; naturally, society is loath to interfere with the worker's freedom of choice in the matter unless the exercise of such freedom by individuals brings decided social disadvantage. The legislation of the past decade or so has leaned in the direction of encouraging and even almost requiring union membership of workers who wish to be employed in occupations that are largely unionized. This has been a consequence of the feeling on the part of the lawmakers that unionism is so important that it must be encouraged by the government even at the expense of the loss of some individual freedom. Such sacrifice of personal freedom for the benefit of society is common, of course, and by no means confined to the labor field.

The desirability of the closed shop may be listed as open to argument, although many economists today approve it as a means to a desirable end.

But, when the closed shop is combined with a closed union, we have quite a different story. A *closed union* is one that does not allow qualified workers to join it without restriction. In extreme cases, such a union might refuse to admit any new members whatever; in other cases, it might restrict the number sharply either by a rule that forbade the initiation of more than a small number of new members or by setting such a high admission fee that in fact almost no one could join. Some unions have set arbitrary restrictions on the number of apprentices who are allowed to learn the trade, thus keeping the number of qualified applicants under control. If a given industry has accepted the closed shop, so that nonmembers cannot work, and if in addition the union refuses to let these nonmembers join the union, then the union has established such a tight monopoly control of the labor market in the industry as cannot well be countenanced. Just as we said in the case of public utilities, where an organization is given great power, society must make sure that the power is not used for individual advantage to the public detriment.

27 · 7. Restriction of Output. One of the most widely criticized of the policies followed by some unions is that of limiting the amount of work that its members may do. This is done in many ways and, as we have already learned, is a policy followed as often by employers as by employees. We have discovered that limitation of output is a phenomenon that commonly appears in conjunction with monopoly control, being under some circumstances advantageous to the monopolist, although never advantageous to one who produces under competitive conditions. Limitation of output, whether practiced by labor unions or by employers or whether induced by short-sighted government policies, is, under ordinary circumstances, detrimental to the general public. The individual gains at the expense of his fellows. The monopolist who reduces the output of a patented article in order to increase his net income harms the public for his own benefit. The labor union that sets up arbitrary restrictions on output in order to raise wages or increase employment helps its own members at the expense of the public. To be sure, the laborer who benefits is himself part of the public, and it is sometimes argued that there is no general loss but merely a transfer from one member of the public to another. The same thing might be argued with regard to the monopolistic producer of the patented article who enriches himself at the expense of others, or of the pickpocket who merely transfers wealth from one member of the society to another. In a real sense the pickpocket does the least harm of the three, because in his case there is actually a transfer of wealth from one member of society to another, while in the other two cases the transfer is but part of the picture. In addition, we must remember, that monopoly, which always tends to restrict output, leaves us with a smaller quantity of goods in the society than we would otherwise have had. There is not a mere transfer, but a reduction in the total

wealth available, as there would be if each time a footpad picked a pocket half the loot vanished into thin air, so that the pickpocket had less gain than the loss that he inflicted on his victim.

Restriction of output by unions has taken many diverse forms. Sometimes it has consisted of "soldiering on the job," "killing time," "taking it easy," or "ca'canny." Sometimes it has consisted in shortening the hours of work.[1] Sometimes unions have adopted rules forbidding the employer to use methods or equipment that would increase output. Bricklayers have limited the number of bricks to be laid in a day; miners have limited the amount of coal to be taken out of the mine by a man in a day; painters have required the use of small brushes in order to increase the time of doing the job; motion-picture operators, musicians, and railroad employees have adopted rules requiring their employers to use more men than really necessary to do a particular piece of work.

Union men have sometimes sought to justify restrictive policies on the basis of health. There can be no question but that some of the restrictions can be justified on that basis. There is, of course, some reason to argue that, if the measures are necessary for health, they should be established and enforced by the public in general rather than by those parties who have a direct interest and who might profit by further restriction than health measures require. But the answer usually given to this argument, and again probably in many cases a valid one, is that the general public is not interested enough in the problem and that, if we were to wait for public clamor to force those restrictions which really have a health basis, we would never get them. The student should remember, however, that many measures instituted for private advantage have been disguised as "health measures" and that manufacturers and merchants have been just as quick to use this disguise for their private programs as have union strategists.

If we admit the validity of the health argument for some cases of restriction, we find that others are "justified" on the basis of the argument that there is just so much work to be done and that if it is done too quickly there will be unemployment. The idea is that the worker must slow down production in order that the job may last and that there may be work enough to go around. Economists have long referred to this doctrine as the "lump-of-work" argument. From the standpoint of the immediate, temporary, and short-run advantage of the individual employee the argument may have

[1] Two comments are necessary here: (1) Reducing the hours of work does not by any means invariably reduce the amount of output. The student will realize at once, if we take an extreme case, that a typical worker will be more productive year in and year out if he works 8 hours a day than if he works 23½ hours a day. (2) There may be gains, both to the worker and to society, which more than offset any reduction in output. This may be true, of course, no matter how the output is restricted, but it is obvious in the case of limitation of hours of work.

some validity; from the standpoint of society as a whole or from any standpoint in the long run, the policy is stupid and harmful. The purpose of our economic system is not to make work, but almost the opposite: to get as large a quantity of the good things of life as possible with the *least* possible work. Soldiering on the job reduces the quantity of good things available and makes the people of the society poorer than they otherwise would be. Restriction of output of desirable things[1] is an inevitable practice of monopoly and, unless specifically offset by some such counterbalancing gain as improvements in health mentioned above, is invariably of disadvantage to society as a whole.

27·8. Opposition to Laborsaving Machinery. For very much the same reason that some unions have practiced the slowdown, some unions have opposed the introduction of new types of laborsaving machinery. Such opposition is based on the same sort of lump-of-work reasoning which leads a man to take twice as long on a job as necessary. The introduction of electric drills in coal mining met bitter union opposition. The use of machines in the making of cigars or glass bottles was likewise opposed; and there have been many other examples. The answer to such opposition has been given very effectively by Henry Ford thus:[2]

> For when were men ever really put out of work by the bettering of industrial processes? The stage-coach drivers lost their jobs with the coming of the railways. Should we have prohibited the railways and kept the stage-coach drivers? Were there more men working with the stage coaches than are working on the railways? Should we have prevented the taxicab because its coming took the bread out of the mouths of the horse-cab drivers? How does the number of taxi-cabs compare with the number of horse-cabs when the latter were in their prime? The coming of shoe machinery closed most of the shops of those who made shoes by hand. When shoes were made by hand, only the very well-to-do could own more than a single pair of shoes, and most working people went barefooted in summer. Now, hardly anyone has only one pair of shoes, and shoemaking is a great industry. No, every time you can so arrange that one man will do the work of two, you so add to the wealth of

[1] It may be wise for a society to restrict the output of undesirable things—smallpox germs and adulterated or harmful drugs. But the hitch comes in those cases where there is disagreement as to the desirability of things. Some people would feel that the world would be better off if there were less tobacco or liquor or lipstick; others would not. Here we get back again to the whole question of human wants and the evaluation of goods. For our purposes here, however, we can dodge that issue with the statement that union members are engaged in making things that some people do want, and if they soldier on the job there will be a reduction in the output of what some people, at least, think is desirable. Fewer human wants will be satisfied.

[2] FORD, HENRY, and SAMUEL CROWTHER, "My Life and Work," pp. 153-154, Doubleday & Company, Inc., 1922.

> the country that there will be a new and better job for the man who is displaced. If whole industries changed overnight then disposing of the surplus men would be a problem, but these changes do not occur as rapidly as that. They come gradually. In our own experience a new place always opens for a man as soon as better processes have taken his old job. And what happens in my shops happens everywhere in industry. There are many times more men today employed in the steel industries than there were in the days when every operation was by hand. It has to be so. It always is so, and always will be so. And if any man cannot see it, it is because he will not look beyond his own nose.

Mr. Ford's statement of the case may be too sweeping in overlooking the temporary unemployment that does take place when men are displaced by machinery. Such unemployment we call "technological unemployment" as distinct from the "cyclical unemployment" that accompanies the swings of the business cycle.[1] When a man has spent years in developing a particular skill and when he suddenly discovers that his work can now be done faster, better, or more cheaply by machine and that he must look for another type of work, he finds that he is no longer a skilled laborer. The type of skill that he has developed is no longer needed, and he either has to take an unskilled job or go back to spend more time and energy and money in developing new types of skill. The change is a difficult one, and it is natural that the worker should fight against it. If society is to get the benefit of the new machine, it is also natural that the displaced worker should feel that society owes him something. Society is benefiting at his cost. The answer is not to refuse to use the better methods, and Mr. Ford is entirely correct in pointing out that in the long run all society benefits from the change. In the meantime we should remember that what union opposition there is in such cases is twin brother to that shown by the corner independent grocer to the new chain store, or by any other vested interest to changes that threaten to reduce the value of the investment. The worker's own training is in most cases his most important and valuable investment. Opposition to social progress by particular groups that stand to lose by the progress can be easily understood even if it cannot be condoned.

27 · 9. Industrial Conflict: The Employer's Weapons. One cannot expect, even under the best of conditions, that union and management representatives will always agree. Of course, if the management refuses even to bargain, there is disagreement from the start; even where there is an honest attempt to discuss issues and reach an agreement, agreement may not be forthcoming. Where there is an active dispute between management and the union, each side has developed various tactics and procedures which experience has shown to be more or less useful in settling the issue. Let us turn our attention

[1] See Chap. 22.

first to some of the more common procedures used under such circumstances by the representatives of management.

One of the commonest of these procedures is the *black list.* This is merely a list of the names of persons whom the management refuses for some reason to hire. Its more questionable features arise when the blacklist is passed around from one employer to another, and when all employers refuse to hire the men whose names are listed. Altogether too often the sole reason for black-listing a man has been the fact that he was an active union member. In fact, the term as used by union men refers to a list of names of persons who are active union members, circulated among employers with the purpose of making their employment difficult or impossible. Black lists are difficult to discover, because they are usually circulated secretly. Today they are seldom actual printed or written lists but are passed along by word of mouth or otherwise. Under common law, the black list was permissible, but many states passed statutes intended to curb or eliminate it, and the present National Labor Relations Act, in Sec. 8, declares that it is an unfair labor practice for any employer to encourage or discourage membership in any union by discrimination in regard to hire or tenure of employment. In the first half-dozen years of experience under this act, five out of every eight complaints filed with the National Labor Relations Board alleged discrimination in employment.[1] Black-listing is probably not so common today as it was prior to the passage of the act, but the ease with which it can be carried on *sub rosa* and the difficulty involved in proving why an employer refused employment to an applicant make it seem unlikely that it can be entirely eliminated.

Another weapon used by some employers has been the *injunction,* which is an order, issued by a court, ordering the enjoined persons to refrain from the performance of certain acts. One who feels that others are threatening to harm him can, of course, wait for the harm to occur and then sue for damages. In some cases, the court may feel that no repayment could compensate for the damage and that there is ground for stepping in in advance to prohibit the threatened action. If, now, the enjoined person disobeys the injunction and performs the act that was specifically forbidden by the court, he is guilty of contempt of court and liable to fine or imprisonment imposed by the court without trial by jury. To be specific, let us take a simple case. If my neighbor threatens to break my neck next time he sees me and if I fear that he is in earnest in the threat, I may apply to the court for protection. It would be futile for the court to suggest that I wait to see how it turns out and that, if he carries out his threat, I may sue him for damages. The damage, once committed, is irreparable, and no suit for damages

[1] MILLIS, HARRY A., and ROYAL E. MONTGOMERY, "The Economics of Labor," vol. III, Organized Labor, p. 595, McGraw-Hill Book Company, Inc., New York, 1945.

can compensate me. Therefore, if the court feels that the injury is really apt to occur, it may issue an order enjoining my neighbor from breaking my neck. If my neighbor now does break my neck, he has committed two offenses instead of one. He has committed assault or murder that he would have committed in the absence of any injunction, but he has also disobeyed an explicit order of the court and has thus committed contempt of court. Although he would have the right of jury trial for the first offense, cases of contempt of court are tried by the court itself without jury. The man who originally made the order—the offended person—acts as both judge and jury in disposing of the case. This aspect of injunction procedure has roused the ire of union men in those cases where injunctions have been used in labor disputes.

Injunctions have been used most commonly in cases where there is violence or threatened violence in labor disputes, as in cases of mass picketing. If irreparable damage is actually threatened, there can be little objection to their use. But there is strong and valid objection to their use for the wholesale prohibition of union activities of a general nature or in cases where the damage from harmful activities could be otherwise appraised and recompensed. Union leaders have felt, rightly or wrongly, that the courts have tended to side with the employer, and to issue injunctions indiscriminately. There is very bitter feeling against their use, and a feeling of heroic bravado on the part of many people who break them. Many states have passed laws restricting the use of injunctions in labor disputes, and the Norris-LaGuardia Act of 1932 specifically prohibits the issuance of injunctions in labor disputes under circumstances specified in the act. Their use in connection with "yellow-dog" contracts is described in the following paragraph.

A third weapon sometimes used by employers has been the refusal to hire workers unless they agreed not to join the union. A contract in which the applicant for work agrees, as a condition of employment, not to join a union, was called by union men a *yellow-dog contract.* In earlier days their use was by no means uncommon. If a particular concern used such contracts as a regular part of its employment policy, every worker employed by it would be a party to such a contract. If, now, a union organizer appeared on the scene and tried to persuade employees to join the union, the employer could complain that the organizer was attempting to persuade the workers to break their contract. The employer would then appeal to a court, arguing that an outside party was trying to get workers to break their contracts, and in many cases the court issued injunctions prohibiting the organizing activity. The union organizer, as explained in preceding paragraphs, was then open to prosecution for contempt of court unless he gave up his activities. This use of the injunction was particularly abhorrent to union men, and their violent dislike of the procedure led them to call such contracts

"yellow-dog contracts."[1] Although there have been many important court decisions on this matter in the past, the matter is one that we can pass over rather quickly nowadays since changes in the law have made such contracts uncommon. The Wagner-Connely National Labor Relations Act of 1935, as we have said before, lists as an unfair practice the discouraging of membership in a labor organization by discrimination with regard to employment or tenure. Several of the more important industrial states have adopted similar laws. Thus the yellow-dog contract is now illegal in those industries which participate in interstate and foreign commerce, and also in intrastate industries in those states which have such laws. A fourth weapon used by employers in labor disputes is the *lockout*. This is a refusal on the part of the employer to allow his employees to work. They are turned out of the plant and the gates closed. Business stops, and the workers' pay stops. The hope of the employer is that the lack of income will force the men to accept the employer's terms. Lockouts have, in practice, been rather uncommon. The employer has been much more apt to dismiss the individual leaders with whom he is disputing than to close down his factory entirely.

A fifth weapon used by employers in labor disputes is the employment of *strikebreakers* to take the places of those men who are out on strike. These strikebreakers are commonly referred to by union men as "scabs." They may be nonunion men who are looking for work and who are willing to take over the jobs vacated by the strikers. On the other hand, they may be professional hoodlums who have no intention of staying on the job but who take temporary employment to aid the employer in breaking the strike. Legally, it makes no difference which is true. The worker who is out on strike always thinks of the strikebreaker as one who is robbing him of his living; as a result, the use of strikebreakers has commonly been accompanied by violence. In case of violence, the local police have commonly sided with the strikebreakers, as they probably are forced to do, since the strikebreaker, after all, is well within his legal rights in seeking employment, and the striker who tries to interfere with him is the perpetrator of the violence. The issues, however, are deeper than this and extremely difficult to appraise. The plain fact is that both employers and employees have often regrettably resorted to violence in labor disputes, but in most cases the legal situation has been such that the employer has had no reason to use violence, and the strikers have been the ones to instigate it. This has naturally prejudiced the

[1] The use of such terms of opprobrium, both by employers and by unions, has been very common. Unthinking people are doubtless influenced into thinking that a given concept is desirable or undesirable according to its name. Thus union men talk of "yellow-dog contracts" and of "scabs," while employers talk of the "closed shop" and of the "American plan" (the latter being their name for the open shop). The student should try not to be biased by nomenclature which is often chosen carefully with the hope that it will bias him. On the contrary, he should think in terms of the issues involved and their social significance.

general public against the workers in such cases, even though to the worker such prejudice has often seemed unfair. It must be said, of course, that society cannot countenance the use of force and violence by any party to a labor dispute. The parties concerned must find other ways to accomplish their objectives, and only the shortsighted person excuses such means as being justified by their ends. Similar to the use of strikebreakers in some ways has been the use by some employers of *labor spies* and *private police forces*. Such tactics have usually been legal, but they have always aroused bitter opposition on the part of union men and have frequently been met by violence. The employer who uses strikebreakers or labor spies should not be condemned as acting illegally, but it is quite possible that his judgment should be questioned.

A final weapon used by employers in combatting unions has been the *company union*. This we have already described.[1] Often, when attempts were made by outside organizers to build up union membership in a plant, the management countered by establishing a union of its own for its employees—a union dominated by and actually operated for the benefit of the management. Advocates of independent unionism have always felt that such unions are but a pretense—a poor excuse for the real thing. They have argued, first, that since only employees of the given company are eligible for membership, the union will have no members who can bargain on equal terms with the employer. The representative of the company union, being himself an employee, is afraid that he may lose his job if he argues too heatedly or even too cogently and successfully. For this reason, union men have insisted that they be represented by outsiders of their own choosing—representatives of a national union, coming in from out of town, and not in any sense responsible to the employer with whom they are to bargain. Although there can be no gainsaying the validity of this position, it is only reasonable to state here, what is almost always overlooked in discussions of this question, that a company union can operate without dominance by the employer and it can bargain through outside representatives of its own choosing. There is nothing inherent in company unionism that makes necessary those particular abuses which have in fact been common. Since the passage of the National Labor Relations Act in 1935, many company unions, theretofore dominated by management, have become in fact independent, and many more independent company unions have sprung up. Section 8 of the Act declares it to be an unfair labor practice for the employer to dominate or interfere with the formation or administration of any union, or to contribute financial or other support to it. This is said by some authors to forbid the formation or operation of company unions, whereas it merely forbids domination of such unions by employers. In some industries the

[1] See Sec. 27 • 3.

truly independent company union has become common and important in recent years.[1]

27 · 10. Industrial Conflict: The Union's Weapons. The activities in which the union engages during a labor dispute are more spectacular, and consequently more widely known, than those used by the employer. Most important and most effective has been the *strike*. A strike is an organized, concerted cessation of work with the purpose of gaining certain ends. The employer and the general public have tended to look upon it as a mass resignation, while the workers themselves are thinking of the cessation as purely temporary and are expecting to go back to their jobs as soon as their point is won. They still think of the jobs as "their" jobs, and it is here that many difficulties arise. One Chicago jurist has stated the outsider's attitude plainly in the following words.[2] "What rights do the [strikers] have? By striking they have definitely ended their employment and become strangers to the plant. They have no rights to protect. . . ." The union member would argue categorically that the jurist's statement is incorrect and that strikers have not ended their employment but have suspended it temporarily.

It is common to recognize several types of strikes. Most easily understood is the *direct strike*—one that arises out of a dispute between employees and their own employer. A *sympathetic strike* is one called by a union that has no direct grievance of its own; it is called by one union to help another union. For example, let us suppose that workers in the stone quarries are striking for higher wages. This would be a direct strike, arising from a dispute between the quarry workers and their employers. But suppose, now, that the stone from the quarry is shipped to a distant city, there to be incorporated into a new building, and suppose that the stonemasons of this city go out on strike, refusing to lay the stone, in an effort to bring pressure to bear on the owners of the distant quarries. The idea is that, if the stone cannot be used in building, it will not be bought; if the quarry owners cannot sell it, they will lose money; and if they lose enough money, they may prefer to come to terms with their own union. The strike of the stonemasons is now a sympathetic strike. They have no grievance against their own employer—no demands of their own which they are making (save the demand that the imported stone shall not be used); but they are trying to assist another union in gaining its ends. The purely sympathetic strike has been pretty generally condemned by the courts as illegal, since it involves material damage to third parties who are not directly involved in the dispute.

A *general strike* is a concerted withdrawal of work in all industries at once. In practice, complete cessation of work has never occurred, and a gen-

[1] For a detailed discussion of recent developments in the company union field see Millis and Montgomery, *op. cit.*, Chap. XV. Much illustrative material is given in this reference.

[2] Quoted in Millis and Montgomery, *op. cit.*, p. 615.

eral strike is thought by most authorities to be impossible; yet it must be realized that even a concerted strike involving several strategic industries, such as the public utilities, transportation, etc., would cripple the economy. If electricity were shut off and if trains stopped running, any modern city would quickly be reduced to chaos. Even a short railroad strike in the United States in 1946 raised social problems of such serious import that emergency action had to be taken, and the strike was quickly broken. The general strike has usually had political, as contrasted with directly economic, motives. It has been the vision of such organizations as the Industrial Workers of the World, and the ordinary union either is not interested in such matters or opposes general strikes as bringing all strikes into opprobrium and disrepute.

A *jurisdictional strike* differs from a direct strike in that it arises from a dispute between unions rather than from a dispute between a union and an employer. If we recall the topics treated in Sec. 27 • 3, we shall realize that craft unions and industrial unions compete for members. A carpenter working in a coal mine is eligible for membership in the carpenters' union (a craft union) and also eligible for membership in the miners' union (an industrial union). The different union organizers will try anxiously to persuade him to join their particular union just as the representatives of various fraternities flock around the star preparatory school athlete during rushing season. Each union wants to get control of a majority of the workers, so that it can be named bargaining agent. The worker cannot well belong to both unions, subject to the jurisdiction of both of them, since their policies might differ. And, of course, the dispute need not involve craft and industrial unions at all. There is often dispute between two craft unions as to which should have jurisdiction over particular work. Should this task be done by a carpenter, a plumber, or an electrician?

In any case, where there is a dispute between unions as to which one has jurisdiction in a particular situation, it is not uncommon for that union which believes that it is stronger to call a strike to enforce its claims. It calls its men off the job in an attempt to halt work and force the employer (who is in this case really an outside party, as far as the dispute is concerned) to "recognize" it as the bargaining agent. Such strikes have particularly tried the patience of employers and the general public, and one of the arguments used by advocates of industrial unionism as opposed to craft unionism is that industrial unionism to some extent reduces jurisdictional difficulties. Even with industrial unions, however, there is the problem as to where the steel industry stops and the automobile industry begins, and jurisdictional disputes between industrial unions are by no means unknown.

In an ordinary strike, the workers involved walk out of the plant and leave their jobs. The employer, as we have seen, usually feels free to replace

the striking workers with strikebreakers if he can get them. To make this impossible, strikers have sometimes stopped work but refused to leave the plant, thus carrying on a *sit-down strike,* or a *stay-in strike.* These were especially popular in the United States in 1937, when there were almost 500 such strikes. Such high-handed tactics on the parts of unions tended to antagonize even those people who were generally sympathetic with union aims and, although a spineless U. S. Secretary of Labor at the time professed to have some question as to the legality or illegality of such proceedings, there has never been any question in the mind of anyone else as to their illegality and undesirability. The U. S. Supreme Court referred to such a strike as a "high-handed proceeding without a shadow of a legal right." In the first case to come before the Federal courts the sit-down technique was held to involve illegal seizure of property and to constitute an act of violence which made it an illegal method of carrying on a strike.

It may be said in general that strikes are not in themselves illegal in the United States, but may become so because of the employment of illegal methods (such as violence), because of the illegality of their purpose, or because there is not sufficient proximity of interest (as in some sympathetic strikes). It should also be said that the strike is still looked at by the courts primarily as a cessation of employment rather than as an interruption of it. The worker is free to resign his job as an individual, and groups of workers may resign in unison; but they then have no legal hold on the job that they have relinquished. The job, according to the law, is thought of as belonging to the employer rather than to the employee and is assigned by the employer to whomever he wishes, subject, of course, to restriction as to discrimination for union activity, etc. As long as the employers, the courts, and a large proportion of the general public think of a job as belonging to the employer while the unions think about a job as belonging to the worker there is bound to be clash of opinion. The student should understand that neither of these points of view is inherently "right"; but that, on the contrary, the problem is one of working out that system of social and legal institutions which promises to yield the greatest long-run social advantage. The problem goes far deeper than merely saying that Mr. Brown does not have to hire Mr. Jones if he does not wish, and that Jones does not have to accept employment from Brown if he does not wish. Such negative freedoms are seldom enough by themselves in social relationships. Men need freedom to do things as well as freedom to refrain from doing things if there is to be substantial accomplishment.

27·11. Picketing. A second weapon used by unions, often in connection with a strike but sometimes when no strike is in progress, is *picketing.* Workers who are out on strike wish still to protect their jobs. They do not wish to allow other men to take over the jobs during their absence. For this

reason they post "pickets" outside the factory to inform those who might otherwise enter that a strike is in progress. Theoretically, the purpose of picketing is to disseminate information—to inform prospective workers that a strike is under way. Actually, a picket line is a method of exerting the pressure of public opinion. The picket urges his fellows not to work during the period of the strike and, unfortunately, often accompanies his urging with threats or abusive language. Persuasion gives way slowly and imperceptibly to intimidation. In cases of "mass picketing," huge crowds of union members and sympathizers throng around the entrances to the plant, pushing and shoving and intimidating workers or others who wish to enter. The courts hold that picketing is legal as long as it is peaceful. As long as it consists merely in informing the public of the facts, it cannot be enjoined without infringing the Fourteenth Amendment to the Constitution protecting freedom of speech. But as soon as the major purpose of picketing ceases to be that of informing and becomes that of intimidating, the picketing becomes illegal. Thus when, in the spring of 1946, a Connecticut union began to picket the homes of plant officials and others, the picketing was held to be illegal on the grounds that it had become mainly an instrument of intimidation. Picketing must be peaceful if the Fourteenth Amendment is to be invoked in its behalf.

In the past, picketing was often prohibited by court injunction. The Norris-LaGuardia Antiinjunction Act of 1932 states, however, that courts shall not issue injunctions that prohibit "giving publicity to the existence of, or the facts involved in, any labor dispute, whether by advertising, speaking, patrolling, or by any other method not involving fraud or violence." This provision is intended to protect unions in the right of peaceful picketing, and the major difficulties now arise in determining what picketing is "peaceful" and what is not. As an outstanding authority in the labor field has put it:[1]

> In the interest of sound policy, the law should permit peaceful picketing in the conduct of lawful strikes and boycotts and define it accurately by statute in so far as possible. Justice and equal treatment under the law require this. In fairness the law must permit both employers and workers to present their cases to those in the labor market and leave it to those in the labor market to decide to accept or reject employment or to decide how they shall bestow their patronage. But in fairness to those in the labor market and in the interest of peace and order, picketing must be limited. In this connection numbers make a difference. . . . While picketing should be permitted "by numbers," mass picketing should not be tolerated. Not only is it unduly coercive; in many cases it makes bad matters worse. "Peaceful" should be defined as accurately as possible, but realistically and not too narrowly. Actual intimidation, really abusive language, and the like, should be forbidden, as well as violence, actual

[1] MILLIS and MONTGOMERY, *op. cit.*, pp. 628-629.

or threatened. Except for these limitations and avoidance of interference with use of the streets or annoyance of customers, picketing in the conduct of lawful strikes and boycotts should not be restricted.

27·12. The Boycott. Strikes are intended to cut off production of goods, and boycotts are similarly intended to cut off the sales of goods. The *boycott* is an organized refusal on the part of a group of people to patronize a given merchant or firm—an organized refusal to buy given products. They are of several different varieties. We may first distinguish between a *consumers' boycott,* which is a concerted refusal to buy given commodities, and a *workers' boycott,* which is a concerted refusal on the part of workers to handle or work upon given materials. Thus we have a workers' boycott if hod carriers refuse to carry bricks made in a nonunion plant, but we have a consumers' boycott if these same hod carriers refuse to wear overalls made by nonunion labor. Legally, the important distinction has been between the primary and the secondary boycott. The distinction is nebulous and confusing and might better be passed over entirely save for its importance in law. A *primary boycott* is one in which the refusal to patronize is directed against the offending employer only; like the direct strike, it involves only the direct parties to the dispute. A *secondary boycott* is one that involves third parties, as does the sympathetic strike. Even in a primary boycott, the boycotting parties usually seek to persuade others to join them in refusing to buy the goods in question; but in a secondary boycott threats are used to cause third parties to join in it. If truck drivers working for a coal company were to go out on strike and, in an effort to bring pressure to bear on the coal company, were to refuse to buy coal from it and seek to persuade their friends to withhold their patronage also, we would have a primary boycott; if the striking truck drivers and their friends were now to demand that stores and business firms in the vicinity also refuse to buy coal from the offending firm under threat that they would boycott the store and other business firms not joining in such refusal, we would have a secondary boycott. Such secondary boycotts are also sometimes called "compound boycotts" and, although legal opinions on the matter vary, they have usually been held illegal. Primary boycotts are usually held legal if their purpose is legal. Courts have been divided as to the legality of the workers' boycott, although again they have usually been held illegal as interfering with the business rights of third parties. The boycott is closely akin to the employers' blacklist. In fact, the black list is a form of boycott.

27·13. Violence and Sabotage. Labor disputes in the United States have been marked with *violence* much more often than has been true in most countries, and in recent years a racketeering element has crept into some unions and reduced them to groups of plug-uglies and hoodlums. Unfortunately, the few cases of violence are the ones that make the headlines and

give a bad reputation to labor disputes in general. Violence has been used by both employers and employees; but the legal structure has been such that the law has been more often on the side of the employer, and one is not seriously tempted to use violence when he already has the law as a partner. The unionist, on the other hand, has often found himself at odds with the law and has attempted to accomplish by violence what he could not accomplish otherwise. The dominant individualism of the United States has put labor organizers at a decided disadvantage not found in countries with a different basic social philosophy. Be the cause what it may, it must be admitted that violence, always inexcusable in labor cases, has been a rather common concomitant in this country. Much of it has been sporadic, unorganized, and fortuitous—the result of hot tempers and without premeditation. Some of it, on the contrary, has undoubtedly been planned in advance and the product of preconcerted action. Murders, maimings, destruction of property, assault and battery, and many other forms of violence have been committed by unions, especially in those cases where the racketeering element has gained control.

One should not get the opinion that such violence is typical or that it is a necessary accompaniment of collective bargaining or of labor disputes; yet it would be shutting our eyes to the facts to omit it from the weapons that have been and are now used by labor in times of trial. Closely allied to it and, like it, illegal, are various subtle forms of sabotage, such as intentional neglect to oil machinery so that it breaks down on the job, purposeful spoiling of products or materials, or even dynamiting or bombing the plant. *Sabotage* consists of willful obstruction of productive processes by employees with the purpose of inducing the employer to accede to their demands. Sabotage is usually, because of its nature, carried on secretly, and there is no way of knowing, therefore, just how prevalent it is. The manufacturer who is having a labor dispute and who finds some morning that a delicate and expensive precision instrument has been ruined because somehow sand has got into the bearings, may not be able to prove who did it or how it happened; but he is apt to draw his own conclusions and possibly may decide that it will be cheaper in the long run to accept his workers' demands than to run the risk of further loss.

27 · 14. Politics as a Worker's Weapon. We have listed as union weapons in industrial disputes the strike, picketing, the boycott, violence, and sabotage. Many of the things that organized labor groups desire they have to accomplish by legislation; as a result, they tend to form a pressure group or lobby like any other group of people interested in influencing legislation. Such lobbies are very common. The farmers, the silver bloc, the manufacturers, the military and naval services, the women voters, the teachers—all these and hundreds of other groups try at one time or another to assist or

impede the passage of legislation, or to influence its form. Yet we seldom find a political party of teachers, silver producers, or retired admirals. In some countries, however, organized labor enters politics as a political party, and in recent years organized labor has taken an increasing interest in political activities in the United States. To date these attempts have been confined largely to working through the regular party organizations, trying to influence the choice of candidates and the writing of platforms by the parties, and trying also to elect those candidates of the regular parties who are thought to be most "friendly to labor." So far, the movement in the United States has not progressed to the point where the average voter thinks of himself primarily as a worker or as an employer. He is faced with many issues and, although he may have definite ideas relative to issues involving organized labor, the problems involving them are usually secondary or even of lesser importance. The voter is usually more interested in the views of a candidate on international affairs, military conscription, or states' rights—even in the appearance of the candidate's wife, the cut of his suit, and the inflection of his voice—than in his attitude on questions of organized labor. As long as this is true, a real labor party would not be likely to be successful in the United States. Today, it would be necessary in most sections of the country for such a party, if it were to be successful, to receive the assistance of farmers, housewives, merchants, and many others who are not union members; and these other groups usually feel that their own problems are far more important than those of organized labor. In fact, most of these groups would feel that their relationship with organized labor was more nearly a competitive than a cooperative one.

27 · 15. Agencies for Settling Industrial Disputes. We have discussed the weapons used by employers and by employees when labor disputes are in progress. Of course, there is a third party interested in every industrial dispute—the general public. Just as the employer and the employee have undoubted right to protection, so the general public has a right to protection against undesirable results of labor disputes. In part, this protection is given by means of legal restriction on the weapons used by the employer and employee—restrictions on the use of the blacklist or the injunction or the yellow-dog contract or strikes or picketing or the boycott. But protection of the general public goes much further. In part, it consists of measures designed to see that industrial disputes do not occur in the first place;[1] in part, of measures designed to settle quickly those disputes which do arise.

If an open break has actually occurred between an employer and a union and if a strike is threatened or in progress, the matter may be conciliated. *Conciliation* involves merely a discussion of the matters under dispute by the participants without the intervention of any third party. If,

[1] Some such measures are mentioned in Sec. 27 • 16.

for example, the employees of the Willington Railway Company are striking or about to strike for higher wages, representatives of the management and the union might meet to discuss the affair and see if they could reach some agreement. This would be conciliation, and the outcome would depend on the willingness, the tact, and the negotiating ability of the parties involved.

On the other hand, under the circumstances depicted, it is possible that some third party, such as the mayor of Willington, might offer his good offices as a mediator. *Mediation* differs from conciliation in that a disinterested unbiased third party is present to act as a go-between. He is often chairman of the meeting, and it is his function to keep the disputants on the subject and to keep their discussion temperate as well as relevant. It is not his duty to pass judgment on the issues, and he has no voice in the outcome. He needs patience and tact, and above all it is important that the parties to the dispute should respect his judgment and integrity. In disputes of national importance the President of the United States has offered his services as mediator; while in lesser cases the mediator may be a governor, a mayor, a representative of a national or state department of labor, a university professor of economics, or anyone else who seems to fit the bill. The important thing to keep in mind is that the mediator acts without any authority whatever. His accomplishment, if any there be, comes from his influence rather than from his power.

In the third place, industrial disputes may be settled by arbitration. In cases of *arbitration,* as in cases of mediation, a third party is called in; but in cases of arbitration the third party acts as judge, deciding what is to be done. If we go back to our hypothetical dispute between the Willington Railway Company and its employees where the latter were asking for higher wages, an arbitrator would listen to the arguments of both sides and would then rule whether wages were to be increased or not. The arbitrator, then, has far more power than the mediator.

Arbitration is of two kinds: *Voluntary arbitration,* as the name implies, is submitted to voluntarily by the disputants. The railway company and the union representatives might agree to submit their dispute to an arbitrator and agree in advance to accept his verdict. Such voluntary arbitration is rather common in the United States. Many labor agreements specify that if any dispute arises during the life of the contract it shall be submitted to arbitration. In the United States employers and employees are not required to arbitrate disputes unless they wish to do so. Some countries have systems of *compulsory arbitration,* in accordance with which every industrial dispute must by law be submitted to arbitration and the disputants must accept the verdict. In such cases, there are usually special industrial courts to hear the disputes and to decide what is to be done. Compulsory arbitration is unconstitutional in the United States and is looked upon with disfavor by both

employers and employees. Workers cannot be forced in this country to accept work at terms that are unacceptable to them, and employers cannot be forced to operate their plants if they do not wish. It should also be said that, even in those places where compulsory arbitration is used, it has been far from successful, and most students would hold that it cannot exist in a society of free enterprise.

Very different from conciliation, mediation, and arbitration is *compulsory investigation*. Under such a system, the parties to an industrial dispute cannot proceed with a strike or lockout—cannot break off their normal relations—until after their dispute has been submitted to an impartial investigating board. The board has no power to enforce any decision, but the method does supply a "cooling-off" period during which the parties to the dispute can think things over before taking final action; it also provides a method for influencing public opinion in the case. Often a strike is won or lost according as the workers or the employer has the sympathy of the general public. The purpose of compulsory investigation is to provide information to the public on which it may base its decision as to the fairness of the demands involved in the dispute.

Often more than one of these methods is used in an attempt to settle a labor dispute. Usually the first attempt is by means of conciliation. If that fails, any one or more of the other procedures may be tried, with the exception, in the United States, of compulsory arbitration.

27·16. Proposals for Preventing Industrial Disputes. We have just been describing various methods for handling industrial disputes after they have arisen. An ounce or prevention here, as elsewhere, is worth a pound of cure, and it is natural that we should interest ourselves in proposals for preventing the disputes entirely. Such proposals are numerous, although many of them are perhaps impracticable.

In Chap. 4 we noted that incomes are very unequal in the United States and that there is much natural unrest, especially among those people who receive low incomes, as a result of this inequality. It may be possible for men to admit, even if their incomes are low, that complete equality of incomes would be undesirable, that some men are worth more than others, and that differences in income serve as an incentive to greater production. But it is hard to get anyone to admit that some men are worth a thousand times as much as others or that the value of men's contributions varies as much as the amounts of their incomes. Therefore, any program designed to lessen the inequality described in Chap. 4 is apt to lessen to some degree the discontent with and the amount of dispute over wages.

Closely allied to proposals for reducing the inequality of incomes are proposals for some system of profit sharing in industry. These proposals usually envisage the ownership of stock by the workers in an industry, mak-

ing them to some degree owners of the business. The dividends on their stock will supplement their wages, and they will share to some extent in the success of the business. To be sure, it is impossible to share profits without running the risk of sharing losses. Those who propose profit-sharing plans often do not realize that profit sharing combined with loss sharing may not be advantageous to the worker and that he may not wish, nor may it be socially desirable for him, to share the risk of operating a business enterprise. Profit-sharing plans have always been much more popular among employers than among employees, perhaps because they give some obvious incentive to increased efficiency in production.

There is hope on the part of many people that careful and scientific study in the field of industrial management may help solve many of the problems arising out of conflict between employers and employees. Certain it is that many of the programs of management in the past have rested on nothing more scientific than old wives' tales. Methods and procedures have been adopted as the result of intuition, superstition, or custom which in turn originated in intuition or superstition. Employers have in many cases decided what wages to offer or what amount of work it is reasonable to expect, etc., without any foundation in reason for their decisions. Naturally, under such circumstances, the results have not always seemed reasonable and have been open to honest dispute. It is the hope of many modern business administrators that scientific study in the field may furnish some basis for deciding these questions on an objective rather than a subjective basis, so that the conclusions will have some demonstrable validity. The unions have, from the start, been very suspicious of these programs of *scientific management* and have often resisted efforts to install systems based on their findings.

Other proposals for eliminating labor problems envisage more or less radical changes in the political and institutional structure of society, to the complete eradication, for example, of the profit system, and the substitution of some other means of encouraging men to work and of evaluating their contribution to the product. Some of these proposals are described in Chap. 36.

27 · 17. National Labor Relations Act. It is out of the question in an elementary text like this one to describe in detail the development of legal provisions concerning organized labor, or even to go into any considerable detail in describing present legal provisions. Many of them have been mentioned in passing at earlier points in this chapter in connection with the particular subject involved. But, at present, by far the most important single act is the National Labor Relations Act, often called the "Wagner Act," signed by the President on July 5, 1935. The act, of course, affects only cases involving interstate or foreign commerce, and is too long to give here *in toto,* but the most important provisions are the following:

Sec. 7. Employees shall have the right to self-organization, to form, join, or assist labor organizations, to bargain collectively through representatives of their own choosing, and to engage in concerted activities, for the purpose of collective bargaining or other mutual aid or protection.

Sec. 8. It shall be an unfair labor practice for an employer:

(1) To interfere with, restrain, or coerce employees in the exercise of the rights guaranteed in section 7.

(2) To dominate or interfere with the formation or administration of any labor organization or contribute financial support to it: *Provided,* That subject to rules and regulations made and published by the Board pursuant to Sec. 6(a), an employer shall not be prohibited from permitting employees to confer with him during working hours without loss of time or pay.

(3) By discrimination in regard to hire or tenure of employment or any condition of employment to encourage or discourage membership in any labor organization: *Provided,* that nothing in this Act . . . or in any other Statute of the United States shall preclude an employer from making an agreement with a labor organization . . . to require as a condition of employment membership therein. . . .

(4) To discharge or otherwise discriminate against an employee because he has filed charges or given testimony under this Act.

(5) To refuse to bargain collectively with the representatives of his employees, subject to the provisions of Sec. 9(a).

The Supreme Court, by a majority of 5 to 4, held the Act to be constitutional on Apr. 12, 1937. The decision was a close one, and it is quite possible that it may sometime be reversed. Strong advocates of organized labor naturally look upon the act as sacrosanct—as the Magna Charta of labor. Opponents of organized labor, on the contrary, view it as a one-sided and unfair act passed for political advantage and validated by a packed court. The truth lies perhaps somewhere between the two extremes. It seems likely that the future will see material amendments to the act, especially if and when there is an overturn of party control in Congress and in the administration; but the direction of the changes cannot be predicted in detail. One cannot say with truth as yet that the purposes of the act have been accepted by a vast majority of employers, many and perhaps most of whom are still reluctant and forced cooperators with its provisions. The general public is naturally ill informed as to the provisions of the act and for the most part uninterested in the issues involved. Many workers, even, are apathetic, and the records of the National Labor Relations Board show that considerable numbers of them fail to vote in the elections held under the act. Those who were already enthusiastic advocates of organized labor, as we have just said, are likewise enthusiastic advocates of the Act, which has given to organized workers privileges that they had sought long and vainly.

27 · 18. Summary and Conclusions. Workingmen have for centuries attempted to band together and act in concert for their mutual advantage, and such organized activity has increased greatly as industry has come to be organized on a larger and larger scale in the form of corporations and consolidations. The workingman has often felt that he was not sufficiently protected by the forces of competition, which works far from perfectly in the labor market and which is at any rate a distant and unseen force. Organization has given to workingmen some element of monopoly power with which they have hoped to offset the power of organized employers. In the disputes that have arisen between employers and unions, each side has developed certain well-recognized weapons. Each side objects to the weapons used by the other, and in each case the law has stepped in to modify the uses of these weapons. Various sorts of agencies have been devised for settling those disputes which occur, and various proposals have been made for eliminating the original causes of such disputes. Recent legislation, both state and Federal, has tended to improve the position of organized labor and increase its power, until there is a fairly general feeling that some amendments will be necessary to bring about a reasonable balance. The issue at the moment is not so much whether changes should be made as what the changes should be and whether they should be made by the "friends" or the "foes" of organized labor.

The student is warned again as we leave this subject, as he was when we took it up, that he must keep clearly in mind the difference between "labor" and "organized labor." We can probably all agree that something needs to be done in our society to decrease the inequality in the distribution of income and to improve the lot of the poorer people. This seems to be pretty generally conceded. The difficult thing is to know how it can best be done. Most people who favor unionism do so because they believe that it will bring about this desired result. They recognize the fact that the individual worker is too small a unit to have any bargaining strength in a world of corporations; that investors, capitalists, and management operate with a united front; and that the union gives the same sort of monopoly power to the worker. What they often overlook is that the union today is often, and perhaps usually, stronger than the employer and that it is now perhaps the employer who needs the protection rather than the worker. A textile mill in a southern New England town has by no means the resources of one of the great textile unions. The union often has more funds and hires men who are more expert in the field of collective bargaining and even in the field of textile economics than the average millowner can possibly command. As mentioned earlier in this chapter, the company union has a bad name because it was often a company-dominated union; but an

independent union organized on a company basis would put the employer and the employee on much more equal footing in bargaining than either the individual bargaining of single workers or the present system of the single employer against the more powerful union.

One unexpected development, which has not received by any means enough attention, is that present union bargaining methods are forcing employers into even more monopolistic practices than they used before. As this is being written, the radio has carried the news that the Master Barbers' Association of Hartford, Conn., has agreed with the barbers on a new wage scale. These employers of barbers have, it seems, been in collusion—taking joint action, possibly action in restraint of trade. It seems quite likely, if the point were forced, that every time the mine owners get together to bargain with the United Mine Workers, the owners are violating the law. The workers are given specifically the right to bargain collectively, but the employers are not. There seems to be no good reason to give the employers such rights, yet, if unions are to be allowed on an ever larger scale, it may become necessary to protect employers by granting them the same right.[1] The point seems to have been pretty much overlooked that, if our real purpose is to achieve equality of bargaining power between employer and employees in collective bargaining, the independent[2] company union may well be the unit that equates the powers, while the national or the international union goes too far in giving power to the workers.

[1] The day after this was written, the author received a copy of the Papers and Proceedings of the Fifty-eighth Annual Meeting of the American Economic Association, *American Economic Review*, vol. 36, no. 2. There he found, on pp. 162-163, a short reference to this matter in an article by George P. Comer of the Department of Justice. A small part of his statement follows: "From the point of view of *economic* restraints of trade, the growing practice of joint or group bargaining by the employers in particular industries may become one of the most serious problems to be faced in that borderland between labor and employer agreements. Normally we think of collective bargaining in terms of an agreement between a particular employer and a particular labor union. Gradually the union activity has spread to entire industries so that an industry union bargains as a unit with the several producers of the industry. The next step which has become more and more prevalent in recent years is for the employers in an industry to get together and present a united front against the union whether of the AFL craft type or the CIO industrial type. If this practice should spread over all the principal industries of the country, we might have the spectacle of industry and labor combining against the consumer, because the easy way out of a wage dispute would be to bind all competitors to the same scale and pass the increases along to the consumer."

[2] By an "independent" company union is meant one that is not dominated by the employer, but one that limits its membership to employees of one concern. It would bargain through representatives of its own choosing, who need not be employees of the concern—but it would be necessary then to stipulate that there could be no concerted action by several of these company unions, and that several of them could not use the same bargaining agents. In other words, collusion between such independent company unions would be restraint of trade in the same sense as collusion between the respective company managements.

Surely it seems true as this is being written that some curb must be placed on irresponsible power of unions. The threat of the coal and railroad strikes of the spring and summer of 1946 have just paralyzed the nation, and many people are realizing for the first time that the unions have in some cases become powerful enough to challenge the government. Although we all wish to improve the lot of the workingman, there is serious question whether it is wise to do this by putting in his hands a monopoly power that he can turn to his own selfish advantage to the disadvantage of society as a whole. It is not sufficient to say in answer to this, as is sometimes done, that the union men are themselves a part of the society. The result of unrestrained monopoly power, whether in the hands of capitalists or workers, is a net loss to the society as a whole. Great power is dangerous, and has always been dangerous. It should be granted grudgingly and cautiously and only from necessity, and it should be hedged about with restraints and safeguards. It is the view of this author[1] that, although the present laws have been passed quite properly with the intent of equalizing bargaining power and improving the lot of the poor, they have gone too far for proper balance and must somehow be revised to give organized labor only as much power as it seems absolutely necessary to give to management in order to get efficient production. We cannot sensibly eradicate the corporate form of enterprise entirely to keep management from building large power units, since small units would be inefficient. But we do try to enforce a reasonable degree of competition among these production units and to keep them from getting larger than necessary for efficiency. There is reason to allow labor organizations to grow large enough to bargain on equal terms with them, but no reason to allow such organizations to gain even more monopoly power than have the corporations with which they deal.

Questions

1. Give several examples of craft unions. Of industrial unions.
2. What are the major differences between the AFL and the CIO?
3. Why is the bargaining power of a single unorganized worker smaller today than it was two centuries ago?
4. Why is it that there is not complete competition in the labor market?
5. What is meant by collective bargaining?
6. What is a closed shop? An open shop?
7. What is a closed union, and what is the disadvantage of combining it with the closed shop?

[1] It is only fair to warn the student that many, many economists do not share this view but, on the contrary, support wholeheartedly the present status of labor legislation.

8. What is the lump-of-work doctrine? Discuss it.

9. Reducing the hours of work does not necessarily reduce the amount of output. Explain.

10. Laborsaving machinery is advantageous to society as a whole, but often disadvantageous to skilled laborers who are displaced by it. Suggest some way in which society might protect the displaced workers.

11. What is an injunction? In what sorts of cases is it used?

12. Describe the yellow-dog contract. What is its present legal status?

13. What is a strike? What is the major way in which employers and strikers would differ in their definitions?

14. What is a sympathetic strike? A jurisdictional strike?

15. Under what circumstances are strikes illegal?

16. Discuss the legality of picketing.

17. What are the major reasons, as you see it, for the fact that organized labor has been less active in politics in the United States than in many other countries?

18. What is the major difference between
    a. Conciliation and mediation?
    b. Mediation and arbitration?
    c. Voluntary and compulsory arbitration?

28. SOCIAL SECURITY

28·1. Desire for Security. Security, the lexicographers tell us, means freedom from fear, anxiety, or care; confidence of power or safety. Ever since the dawn of history, men have had their cares and worries. Early man was in constant peril of his life from wild beasts, and constantly pressed for food and shelter. The progress of economic organization has been such that modern man is seldom in fear of starvation; anxiety with regard to filling the basic animal wants is no longer omnipresent. Yet new anxieties have arisen to replace the old ones, and each of us today has his own personal fears, anxieties, and cares. These may be tied up with the approach of another examination or marking period, or they may stem from some chance remark of a friend. The fears of childhood often seem foolishly unnecessary to the adult, and presumably many of the fears of adults are equally foolish. Yet, fears are real to those who experience them.

Of the many fears and anxieties to which modern man is subject, several are connected with his economic life, arising from his efforts to get a living for himself and his family. Men are hungry for safety and security, longing for freedom from anxiety as the youngster alone in the dark longs for his mother's protecting presence. Freedom from the threat of economic disaster is commonly referred to as *social security,* and we now turn to a consideration of men's efforts to achieve freedom and confidence in this field. Today, when most families are entirely dependent on the economic system for their livelihood, there is constant fear that the income earner may get sick, suffer injuries, be laid off or dismissed from his employment, die and leave a dependent family, or, living beyond the time when he is able to earn, face a destitute old age. Although some people are more amply provided for than others, no one is completely safe from economic disaster.

To be sure, there is some question whether men would really enjoy security if they had it, and whether or not it would be good for them. It has been argued that men would find life drab and monotonous if no economic problems faced them—if they could be certain in advance that

every want would be satisfied. This matter was discussed briefly in Sec. 1·8. There are also those who believe that adversity is good for man's moral fiber and that easy living makes men soft of spirit and morals, as well as of body. These people argue that security would be bad for men—that, if they felt no anxiety either for the present or the future, they would not be industrious and productive and frugal. Although it may seem peculiar to the student at first to hear it suggested that security would be undesirable, there is nevertheless a good deal of weight to this position.

On the other hand, security, like income, is at present unequally distributed. Some people have a great deal of it and others very little. This depends partly on their psychological outlook on life. It depends on whether they happen to be introverts or extroverts, and on the state of their digestion. We all realize that, if we were to find two men of equal wealth, equal income, equal family and business responsibilities, and equal prospects for the future, one might be carefree while the other was overburdened with doubt and care and fear. Although one's frame of mind has much to do with the feeling of security and although men of equal prospects differ in their outlook, nevertheless we must realize that men do not all have equal prospects. Some of them have much more reason than others, based on any objective criteria, to fear the future. And it is these differences in men's prospects rather than in their psychological make-ups that we are now to discuss.

28·2. When Is a Man Secure? Security is a relative thing, and there is no accepted objective basis for deciding how much a man needs in the way of income or savings to be secure. Yet we can give some rough indication of the nature of the problem without attempting to set up absolute standards.

In our society we accept without question the obligation of caring for those people who are unable to maintain themselves. No one seriously proposes that the sick or unfortunate be left to starve. The old, the chronically ill, the mentally incompetent are looked upon as proper public charges if they are unable to support themselves. A program of social security is thought of as one which, as a minimum, aims to keep people from becoming objects of public charity—to maintain their economic independence in the light of whatever standards and requirements their society sets up. These standards, as we have seen,[1] differ greatly in different communities. The minimum acceptable level of living in one society might be looked upon as a level of comfort or luxury in another. We do not have social security in any society as long as any considerable number of people in it have to depend on public charity to maintain whatever standard is considered the minimum acceptable there. We must perhaps expect in any society that a few of the most unfortunate people will be physically or mentally unfit to support themselves, but

[1] See Chap. 1.

it is the purpose of a system of social security to make sure that persons of reasonable competence and foresight should be able to support themselves, not only through the period of their own working life, but through the period of their old age. In fact, most systems of social security seek to improve on the foresight of the average man and require him to make provision for various contingencies where his own foresight would not have led him to do so. Social security is not supposed to be public charity, nor a gift from the state. It is a substitute for charity, and under it each man pays his own way. In practice, however, modern systems of social security are sometimes tied in with charitable relief in such a way that the two are difficult to separate.

28·3. Workman's Compensation. Under the common law, workers injured as a result of the normal hazards of their occupations were entitled to sue their employers for recovery of their losses. In case of such a suit, the court attempted to discover who was responsible for the accident—who had been to blame. If the entire responsibility fell upon the employer, the injured worker could recover; if all or even any part of the responsibility fell upon the worker himself, he could not recover. The entire problem was considered to be a problem of *negligence.*

In actual practice, workers found it almost impossible to recover for injuries under this system. The employer had an adequate defense if he could show *contributory negligence* on the part of the employee; if he could show *negligence of a fellow servant;* if he could show that the injury arose from a hazard which was normal to the business and which the worker might reasonably have anticipated when he took the job—the doctrine of *assumption of risk.* In any of these cases, the injured worker himself had to bear the burden resulting from the injury.

With the development of large-scale business enterprise and the use of power machinery, industrial accidents became more serious, and at the same time the complications of industry made it harder and harder for the employee to prove that the negligence that caused the accident was solely and purely that of his employer. The old common-law rule of negligence became more and more unworkable. As a result, a movement began, starting in Germany and spreading slowly through Europe and only very tardily into the United States, to compensate for industrial accidents without any regard to negligence or without deciding who was to blame for them. These modern systems provide standardized benefits for various sorts of accidents, with the benefit varied somewhat in accordance with the amount of financial loss incurred by the worker, although with the compensation usually somewhat lower than the full amount of the loss. Today industrial accidents are regarded as one of the normal costs of business, chargeable properly to the business rather than to the particular employees who are unfortunate enough

to incur them. The employer pays the compensation provided by law and he may, in turn, protect himself through some sort of insurance.

In the United States, present workman's compensation laws are adopted and administered by the individual states, varying from state to state. In some cases, they are compulsory, every employer being required to accept the provisions of the act and to pay the benefits there established. In other cases, employers are allowed to choose whether they will come under the act or continue to operate under the old common law with regard to liability. If they elect the latter, most statutes provide that they may no longer fall back on the old defenses of negligence. Most states also restrict the compensation laws to certain fields of endeavor; excluding, for example, farm workers and domestic servants. Sometimes establishments with very small numbers of employees are likewise exempted. Rates of compensation vary but are always some percentage of the wages that the injured worker was receiving. There are also usually restrictions as to the length of time for which compensation may be paid and as to the maximum amount of the payment. In order to guarantee that the required sums will be paid, most states require the employer to carry insurance covering his risk, sometimes in a state insurance fund and sometimes with private insurers. Some states administer their systems of workmen's compensation by means of a board or commission to which complaints may be taken; other states leave the matter to the employer, the employee, and the courts. Under this latter system, if the employer and employee are unable to agree, they may take the case to court for adjudication.

The systems of workmen's compensation differ greatly in the different states, but the important element of modern systems is the assumption of the risk of loss from accident by the industry, making the cost of accidents a part of the cost of production, just like the costs of labor and materials. It is a cost that is paid immediately by the employer but, like other costs, added to the price of the goods and passed along to the ultimate consumer. Yet, even under modern systems, only part of the cost is so carried. The rate of compensation is never so high as the wages that the worker was receiving, and usually amount to from one-half to two-thirds of those wages. In this way from one-half to one-third of the economic loss is thrown immediately upon the injured worker and his family. Also limitations of payment period and of total amount of compensation payments tend to throw part of the burden on the worker. The worker, consequently, has not been entirely relieved of responsibility for accident by the removal of the older penalties for negligence; he still has a distinct economic incentive to be prudent and careful. It has been estimated that injured workers still bear more than half the money cost of their injuries and that in the less liberal states the proportion borne by workers is even greater. Yet the new system does avoid the

heavy costs of litigation that were involved under the old common-law procedure, they provide a certain and definite recompense, and they make it possible for the workers to obtain needed medical attention promptly.

28·4. Unemployment Compensation. Among the risks of modern industry is that of unemployment. To be sure, the victim of an industrial accident is not employed, nor is the typical octogenarian or six-months-old infant. Strangely enough, in the jargon of the economist to be unemployed is not the same as not to be employed. Many people who are not employed would not be listed among the unemployed. Technically, *an unemployed person is one who is able and willing to work, but who is not working and not able to find work for which he is qualified under conditions that are reasonable as judged by local standards.* Under this definition, one who is mentally or physically incapacitated is not unemployed, nor is an able-bodied skilled artisan who is taking a voluntary vacation or who is out on strike. A workman who loses his job, but who could immediately start work on another job of similar nature and at the same wages, is not unemployed even if he decides to look around for other opportunities for a time before accepting work. If a skilled electrician loses his job, and the only thing that he can find is work at some unskilled low-paid occupation, he is unemployed while he continues to seek for work for which he is especially qualified or, if this skilled electrician can find a new job as an electrician but only at wages that are half those commonly being paid for such work, he is unemployed while he continues to look for work under conditions that are reasonable, as judged by local standards.

Unemployment arises from many causes. Perhaps the most common are the *cyclical unemployment* which occurs during the depression phase of the business cycle[1]; the *seasonal unemployment* which comes regularly at the slack season in seasonal businesses; the *technological unemployment* which is often the temporary result of the introduction of new methods of production; and the *normal unemployment* which arises in any society as newcomers to the labor market look for their first jobs, as factories burn down and workers seek new positions, as competent workers are discharged for various causes. Thus the burden of unemployment rests much more heavily on society (especially, on the unemployed members of society) at some times than at others. The normal unemployment has been estimated at about 6 to 8 per cent of the total labor force and, at times of peak cyclical unemployment, it has risen to perhaps a quarter of all wage earners excluding agriculture. In absolute numbers the normal unemployment is probably about 1½ million workers, while in the depression of the early thirties the number of unemployed is estimated to have reached 13 or 14 million. The fear of this unemployment had led workers to adopt many of the practices, such as limiting output,

[1] See Chap. 22.

opposing the introduction of laborsaving machinery, etc., which were mentioned in the preceding chapter.

Various schemes have been tried or suggested for reducing unemployment. Vocational education has helped workers acquire new skills. Private and public employment offices have been opened to help unemployed persons learn of vacancies. Private industry has been urged to stabilize employment by developing side lines that use workers during the period when the major product is undergoing its slack period or by following a practice of continuing employment during the slack period and making goods for stock. The government itself has usually followed a policy of undertaking public works in times of general unemployment in order to absorb the unemployed at that time. Many of these schemes presumably operate to reduce the amount of unemployment, but there is still a considerable amount of it left, and a social security program naturally attempts to take care of the burden of this residual unemployment.

Social security programs have ordinarily tried to meet this need by means of unemployment insurance. This, like other insurance, requires the payment of premiums in advance out of which benefits can be paid in time of need. Early plans were voluntary. The city of Ghent in Belgium established such a system as long ago as 1901. Under the Ghent system, the city cooperated with trade-unions, the city paying half the benefits out of public funds and the unions paying the other half out of assessments levied against their members. Any such plan, however, reaches but a minority of the workers and does not touch the major social problem. Consequently, there has been a tendency in recent years to develop systems of compulsory insurance, supervised and at least partly financed by the government.

In the United States, unemployment insurance, like workmen's compensation, is under the jurisdiction of the individual states, and the plans differ somewhat from state to state. The Federal government, however, in its Social Security Act of 1935, brought about some uniformity in state statutes by a method that has become increasingly common in recent years—a system that amounts in effect to bribing the states to do those things which the Congress has not the power to compel them to do. In this case the Federal government offers to pay the administrative cost of any state program if the state will meet certain requirments. In addition, the Federal government lays a tax of 3 per cent on pay rolls, the tax being paid by employers.[1] But the government credits those employers who pay into state unemployment insurance plans up to 90 per cent of the amount owed the government. For

[1] Actually this tax covers only the first $3000 paid to any particular employee. If an employee received $5000 in a year, the employer would pay a tax of 3 per cent on only $3000; if another employee received $2000 per year, the employer would pay 3 per cent on the entire $2000.

example, if an employer has total pay rolls of $12,000 and if all his employees are receiving wages of $3000 per year or less, the 3 per cent tax owed to the Federal government will amount to 3 per cent of $12,000, or $360. If the employer has paid $100 into a state unemployment insurance system, he can deduct this amount from the tax owed the Federal government, and pay them but $260. Similarly, he could deduct $200 if he had paid that amount to the state. But the maximum that he could deduct would be 90 per cent of the $360 he owes, or $324. If he has paid $324 into a state fund, he will owe the Federal government $36; if he pays more than $324 into a state fund, he will still owe the Federal government $36. It is plain that the employer pays no more as the state tax gets larger and larger until the state tax reaches 90 per cent of the amount owed the Federal government; and this would be 90 per cent of 3 per cent of the pay roll, or 2.7 per cent of the pay roll.

All the states have now established unemployment insurance plans in cooperation with the Federal government. In the fiscal year 1944-1945 the Federal contributions to the states for the administration of the state programs amounted to 30 million dollars, and the trust funds accumulated by the states amounted to 6.7 billion. The states are allowed considerable leeway in the details of their acts, but minimum standards are set by the Social Security Act. Funds collected by the states are deposited in a Federal Unemployment Trust Fund in the U. S. Treasury. Certain kinds of workers are exempted from the act, some of the major ones being agricultural workers, domestic servants, the crews of American and foreign vessels, certain government employees, newsboys under eighteen, and employees of certain charitable, religious, and educational institutions. Roughly 45 million workers had wage credits in state systems at the close of 1944.

It is evident that, if the employers of some states were taxed for unemployment insurance while those of other states were not, the employers in the former states would be put at a competitive disadvantage. It is for this reason, among others, that the Federal government steps in to make it advantageous for the several states to adopt roughly uniform plans. The Federal plan also requires that all benefits are to be paid through public employment offices. It is the purpose of this provision to insist that the unemployed person be willing to take a new job if one is available. Each week when he reports for his unemployment benefits, he can be offered a new job if one is available, and the supposition is that, if he refuses to take work for which he is qualified, he will not be allowed unemployment compensation. It must be confessed that this provision of the act has not been rigidly enforced, and there has been a considerable amount of justified complaint at the close of the second World War that the system of unemployment compensation was encouraging idleness, by paying people unemployment benefits, which they accepted in preference to going back to work.

The American system is financed entirely by taxes on employers. Employees are not assessed. There is difference of opinion as to the wisdom of such a provision. It is likely that there is little economic difference in the long run. The employer merely pays less in the way of wages and more in the way of taxes. But there are some who think that the worker would be more keenly aware of the costs involved if he bore part of the tax directly rather than indirectly. The present system is, however, probably much easier to administer than one in which the burden of the tax is divided between employers and employees.

Although the state plans differ in detail, a typical one works somewhat as follows: If a man becomes unemployed, no benefits are paid during an initial waiting period of from 2 to 4 weeks. He then starts to receive benefits of 50 per cent of his normal weekly wages, but not more than $20 nor less than $5 per week. The length of time for which he may receive the benefits depends usually on how long he has been employed in an insured occupation, but cannot exceed, say, 16 weeks in any one year. All the figures in this paragraph are illustrative only and differ from state to state; yet they are close enough to those actually in use to give a fairly good idea of the operation of the system in most states.

Few people believe that the present system of unemployment insurance can take care of any large proportion of the cyclical unemployment that accompanies a severe depression. It is by no means a cure-all for unemployment. Yet within its limits the act seems to work well and should do a reasonable job of caring for normal and seasonal unemployment. There have been disputes among the experts as to how the reserves can best be invested, and with regard to the desirability of requiring employees to bear a portion of the direct cost. Some authorities have felt that either a direct Federal system, or one based on a direct Federal subsidy to the states without the pay-roll tax, would have been preferable. There is doubt as to the constitutionality of such methods. There is also criticism of a plan under which states with very different unemployment records charge the same assessments. Under such circumstances, it is surely true either that the states with little unemployment are charging too much or that the states with the worst unemployment records are apt to find their funds insolvent. Despite these criticisms, the student should understand that some plan of publicly administered and publicly subsidized unemployment insurance seems to be with us to stay and that, while the record is too short for final judgment, the present system seems to be working reasonably well.

28 · 5. Old-age Pensions and Old-age Insurance. In our discussion of the human resources of the United States we noted[1] that the average age of our population is increasing, and that with the passage of the time a larger and

[1] See Chap. 12.

larger proportion of our people are in that nebulous period which we call "old age." With this change, the problems incident to old age are becoming more and more important in our society, and present age trends lead us to suppose that their importance will continue to increase for some time. Old age itself, of course, is not a problem, but rather the infirmities that accompany it. For the moment, we are especially interested in the fact that most people who attain "a ripe old age" are no longer employable. Their income is cut off, although their expenditures continue. Unless they have saved during their productive period, they find themselves dependent either on their relatives or on society. It is not old age, but old-age dependence, that raises economic problems. The relatively low earnings of many workers make it difficult or impossible for them to lay anything by to support them when earnings cease, and the relatively early age at which many workers become unemployable under modern production methods tends constantly to prolong the old-age period. At the time of the 1940 census, one person out of every five in the United States was sixty years old or over. The National Resources Planning Board estimates that by the year 2000 three out of every eight persons in the United States will be in that age group.[1]

Society formerly cared for its aged by means of poor relief: almshouses, poor farms, and public or private charity. The method was not only psychologically bad for the indigent aged themselves, but it turned out to be a very expensive way of handling the problem. In recent years, the tendency has been to handle it by one or another of two methods: either by a *pension* paid weekly or monthly without any previous payment into a reserve fund by the recipient, or by an *insurance* system into the reserve fund of which people pay during their productive period and from the reserves of which they draw an annuity when they become aged.

Noncontributory pensions are paid by many corporations to former employees and, in addition, many of the states make provision for old-age pensions payable from public funds. The Social Security Act of 1935 made further provision in this direction. Under this act the Federal government bears half of the expense of approved state pension systems up to a maximum of $20 per month per pensioner.[2] All the states and territories have approved plans in operation, and in June, 1945, over 2 million persons received over 60 million dollars from these funds, an average payment of $29.46 per month.

Far more important than the pension plans, however, is the old-age insurance phase of the Social Security Act. Under this so-called "Federal Old-age and Survivors' Insurance" section of the act, deductions are made regularly from employees' pay envelopes in the amount of 1 per cent of the

[1] "Statistical Abstract of the United States, 1944-1945," Table 24, p. 27.

[2] Twenty dollars is the Federal share. The state adds another $20 to make a maximum of $40 per month.

wage.[1] Employers are charged a pay-roll tax of the same amount. These funds, unlike those for old-age pensions, unemployment compensation, and workmen's compensation, are administered by the Federal government rather than through Federally aided state plans. Insured persons become eligible for payments when they reach the age of sixty-five, if they are fully insured and file a claim for benefits. Payments are made not only to insured workers, but to their wives and children (if the former are over sixty-five and if the latter are under sixteen or, if in school, under eighteen), and to dependents of deceased workers. Thus the amount of the benefit depends on the size and age distribution of the family, running for the entire family from a minimum of $10 to a maximum of $85 per month. The amount of the benefit also depends on the amount of the worker's earnings and the length of time during which he has contributed to the fund. The benefit amounts to

1. Forty per cent of the first $50 of average monthly wage, plus
2. Ten per cent of the next $200 of average monthly wage, plus
3. One per cent of the amount obtained under the preceding two rules for each year in which the worker earned $200 or more in a covered employment.

This will be more easily understandable if we illustrate it. Let us imagine that a worker retires at the age of sixty-five after having been covered by the act for 30 years. His average monthly earnings have been $200. His annuity would be

40 per cent of $50	$20
10 per cent of $150	15
(1 per cent of $35) × 30	10.50
Monthly annuity	$45.50

He would be entitled to this annuity for the rest of his life. He would also receive half this amount ($22.75) each month if he had a dependent wife over sixty-five years of age, and another equal amount ($22.75) for each dependent child under eighteen, but with a maximum for the entire family of twice the pension that would be due a single worker, or a maximum of $85.

When a fully insured worker [2] dies either before or after retirement, the survivors receive the following benefits:

1. Three-fourths the worker's benefit for a widow age sixty-five or over, or for any widow regardless of age if she has dependent children.

[1] This deduction started at the rate of 1 per cent, and the original act called for increases of ½ per cent every third year until it reached 3 per cent in 1949. Amendments, however, have so far (1946) kept the deduction at the original amount of 1 per cent.

[2] A worker is "fully insured" if he has received as much as $50 in covered employment in each of 40 calendar quarters or as many as half the calendar quarters between his twenty-first birthday and the time of his death with a minimum of six calendar quarters. Special provision was made for persons who were over twenty-one years of age on Dec. 31, 1936.

2. One-half the worker's benefit for each unmarried dependent child under eighteen years of age.

3. One-half the worker's benefit for each dependent parent if there is no surviving widow or dependent child.

4. If there is no survivor entitled to the above benefits, a lump-sum death payment equal to six times the worker's monthly benefit is paid to certain relatives, or to the person paying burial expenses.

In the case of our recent numerical illustration, the worker was entitled to $45.50 per month. If he had a wife and a dependent child his pension would have been

For himself	$45.50
For his wife	22.75
For his child	22.75
Total	$91.00

But the regulations provide that the minimum payment shall be $10 per month and the maximum shall not exceed:

1. $85, or
2. Eighty per cent of the average monthly wage, or
3. Twice the amount that would be paid to a single worker.

In this case 80 per cent of the average monthly earnings would be 80 per cent of $200, or $160, and the amount just computed does not exceed twice the amount which a single worker would get; but it does exceed $85, so our worker would actually get $85 per month. If he died leaving a widow over age sixty-five and a dependent child, the widow would receive

Three-fourths of his pension which is three-fourths of $45.50.	$34.12
One-half his pension for the child, or one-half of $45.50.	22.75
Total	$56.87

The widow would receive $56.87 per month as long as the child was below eighteen years of age, and thereafter she would receive $34.12 per month for the rest of her life.

The Social Security Board keeps a wage record for each employee covered by the act. At the end of June, 1945, there were such accounts for over 70 million persons, each identified by his own social security number. The status of the plan on that date is shown in Table 28 • 1.

The old-age insurance provisions of the Social Security Act cover only certain classes of workers, excluding agricultural workers, domestic servants, employees of the Federal, state, and local governments, employees of religious, charitable, and educational institutions, and some others. These people are excluded in part for political reasons, and in part because of the difficulty of getting accurate records of their wages. Of course, many of the

governmental employees, and some of the others, are cared for through pension schemes or through insurance plans established by their employers or by themselves.

There are various difficulties and problems that must be faced in operating the present plan of old-age insurance. One of the most difficult is that it necessitates the collection of tremendous reserves, and it is difficult to know how these reserves can be productively invested to earn the interest necessary if the benefits are to be paid. If invested in government securities, there must be a continual tremendous public debt; if invested in private industry, a vast proportion of the nation's productive resources will soon belong to the government. We shall reach a situation of socialism via the route of old-age assistance. Yet these difficulties must be faced and these

Table 28·1. Status of Old-age Insurance Plan, June 30, 1945

Beneficiaries	Number of beneficiaries	Amount of monthly benefits
Retired workers	519,186	$12,357,934
Aged wives	153,521	1,937,931
Children	378,315	4,692,331
Widows with children	144,923	2,870,973
Widows without children	83,007	1,672,764
Parents	5,727	74,828
Total	1,284,679	$23,606,761

problems solved. The plan is at present operating satisfactorily on a businesslike basis, and the benefits are so great that there can be no question as to its continuation.

28·6. Other Social Security Programs. It is hard to tell exactly what activities ought to be listed as part of a nation's social security program and what parts should be classified otherwise. One can take the easy way of listing merely those things for which provision is made in the Social Security Act of 1935; yet, as is true of most acts, some things are there included which might not be classified on abstract grounds as social security, while some things that might be so classified are omitted.

We might start by asking ourselves, "How can one tell a social security program from any other program?" In the first section of this chapter we said that social security meant freedom from the threat of economic disaster, yet that definition is so broad and general as to be almost useless when attempts are made to apply it to particular cases. What is an economic disaster? How free can one ever be from the threat of economic disaster in a world like ours? One cannot insist on Utopia. One does not expect in a

social security program to give people everything they want—to satisfy all their desires. No one seriously suggests that we should set up a program of insurance which would make it possible for every family to attend the movies twice a week, or that would make it possible for every family to have five servants and a steam yacht. What are the things provided under social security programs, and what things do we omit?

There is a good deal in common between the modern idea of social security and the idea of charity. We are, most of us, quite willing to have some degree of inequality in our society. There is even reason to suppose that some degree of inequality is desirable.[1] We are also well aware of the fact that, even if incomes were equal, people would not consume the same things because of differences in their tastes. Some people would attend the movies twice a week or more, and others would be pleased to be spared. Some people would gladly spend money for servants or steam yachts, while others, even if they had the funds, would prefer to put them to other uses. Yet there are some things so fundamental that we feel everyone should have them, even if they do not want them or will not make provision for them voluntarily. If parents are unwilling to educate their children, society insists that it be done anyway, at public expense. If men do not provide food for themselves, the public provides it at public expense rather than allow them to starve. In any society there are some things that we do not leave to individual whim and preference. We feel that everyone must have at least these things to be human and that, if men do not supply these things voluntarily for themselves, we must either force them to do so or provide them by other means.

These absolute necessities have long been furnished by public or private charity. Food, clothing, crude shelter, the most urgent sorts of medical care, etc., are examples. And it is these same sorts of things that now make up the basis of a social security program. We know that society must somehow provide for the maintenance of men who are out of work because of accident, old age, or the unemployment incident to a depression. The question is not whether these men should be supported, but how can they best be supported. These phenomena have a social cost inherent in them unless we intend to let men starve, and the problem is one of deciding how we can minimize the amount of the cost and how the cost can best be borne. Social security programs do not pretend to give men everything they want, but merely to provide those things which we think of as the absolute minimum necessities of human life. Social security programs attempt to supply these things, not by handling each case as it comes along but by advanced planning and by organization that specializes in the problems involved.

The student will readily understand, under the circumstances, that

[1] See Sec. 4 • 5.

social security programs deal with minima. No attempt is made to give pensions or unemployment benefits or accident compensation that will allow the recipients to live "the life of Riley," nor even to live an average life. They are purposely pitched at the minimum acceptable level.

For this reason there has been little general enthusiasm for some of the suggestions that have been made for the addition of other services to the social security program. To be sure, it is quite likely that some things could be, should be, and will be added, but many of the proposals are by no means in terms of minima. There are always those who would like to regulate the lives of others, confident that they know enough about other people's desires and advantage so that they can do a better job than the person can do himself. These persons, suffering from what has been aptly called the "Messiah complex," would write more and more of consumers' decisions into the statute books. With the best of intentions they would make more and more decisions for consumers instead of letting consumers make the decisions for themselves. In a "free society," if we may borrow a phrase from a recent widely known Harvard University report, it is necessary to resist this constant temptation and to make men's decisions for them only in those cases where society cannot tolerate any decision save one.

Many other programs could be listed here as examples of actual or proposed social security measures. The legal minimum wage is often so listed. its advocates would argue that there are some wages so low that no person should be allowed to accept them—that they will not support life at an acceptable level. If an industry cannot pay higher wages, the industry should not be allowed to exist. In such cases, the law has stepped in to specify minimum wages. To be sure, this means that some men, who are not productive enough to earn such wages, are thrown out of employment. If a man is so unproductive that he can produce only 10 cents' worth in an hour and if the law provides that no one is to be paid less than 35 cents an hour, then the unproductive man will not be employed. No employer can afford to hire him. But the advocates of the minimum wage would have a good deal of ground for contending in such a case that the man can better be cared for by some other means than by having him work at wages that will not support him, even at the minimum acceptable level. Various programs of health insurance, life insurance, etc., have also been suggested or tried as social security measures. One can hardly expect agreement as to which of these programs is desirable. The tendency in many other countries has been to develop them far beyond anything that we in the United States have yet been willing to accept. This, however, is in part a reflection of the fact that many foreign countries have been willing to tolerate much more interference of the state in what we consider "private affairs" than have the individualistic people of the United States.

Questions

1. On what ground can it be argued that complete security would be undesirable?
2. Describe the doctrine of negligence as it applied under the common law to industrial accidents.
3. Why are employees in certain lines of endeavor, or in small establishments, exempted from the workmen's compensation laws?
4. "It has been estimated that injured workers still bear more than half the money cost of their injuries." Does this seem fair to you? What arguments are there in favor of requiring the worker to carry part of the cost? What are the disadvantages?
5. Which of the following would you classify technically as unemployed:
   a. A student attending college?
   b. A farmer who works for himself?
   c. A housewife?
   d. A man on summer vacation?
   e. A man who has retired?
   f. A dental technician who has been unable to find dental technician's work but who has been offered a job clerking in a five-and-ten-cent store?
   g. A dental technician who is not working but who could have had a dental technician's job if she had been willing to work for half the normal pay?
   h. A man who has been fired for insubordination and who has been unable to find new work?
6. What are some of the causes of normal unemployment?
7. If you are willing to admit that, in practice, we can never expect to have every single able person working all the time, what percentage of normal unemployment would you think could be tolerated? Your answer will have to be largely guesswork; but think in these terms: How much of the time can the typical or average worker afford to be unemployed—one day a year? One week a year? One month a year? How much?
8. Explain how private or public employment offices reduce unemployment.
9. Why is unemployment compensation under the jurisdiction of the states rather than the Federal government in this country? Would it not have been easier for Congress to pass one uniform law rather than to try by indirection to influence state laws?
10. In most states employers are taxed 2.7 per cent of their pay rolls as an unemployment insurance tax. Where did such an unusual figure as 2.7 per cent arise?
11. Explain why the Federal law requires the payment of unemployment benefits through public employment offices in all approved state plans. Why not some other agency for payment?
12. Suppose a man who has been working in an industry where he is covered by unem-

ployment insurance loses his job and wants to receive unemployment benefits. What does he have to do? Why is each restriction there?

13. What is the difference between old-age pensions and old-age insurance?

14. Is old-age insurance a type of insurance intended to protect against having an old age?

15. If a worker covered by old-age insurance starts to work under the act at the age of forty and retires at the age of seventy, having received an average income of $150 per month throughout the 30 years, what will be his retirement income if he is single? If he has a wife aged sixty-seven?

16. What, as you understand it, are the difficulties with investing the reserves of old-age insurance?

17. What is the meaning of the statement that "social security programs deal with minima"?

18. Opponents of legal minimum wage legislation point out that such legislation is likely to increase unemployment. Do you think that the statement is correct? If not, why not? If so, how would you answer the argument?

29. ECONOMIC PROBLEMS OF AGRICULTURE

29 • 1. Position of Agriculture in the Economy. It may seem peculiar, in a general textbook, to single out one industry, such as agriculture, for special treatment; but agriculture holds such an unusual place in the American economy as to make some understanding of its problems a prerequisite to an understanding of the general over-all economic situation. In the ordinary sense, to be sure, agriculture is not a single industry, but a combination of varied ones. There is as much difference between dairy farming and the raising of sugar beets as there is between the operation of a blast furnace and a textile mill. Yet, certain of the economic problems of agriculture are common to all its parts, making it reasonable for us to examine its problems as a unit. And these problems have been so persistently troublesome that a whole body of specialized subject matter known as "agricultural economics" has come into being in connection with them, and a vast amount of political and legislative attention has been given to them. The "farm problem" has been constantly in the spotlight on the American stage, and for the past 25 or 30 years has held the attention of our major statesmen as well as of technically trained economists.

The importance of agriculture arises in part from the nature of its products—largely basic raw materials of food and clothing. A considerable proportion of those commodities, which we call "the necessities of life," come originally from agriculture. Agriculture is what we called in Chap. 10 a "primary industry," and among the primary industries it is by far the most important. Those people who think of the farm problem as something far removed, in which they have no interest, overlook the fact that agriculture touches them every time they sit down to a meal, that it produces the raw materials for their clothing and their houses, and that they are never far removed from the influence of its problems.

It is not alone the basic importance of the products that forces us to be interested in agriculture. The very number of people engaged in that pur-

suit gives it a peculiar place in our economy. To be sure, this number is not proportionately so large as it used to be. In colonial days some 90 per cent of our population was engaged in agriculture.[1] This means that the services of 9 men were then required to produce the raw materials for food and clothing for every 10 men, leaving only 1 of the 10 men free for other occupations. But as time has passed, agricultural production has become much more efficient. New and more fruitful varieties of crops, more productive breeds of livestock, the introduction of farm machinery, scientific developments in the fields of fertilization and pest control, these and many other changes have brought an industrial revolution in agriculture to match that in manufacturing. And with these changes a man's output has been increased

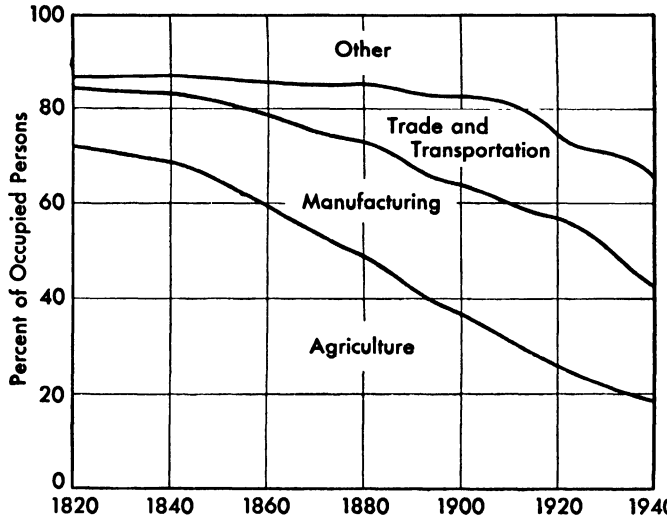


FIG. 29 · 1.—Distribution of the employed population among the various industries. United States, 1820-1940. The data for 1820-1920 are based on P. K. Whelpton, Occupational Groups in the United States, *Jour. Amer. Stat. Ass'n.*, vol. 21, p. 340, (1926). The data for 1930 and 1940 are from the census, and while not strictly comparable, serve to indicate the general trend.

many times, so that a smaller and smaller proportion of our people need be engaged in agriculture, and more and more of them can be set free for other kinds of work. This change is startlingly evident in Fig. 29 · 1, which shows the percentage of our entire employed population in each of several lines of endeavor since 1820. Today, roughly a fifth of our employed population is in the field of agriculture. One man produces enough to support himself and four others, leaving the other four free to enter other occupations and to satisfy types of wants that would have to go unsatisfied were

[1] WRIGHT, CHESTER W., "Economic History of the United States," p. 86, McGraw-Hill Book Company, Inc., New York, 1941.

the farmer less efficient. Those people in other fields who look askance at farming as being a "rube" occupation "out in the sticks"—as being somehow a second-rate field of work—have not realized that the only reason that they themselves are not farmers is that the people in agriculture are doing the job well enough to release them for less vital kinds of work.

Agriculture is important to us because of the fundamental nature of its products, and because of the fact that one out of every five employed persons works in this field. But it is also important because of the buying power of farm families, and the effect of their demand in the industrial and the consumer-goods markets. In 1943, the total national income was estimated at 149.4 billion dollars. Of this, 13.5 billion or 9 per cent went to

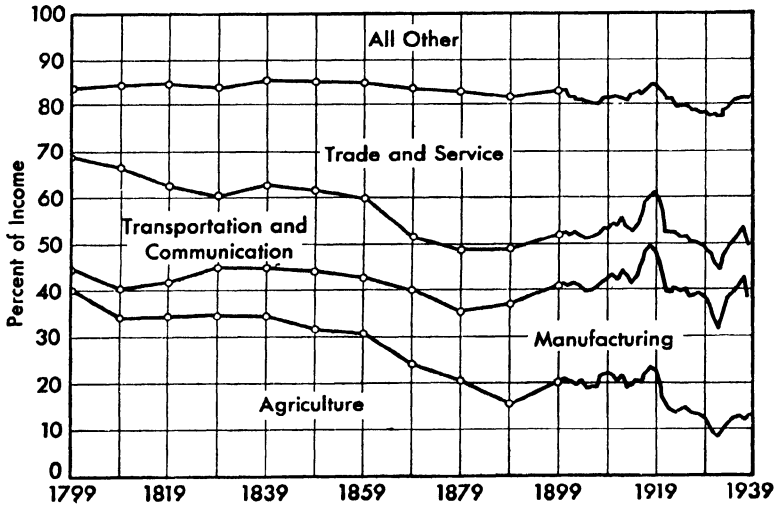


FIG. 29 • 2.—Percentage distribution of national income of the United States by industries, 1799-1938. (*National Industrial Conference Board, Studies in Enterprise and Social Progress, N. Y.*, 1935, *p.* 96. *Figures before* 1899 *are by decades, and thereafter are for single years.*)

agriculture.[1] Nine cents out of every dollar of buying power were in the hands of farm families, and this buying power was felt in the market for stoves, automobiles, dress goods, chinaware, and throughout the economy. Anything that seriously affects the buying power of farmers is reflected in the demands for all sorts of industrial products. The historical changes in the buying power of agriculture since 1799 are shown in Fig. 29 • 2. Here we see that, although agriculture has shrunk from its post-Revolutionary War position of receiving 40 per cent of the nation's income to a present position of receiving roughly one-tenth of the income, it is still of such decided importance as to require our careful attention in estimating the demands for the products of all other industries. Although the income of people in agri-

[1] Figures are from the "Statistical Abstract of the United States, 1944-1945," p. 404.

culture is nowadays only a half to a third as large as the incomes of those in manufacturing, nevertheless agricultural income is far greater than that of any particular manufacturing enterprise.

Figures 29 • 1 and 29 • 2 are drawn to the same scale, and the student will find a comparison of the two figures instructive. The proportion of the employed population engaged in agriculture, as shown in Fig 29 • 1, is larger in every year than the proportion of the national income received by people in agriculture. This indicates that, while the total income received by farmers is large, the average income is low. In fact, in recent years there have been roughly 10 million persons in the labor force in agriculture, and the total agricultural income has run in the neighborhood of 6 to 8 billion dollars,

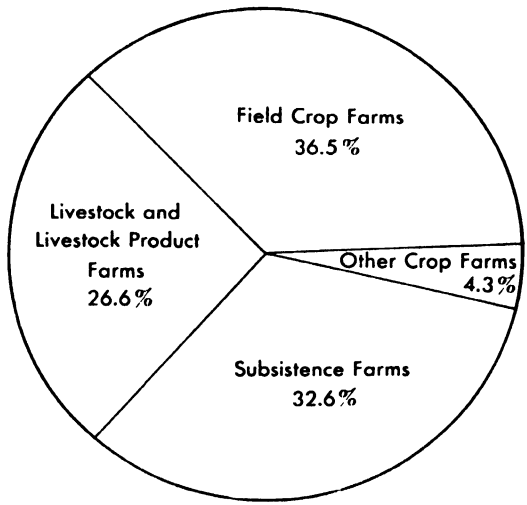


FIG. 29 • 3.—Types of farms in the United States, 1939. (*Data from Statistical Abstract of the United States*, 1944-45, *p.* 609.)

which means that the average income per agricultural worker (including farm owners as well as hired hands) is only about $600 to $800 per year. If we add to this the fact that rural people on the average have poorer medical and educational services than the people of the cities, and often also poorer housing,[1] we can see that the real incomes are low indeed. To some extent these disadvantages may be offset by better diets, including fresh food from the farm, and by a more independent type of life; yet on the whole it must be admitted that farm incomes are, on the average, low. More will be said on this point later.

The nation's farms are themselves of many different kinds—wheat

[1] At the time of the 1940 census there were 5.8 million farm houses, of which only 2.0 million were served with electricity, only 1.5 million had telephone service, etc. See "Statistical Abstract of the United States, 1944-1945," p. 641.

farms, dairy farms, cotton farms, etc., but we can give a rough idea of the relative importance of various types of farming by Fig. 29 • 3, which shows the situation in 1939. This figure indicates that roughly a quarter of our farms get their major income from livestock and livestock products, while something over a third rely mainly on field crops, such as corn, wheat, potatoes, cotton, etc. A very small proportion of the farms, less than 5 per cent, rely on other crops, such as tree crops, vegetables, and forest products. About a third of all the farms are "subsistence farms," which are not run on a commercial basis of raising products for sale but which raise products primarily for home consumption.

29 • 2. Peculiarities of the Agricultural Industry. Agriculture is one of the best examples of a small-scale industry. This is indicated by the figures of Table 29 • 1, which show the average number of workers in various industries in 1935. Here we see that the average farm employs but two workers, including the owner of the farm himself and any members of his family who work on the farm in addition to hired help. This places farming, as far as size of business is concerned, with the firms offering services, such as barber shops, shoe-shine parlors, tailor shops, watch-repair shops, etc. The reason for this small size in agriculture is a simple one. The manager of a factory is able to make decisions that last for some time. He can, for example, adopt a scheme of operations which assigns men to certain kinds of work which they are to do day after day for months and perhaps years on end. On the farm, however, the summer's work and the winter's work differ, the work of a sunny day and that of a rainy day differ, and the work of the morning often differs from that of the afternoon. The manager of a farm plans as carefully as he can and lays out his scheme of operations; but every change in the weather makes it necessary to change his plans and to reassign his workers and his equipment. The manager of an industrial plant may be able to supervise effectively the work of 100 subordinates; but farm management is so much more complex that the typical farm manager is unable to supervise effectively the work of more than a very few men. This is merely another way of saying, what we have pointed out before, that in farming, as in most primary industries, the point of diminishing productivity is reached early, and the ordinary business operates under conditions of increasing cost. The economies of large-scale operation disappear very early, and the big farm has no advantage in efficiency over the small one.[1]

[1] Here again it must be emphasized that different kinds of farming are very different. In some lines, large-scale corporation farming seems to show some promise. The average canteloupe farm in the Imperial Valley of California produced $119,000 worth of canteloupes in 1939, and the average grower of shade tobacco in the Connecticut Valley district produced $67,000 worth of tobacco in 1939. These might be classed as "big business," or, at least, not as tiny businesses; but they are definitely the exception in agriculture. Figures in this footnote are from *T.N.E.C. Monograph* No. 23, p. 12.

As a sort of corollary to the fact that agriculture consists of a vast number of relatively small units, the competitive forces have been much more pro-

Table 29·1. Number of Businesses, Numbers of Workers, and Average Workers per Enterprise in Various Industries, United States, 1935*

Activities	Enterprises	Number of workers	
		Average for year, thousands	Average per enterprise
Total	9,334,437	34,103	3.7
Agriculture	6,242,500	12,408	2
Nonagricultural	3,091,937	21,696	7
Manufacturing	162,000	8,536	52.7
Mining and quarrying	10,135	864	85.2
Transportation	63,596	1,435	22.6
Class I railroads	144	994	6,903.3
Public utilities	46,478	626	13.5
Construction	69,838	478	6.8
Trade	1,627,411	6,652	4.1
Wholesale	176,756	1,375	7.8
Retail	1,450,655	5,277	3.6
Financial	119,723	913	7.6
Banks	19,581	266	13.6
Service	891,098	1,770	2
Personal services	355,976	661	1.9
Business services	29,859	198	6.6
Repair services	97,654	172	1.8
Custom industries	85,270	160	1.9
Miscellaneous services	5,949	23	3.9
Lawyers, physicians, surgeons, dentists	316,390	555	1.8
Miscellaneous activities	101,658	737	7.3
Advertising agencies	1,212	14	11.2
Amusement places	37,677	189	5
Private cemeteries	2,060	15	7
Hotels	28,822	316	11
Office buildings	9,061	87	9.6
Professional service	9,045	50	5.5
Radio broadcasting	564	15	25.8
Tourist camps	9,848	16	1.6
Warehousing	3,014	34	11.3
Private water works	355	3	9.8

* This table is taken from "Studies in Enterprise and Social Progress," p. 234, National Industrial Conference Board, New York, 1939.

nounced in agriculture than in most industries—the monopoly element has been much weaker. This has been particularly true if we limit ourselves to

the economic problems on the farm itself, and do not consider the problems that arise in connection with the marketing, storing, and processing of farm products or in connection with the production of farm equipment. Even in the sphere of actual farm operations, there has perhaps been some tendency to increase the size of business, especially as farms have become more mechanized. But the typical farm product is produced under conditions as nearly those of complete competition as we are apt to find in our economy, except in those cases where the government has stepped into reduce the degree of farm competition. As we shall see later in this chapter, this governmental reduction of competition has been rather common in the United States in the years between the two World Wars, and it shows no immediate signs of disappearing.

29 • 3. Characteristics of Demand for Agricultural Products. While we are considering the peculiarities of the agricultural industry, we should pay some attention to the characteristics of the demand for farm products. Most agricultural products are raw materials of food and clothing, and hence would be classed as necessities of life. In fact, between 90 and 95 per cent of them are used either as food or as feed, the important exceptions being such things as cotton, wool, and tobacco. But the demand for necessities tends to be inelastic.[1] Changes in price have relatively little effect on quantities purchased by consumers; and, conversely, changes in quantities offered for sale have relatively great effects on prices.[2] Thus we find in many cases that the farmer's total income from a large crop is smaller than his total income from a small crop. This is by no means uniformly true, of course, since agricultural products differ considerably in the elasticity of demand. Although we would expect a very inelastic demand for food, we would expect a greater elasticity in the demand for meat, and a still greater elasticity in the demand for beef.[3] In many cases, one agricultural commodity is to a considerable degree a substitute for another, and beef, pork, lamb, and poultry can be interchanged to a considerable extent. These possibilities of substitution increase the elasticity of demand and raise important and inter-

[1] See Chap. 7.

[2] The textbooks seem ordinarily to treat these two statements as being of such a nature that if one is true the other must be. As a matter of fact, it is quite possible for the change in dependence of the variables to alter the entire nature of the functional relationship. Price control and quantity control may have very different effects, especially since the attitude of the public toward them is very different. To say that demand is inelastic actually means that changes in price have relatively little effect on the quantity taken, and implies nothing about the effect on price of alterations in the quantity offered for sale. Yet this point, while of fundamental importance in advanced price theory, where, by the way, it seems strangely neglected, need not be labored in an elementary treatment, and the sentence above can stand as being roughly true on purely empirical grounds even if there is no need for it to be true on theoretical ones.

[3] For an explanation of this fact, see Sec. 7 • 6.

esting problems of cross-elasticity.[1] If we consider individual commodities, we may find some elastic demands. Thus the National Industrial Conference Board, after studying all available statistics since the 1860's, concluded that, "The agricultural history of the last sixty or seventy years in the United States indicates that there is no basis for the assumption that reduced crops of the leading products tend to increase farmers' income." [2] This is another way of saying that the data give no basis for the assumption that the demands for agricultural products are inelastic. Despite this careful study, it is only fair to say that the preponderance of what evidence we have from many studies seems to indicate that the demand for farm products is less elastic on the average than the demand for manufactured products.

Trends that we have already discussed have brought recent changes in the demand for agricultural products. In the first place, the reduction in the rate of growth of our population, mentioned in Chap. 12, has had an obvious effect. The demand for farm products depends to a considerable extent on the number of consumers, which is equivalent to saying that it depends on the size of the population. But not only has the rate of population growth slowed down; at the same time there have been important changes in human wants. There has been a noticeable shift in preferences for food, with a tendency to eat more fruit, green vegetables, and milk, and to rely more for energy on the highly concentrated sources such as sugars. With the change in types of occupation, there has been a reduction in the actual weight of food eaten per person and almost certainly also in its caloric value. While these changes have been taking place in diet, there have been also changes in dress which have had important repercussions in agricultural markets. The total yardage of cloth worn by women today is far less than that worn by their great grandparents, and the kinds of material have also changed. Clothing is lighter in weight and of reduced amount.

Yet one should not overemphasize these changes in the demand for agricultural products. They have occurred, and they have had important effects; yet, in general, it can probably be said not only that the demands for farm products are relatively inelastic, but also that they are relatively stable; that is, they change only slowly with the passage of time. Food habits change, but not rapidly. The farm products on the market today are much more nearly like those of three or four generations ago than would be true of most manufactured goods. If we are to compare and contrast agriculture with other industries, rather than with some unchanging standard, we shall have to say that the demands for farm products change but little and then slowly.

29 · 4. Characteristics of the Supply of Agricultural Products. The farmer

[1] See Sec. 7 · 8.

[2] "Studies in Enterprise and Social Progress," p. 198, New York, 1939.

has less control over the quantity of his output than one ordinarily finds in other industries. Although he can, of course, decide how many acres to plant, the amount of the product depends to a very considerable degree on natural forces beyond his control: the temperature, the rainfall, the prevalence of insect pests, etc. The extent to which these forces influence agricultural production is seldom appreciated by those without farm experience. The yield of corn in the United States was 15.7 bushels per acre in 1934 as compared with 24.0 the year following. This was an increase of 53 per cent in one year. The yield per acre of wheat varied from a low of 12.2 to a high of 19.8 bushels per acre in the decade 1935 to 1944. In the same decade the yield per acre of potatoes varied between 109 and 139 bushels per acre; the yield of cotton between 185 and 295 pounds per acre, etc. These wide variations in output were to some extent the result of human planning, being caused by variations in cultural practices; but to a very considerable degree they were fortuitous. The farmer may plan his production program carefully in an effort to select the most profitable volume of output, but when harvesttime comes he need not be surprised to find that his output is 10 or even 25 per cent above or below that which he had planned. The manufacturer, on the other hand, is able to control his output within much narrower limits. The automobile manufacturer can decide at the beginning of the year to turn out 5000 cars, and he can then adjust his program in such a way that he turns out exactly that many if he wishes.

Not only is the outcome of the production process uncertain in agriculture, but the process itself is often a very long one as compared with those of industry. To be sure, there are great differences between manufacturing industries in this regard, but in most cases a manufacturer who decides to increase his output can make the increase effective within a rather short time. Within a few weeks or months from the time his decision is made, the increased flow of goods is reaching the market. In agriculture it is seldom possible to alter output in less than a year; often it takes much longer. When the potato crop has been planted, relatively little can be done to increase the yield for a year. If dairy farmers want to expand their output, they have to wait several years to raise additional cows. If apple farmers wish to expand, they may well find that it takes 8 or 10 years to plant an orchard and bring it into bearing. The long period of time that elapses between the making of production decisions and the effective culmination of those decisions in the field of agriculture makes maladjustment likely.

Another factor influencing the supply of farm products is that many of them are perishable commodities. Of course, there are many exceptions, but fresh fruits and vegetables, meats, milk, etc., require special and expensive handling for their preservation, and in most cases must be sold very soon after they are once ready for market. We learned in Sec. 14·11 that the

supply of perishable products is usually inelastic, and the immediate short-run supply[1] of many agricultural products shows the inelasticity that one would expect. The combination of this inelastic supply with the inelastic demand, mentioned in the preceding section, has characteristic effects on the behavior of agricultural prices which we would expect from our study of Sec. 15 • 8. We can illustrate it effectively by means of Fig. 29 • 4. Part *A*

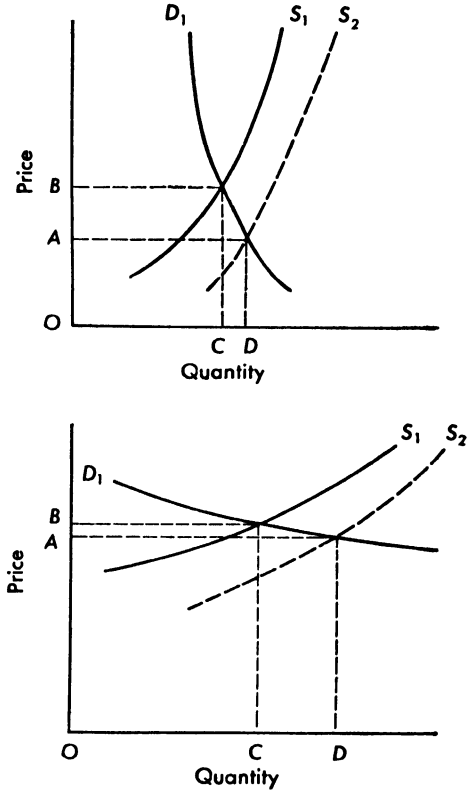


FIG. 29 • 4.—Effect of elasticity of demand and supply on changes in price and output.

of this figure, at the top, shows an inelastic demand curve D_1, and an inelastic supply curve S_1. With this demand and supply, farmers will sell a quantity OC at a price OB. If, now, the supply increases somewhat (as, for example, because of an unexpectedly good harvest) to the broken curve S_2, leaving the demand where it was before, farmers will sell a quantity OD at a price OA. The change in supply has brought a major change in price and only a minor change in the quantity sold. But in part *B* of Fig. 29 • 4, we have a

[1] The student who is hazy as to the distinction between "long-run" and "short-run" supply should reread Sec. 14 • 11.

combination of elastic demand and elastic supply. Here, if the supply changes from S_1 to S_2, the price drops only slightly from OB to OA, while the quantity sold on the market increases sharply from OC to OD. The student will discover by drawing similar diagrams that the same type of results would have been found had we shifted the demand curves rather than the supply curves. *Where demand and supply are both inelastic, changes in either demand or supply bring major price changes and only minor changes in the volume of sales; but where demand and supply are both elastic, changes in either demand or supply bring major changes in the volume of sales and only minor changes in prices.* The characteristics of the demand for and the supply of agricultural products are such that we tend to get major changes in prices and only minor changes in output from year to year. The manufacturing concern, especially if it has some degree of monopolistic control of the market for its product, often follows a policy of "administered prices," holding the price constant or approximately constant from year to year and allowing considerable fluctuation in the volume of business. The farmer, having no control over his highly competitive and unorganized market, and being faced with considerable inelasticity both in demand and supply, follows willy-nilly the other procedure of widely fluctuating prices and only narrowly fluctuating quantities.

Many authors have suggested that the supply of farm products is also materially influenced by a supposed large element of fixed costs in agriculture. They suggest that the investment in land, buildings, and farm machinery is fixed, and that the farmer therefore neglects in the short run most of his cost and considers only the small element of variable cost involved. Even his labor, they would argue, is largely a fixed cost, since if he does not put it into production there is nothing else that can be done with it. Now this invoking of fixed-cost theory, familiar to us from our study in Chap. 17, to explain the agricultural situation is to some extent pertinent, but only to a minor degree is it pertinent in the sense just suggested. When the economist speaks of fixed or variable costs, he is thinking of the cost of producing particular products, and in this regard there is a smaller degree of fixed cost in agriculture than in most lines of manufacturing or other business. The farmer who is growing wheat can usually use the same land and much of the same machinery and equipment and many of the same buildings to grow corn or oats. The farmer who is growing celery can use the same land and much of the same equipment to grow string beans or lettuce. To be sure, these different crops have somewhat different requirements for soil, climate, machinery, etc.; but it is usually far easier for the farmer to shift to another type of production than it is for the manufacturer. The owner of a textile mill cannot use his looms for the manufacture of automobile hub caps. If he does not use them in weaving cloth, he can not use them at all. But a farmer who finds one crop unattractive can, if it seems to him wise,

usually shift many of his productive resources to others. The investment in these shiftable resources cannot properly be called a fixed cost of producing any of the individual crops. In this sense, many farmers actually have a smaller proportion of fixed cost than does the average producer in other fields.

The important point to make is not so much one of fixed costs, as that, when farming is depressed, the farmer cannot meet his problem by reducing the employment of labor—by laying off men. A very large proportion of the farm labor supply is furnished by the farmer himself and the members of his family.[1] The manufacturer, who is finding it unprofitable to maintain his level of output, can reduce output by laying off workers. The farmer, faced with the same problem, has no one to lay off but members of his own family. The manufacturer passes a part of the burden of depression to his employees. The farmer has no one to whom he can pass it. He is forced to carry it himself. It is not so much the costs of producing particular farm crops that is fixed as it is the farmer himself and his family who are fixed. Their business is also their home and their way of life, and they are less mobile than are workers in many other occupations. It is, consequently, very difficult for them to reduce output when prices drop and profits shrink to a trickle and disappear. This, of course, is an element more important in influencing long-run supply than short-run supply.

29·5. Foreign Trade in Agricultural Products. As we discovered in Chap. 23, the early export trade of the United States was almost exclusively in raw materials and farm products and, although there has been a definite tendency to shift toward more and more reliance on manufactured goods, nevertheless farm products still make up an important part of our outgoing foreign commerce. In the 5-year period, 1931 to 1935, approximately 45 per cent, and in the following 5 years roughly 30 per cent of our exports, were farm products. For the moment we are more concerned with the importance of the export business to the agricultural industry than we are with the importance of the agricultural industry to the export market; we care more about the percentage of agricultural products that go abroad than about the percentage of the products going abroad which came from farms.

Naturally the importance of the export market differs for different products. Some are consumed exclusively in the domestic market, while some are shipped in large part abroad. Figure 29·5 shows the situation for several important farm products for a period of 70 years. Although the situation has varied from year to year and although wars have twice almost cut off foreign markets for the time being, nevertheless we see that roughly 50 to 60 per cent of our cotton, 40 per cent of our tobacco, a quarter of our wheat,

[1] When the Census Bureau checked farm employment for the week of Mar. 24-30, 1940, they found a total of 1,753,000 hired hands working on farms as compared with 7,941,000 members of farm families working. Data from "Statistical Abstract of the United States, 1944-1945," p. 633.

and a sixth of our pork and its products have gone into the export trade. Surely farmers who produce significant amounts of these products find themselves largely dependent on the world market and on the international situation—on tariff acts, trade quotas, monetary policies, and war. But these commodities are not typical. A better idea of the general situation is given by Fig. 29 • 6, which shows similarly the proportions of all farm products sold at home and abroad. Here we note that an average of perhaps a seventh of our whole farm output is taken by the foreign market.

The student's first reaction is apt to be that a seventh of the output is a small and unimportant share—that the six-sevenths sold at home are far more important than the one-seventh sold abroad. To be sure, he will be

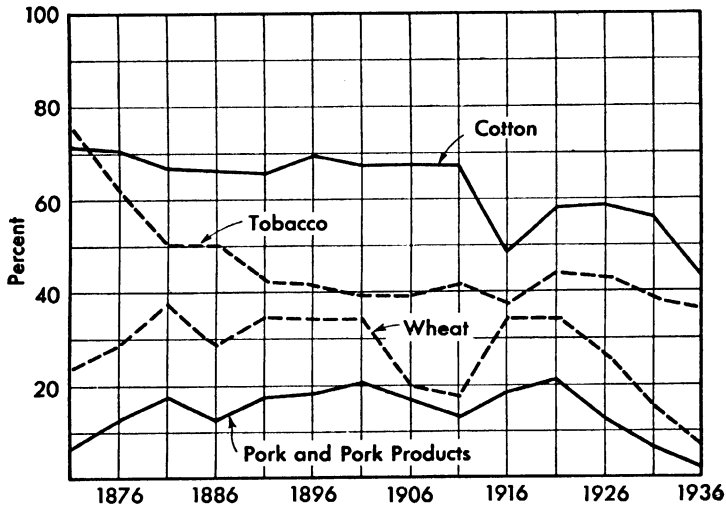


FIG. 29 • 5.—Percentages of gross income from various farm products which came from the export market, United States, 1871-1936. Each value given on the chart is a 5-year average centered at the middle year. For example, the values for 1871 are averages for the 5 years 1869-1873. (*Data are from T.N.E.C. Monograph No.* 23, *p.* 6.)

right in this latter thought. But the fact that the foreign sales are but a sixth as important as the domestic ones does not mean that the foreign sales are unimportant. We have learned on several occasions that it is not the total or the average output that is important in an economic sense as much as the marginal output. Even a small marginal change may be of great importance in the total situation. A very slight increase in output may affect seriously the price of the entire output. When the foreign market is as important as it is in the case of many farm products, we must realize that changes in foreign demand are of vital importance to domestic growers even though but a seventh, or even less, of the output is involved. The price received by domestic producers of agricultural commodities is to a very considerable

extent a world price, made in world markets, in competition with commodities produced in all parts of the globe. Some perishable commodities, such as fresh milk, have but local markets and are effectively insulated from world conditions. But most of our farm products are staple commodities used all over the world and produced in widely scattered areas, and as a result our farmers are peculiarly dependent on and interested in world conditions. It is a strange commentary on popular understanding of economic matters that our rural communities have been to a considerable extent our centers of economic and political isolationism when, as a matter of fact, their dependence on economic and political conditions abroad is far above that of the average citizen.

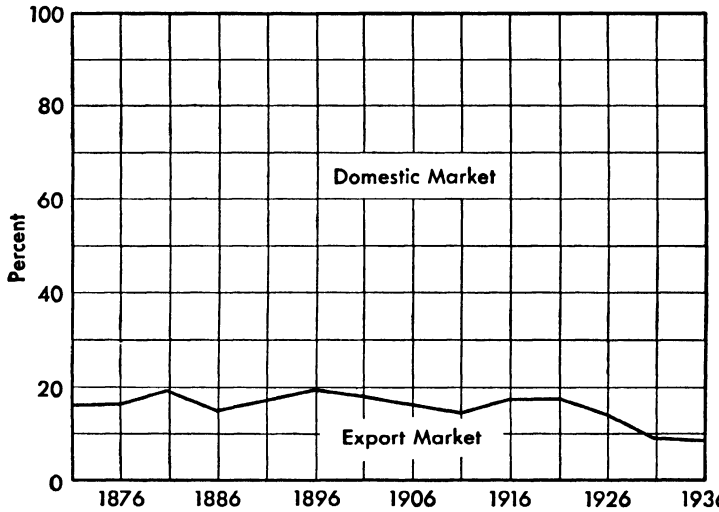


FIG. 29 • 6.—Relative importance of the domestic and the export market for farm products, United States, 1871-1936. Each value given on the chart is a 5-year average centered at the middle year. For example, the value for 1871 is the average for the 5 years 1869-1973. (*Data are from T.N.E.C. Monograph No.* 23, *p.* 6.)

Especially is this problem an important one when we recall the inelastic nature of the demand for farm products mentioned in Sec. 29 • 3. Because of this lack of elasticity, even small changes in the quantity of goods thrown on the market bring about major changes in prices.[1] Therefore, even small changes in the foreign demand for these goods bring repercussions of tremendous importance in the domestic market. A new tariff or any other new trade policy abroad may affect the incomes of American farmers out of all proportion to the actual physical quantity of goods involved. The data

[1] Reference should perhaps again be made to the footnote on p. 692, where apology is made for the laxity of this statement; again we can excuse it with the empirical fact that this dependence of price on quantity does seem to exist, even if it is not technically what we mean by elasticity of demand.

of Figs. 29 · 5 and 29 · 6 show vividly the effects of business depression and of war, and it is yet too early to judge whether or not the economy of the post-war world will allow the development of free trade to such an extent that the old markets can again be regained by our farmers.

29 · 6. Problem of Farm Tenancy. Each country has developed its own peculiar system of agriculture. In England much of the agricultural land is in large estates and is cultivated by hired hands. On the continent of Europe, agriculture is commonly practiced by peasants who operate their own small farms without benefit of modern farm machinery. In the United States, we have usually thought of our agriculture as being typified by the independent farmer who owns his own farm and works on it himself, often with the help of his family. It has commonly been felt that this system of operation by the owner has certain advantages, both economic and sociological. The farmer who owns his own land is independent in mind and in character. He has every incentive to maintain and to improve the productivity of his farm and its equipment, while the tenant may find it advantageous to get what he can out of his hired farm in the immediate present without much regard to its deterioration in long-run value.

In a country where land is plentiful and cheap, as it was in the United States at and before the time of the Civil War and perhaps even up until the beginning of the twentieth century, there is no reason for tenancy. Men prefer, other things equal, to own their farms. As land becomes scarce and expensive, ownership becomes more and more difficult. The young man who wants to start out in farming finds that he has not the capital to buy one, and he hesitates to incur the large debt that would be necessary if he bought it on credit. As a result, he tends to start off as a tenant, renting a farm from someone else. Especially this is true where, as in England, land ownership has been looked upon both as a fine investment and as a mark of social distinction.

Despite the fact that farm ownership has been considered more desirable than farm tenancy in the United States, and despite the fact that various government projects have sought to stimulate ownership as opposed to renting, the proportion of tenancy has increased regularly ever since the first census of tenancy was taken in 1880. At that time, roughly a quarter of the farms of the United States were operated by tenants, and three quarters by owners. In more recent years the proportion operated by tenants has been as follows:

Year	Per cent
1910	37
1920	38
1925	39
1930	43
1935	42
1940	39

Thus, for the country as a whole about two farms out of every five are operated by tenants. But the situation is by no means uniform in different parts of the country. Tenancy is especially common in the southeastern part of the United States, and actually two-thirds of Mississippi farms were operated by tenants in 1940. The high tenancy rate runs up through the rich agricultural Midwest, where land values are high, but falls sharply as one approaches the West coast or the Northeast. The situation as it existed in 1940 is shown roughly by the "contour" map of Fig. 29 • 7.

Rural sociologists have spent a good deal of time in studying what they call the "agricultural ladder"—the steps by which people typically reach the position of farm ownership. In the United States, it is more and more becom-

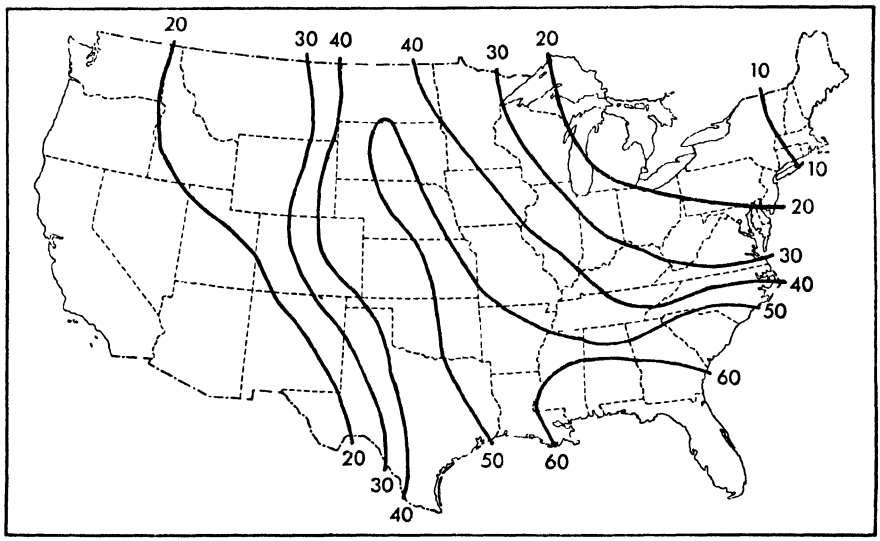


FIG. 29 • 7.—Prevalence of farm tenancy in the United States, 1940. The figures show the percentages of farms that are operated by tenants in the localities through which the lines pass. Areas between lines have tenancy rates between the figures on the enclosing lines. Only the broad general picture is shown, with no attempt to indicate minor local variations. For example, tenancy in Florida is actually rather low, although not so indicated on the map. The high tenancy rate in the Southeast is evident, as is the low tenancy rate of the Northeast and of the far West. (*The map is based on data from the Statistical Abstract of the United States,* 1944-45, *p.* 620.)

ing true that tenancy is a step on this ladder intermediate between that of farm laborer and that of farm owner. The young country lad starts out as a hired hand on his neighbor's farm. He gets married and rents a farm of his own, saving his money until he finally is able to buy a farm. And then, as he reaches old age, he often retires and rents his farm to a new tenant, starting the cycle again. It has been common to decry the tendency for farm tenancy to increase; yet it is a natural and inevitable situation in a society

in which land values are rising. There is little advantage in wasting energy wailing for the return of the "good old days" when almost every farmer owned his farm. Instead, we should be spending the energy in finding ways to improve the relationship between landlord and tenant—in working out types of leases and forms of credit that minimize the evils of tenancy, and in learning more about the other economic and sociological problems involved. Farm tenancy seems to be with us to stay.

29 • 7. The Farm Problem: Historical Background. It is impossible for us to take time here to discuss in detail the history of agriculture in the United States, even though some superficial familiarity with the major developments is necessary for any understanding of the current situation. Suffice it to say that the United States has developed from a nation almost entirely agricultural to one in which agriculture, although still a major type of economic activity, is but one of many important lines of endeavor.[1] During our early history, the government encouraged by various means the settlement of the public domain, and land was "tamed" and put under the plow at a rate that now seems hardly possible. There was some waste involved in the policies followed, but they did succeed in their main objective, which was that of "manufacturing farms." During this period of settlement, agriculture suffered its ups and downs with the swings of the business cycle, and especially toward the close of the nineteenth century farmers experienced a protracted period of acute economic hardship which had political repercussions in the granger movement, the populist movement, the free silver movement, etc. Repeated increases in farm efficiency made it possible to reduce continually the proportion of the population on farms without reducing the per capita production of farm products, and improvements in transportation made possible a greater and greater degree of regional specialization of production which also increased productive efficiency. With the growth of population and with the concentration of that population into large cities, agriculture became commercialized. Instead of producing products mainly for home use, the practice of production for sale became common. And with the other changes, agriculture was rapidly becoming a skilled occupation, not under the control of untutored peasants, but guided by men who needed and who increasingly got an understanding of chemistry, nutrition, genetics, botany, anatomy, and economics.

29 • 8. Present Farm Problem. Although present-day agriculture is still greatly influenced by the changes just described, nevertheless the real farm problem about which we speak today stems back to the time of the first World War. The outbreak of the European aspects of this war found the agriculture of the United States in a period of fairly stable prosperity. With the beginning of hostilities in Europe, there was a sudden increase in the

[1] See again Figs. 29 • 1 and 29 • 2.

demand for our farm products, since the warring nations were no longer able to supply their own. With our own entry into the conflict the demand was increased, and at the same time the number of men available for farm work was sharply curtailed. Naturally, the prices of most farm products increased materially, and farmers found it profitable to increase their outputs. Despite a reduction of between 7 and 8 per cent in the number of persons working on farms, the acreage in crops expanded by about an eighth and the number of livestock by perhaps a sixth or a seventh. Increased earnings on farms, and the prospects in the minds of enthusiastic farmers that these earnings would be maintained or even increased, brought about a boom in land values. The average value of farm land per acre for the entire country rose from $32.40 in 1910 to $57.36 in 1920, an increase of over 75 per cent. Between 1915 and 1920, farm real-estate values per acre doubled in Minnesota, Kentucky, and Tennessee; they rose 118 per cent in North Carolina, 107 per cent in Louisiana, 125 per cent in Mississippi, 135 per cent in Arkansas, and 144 per cent in Georgia. These are the highest figures, but every state without exception showed an increase, and for the United States as a whole the increase amounted to 65 per cent, or an increase of about two-thirds in 5 years.[1]

With these inflated land values men went farther and farther into debt to buy farms. Farm mortgages became heavier and heavier, increasing by about 75 per cent between 1916 and 1920, until over a third of all farms carried mortgages, and 29 per cent of their value was covered by mortgage. Mortgage interest payable on farms increased 83 per cent between 1915 and 1920, and during this same 5-year period, while the interest burden was becoming heavier and heavier, taxes on farms property increased 96 per cent. Had we taken our base for comparison a year or two earlier rather than in 1915 we would have found that the burden of both mortgage interest and taxes more than doubled. During this period farmers were enjoying the exhilarating thrill of inflation—of a false prosperity.

In 1920, came the headache of the morning after. The general level of wholesale prices had reached a peak in May, 1920, with the price index at 244 (on a base of 1910-1914 = 100). By the following May, it had dropped to 141, a reduction of over 40 per cent in a year. "No such drop in prices," say Warren and Pearson,[2] "had occurred in the United States since the Revolutionary War." Business was hard hit. Bankruptcy was common. There was a short, sharp period of unemployment. But, perhaps most important of all to the farmers, was the fact that as European countries, recovering from the war, slowly brought their own farms back into cultivation, the export demand for our farm products was sharply curtailed. Our agricultural ex-

[1] Data from "Statistical Abstract of the United States, 1944-1945," p. 605.

[2] WARREN, G. F., and F. A. PEARSON, "Prices," p. 359, John Wiley & Sons, Inc., New York, 1933.

ports, which had amounted to 3.9 billion dollars in the fiscal year ending June 30, 1920, fell in the three following fiscal years to 2.6 billion, 1.9 billion, and 1.8 billion. As we noted earlier in this chapter, even when the export demand is but a small percentage of the total demand, it is apt to have a major effect on prices. And so it did. The index number of prices of farm products, which had stood at 211 (on a 1910-1914 base) in 1920, fell to 124 in 1921.[1] But the change is more easily visualized in the case of particular commodities. Table 29·2 shows the prices received by farmers for various farm products in May, 1920, and in May, 1921, and also shows for each commodity the percentage by which prices fell during the year. On the

Table 29·2. Farm Prices of Certain Agricultural Products, United States ,May, 1920, and, May, 1921, with Percentage Fall in Price of Each*

Commodity	Farm price, cents		Percentage drop in price
	May, 1920	May, 1921	
Wheat, per bu.	251	111	56
Corn, per bu.	170	60	65
Apples, per bu.	285	142	50
Potatoes, per bu.	394	68	83
Cotton, per lb.	37 7	9 4	75
Beef cattle, per cwt.	897	598	33
Butter, per lb.	56	34	39
Hogs, per cwt.	1344	762	43
Wool, per lb.	50	16	68
Eggs, per doz.	37 5	19 4	48

* The data of this table are taken from the yearbook of the U S Department of Agriculture for 1923.

average, the commodities listed in the table suffered a cut in prices of over 50 per cent.

As noted in Sec. 21·6, falling prices would bring no disadvantage if all prices fell simultaneously and equally; but they do not. While farm prices were falling in the early 1920's, the heavy interest on farm mortgages and the heavy taxes on farm real estate were staying at their wartime level or even still increasing slightly. The prices of manufactured goods that farmers bought were falling, but falling more slowly than the prices of things that farmers sold. Figure 29·8 shows the situation graphically. During the war years, farm products had advanced more rapidly than had the prices of

[1] This drop in farm prices from 211 to 124 appears to be, like the drop in the prices of all commodities, a reduction of just over 40 per cent. But it should be pointed out that the figures that we used for all commodities were monthly figures, and we measured the drop from the highest single month in 1920; while the farm figures just given are yearly averages, and we have measured the drop from the 1920 average. This makes the reduction in agricultural prices appear smaller than it actually was.

those things which farmers buy, and farmers as a result had found their condition improved. When prices broke in 1920, the prices of farm products fell further and faster than the prices of things that farmers bought, so that farmers were in a disadvantageous position. Although the amount of this disadvantage varied from year to year, actually disappearing altogether temporarily in 1925, nevertheless the general situation of farmers was not again so good as it had been before the first World War until the coming of the second World War. For two long disheartening decades, farmers found themselves constantly at a disadvantage. At the low point, in 1932, farm products had a purchasing power only two-thirds that which they had

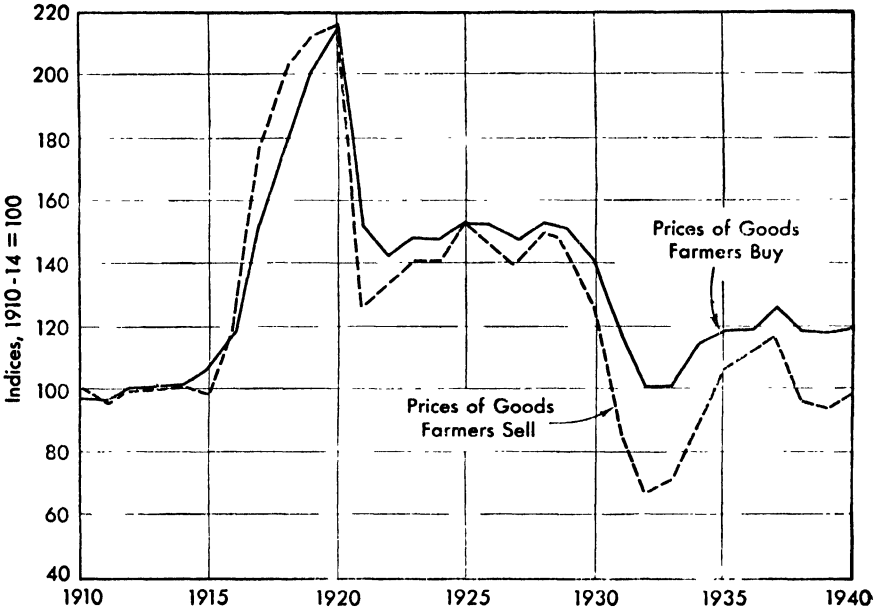


FIG. 29 • 8.—Index numbers of prices paid by farmers for things purchased by them, and of prices received by farmers for things which they sold, 1910-1940. Base period is 1910-1914.

enjoyed in 1910-1914. It took, in other words, half again as much in the way of farm products to buy a given quantity of those things farmers buy; farmers had to work half again as hard or half again as long to buy what they had purchased before the first World War. The fact is, of course, that they did not buy so much. Their level of living fell. The percentage of the total national income going to farmers diminished sharply, as we see in Fig. 29 • 2. It was at this time that the present farm problem arose—a problem that consisted of low incomes, low purchasing power, and falling levels of living for a considerable proportion of the economy. The so-called "prosperity" of the twenties was not shared by the farmer. The businessman who

felt the awful weight of depression in 1930 was beginning to share an experience that the farmer had already known for a decade.

29 • 9. Period of Maladjustment. When prices in a given industry drop below the general level and when profits disappear, it is in general an indication that some of the people in that industry should go elsewhere. This is the method by which we tell enterprisers in a "capitalist" economy that their services would be more valuable in other lines. Agriculture had expanded rapidly and radically in response to the high prices and profits of the war years. This expansion had been desirable. But now, as foreign demand returned to normal, it was also desirable that domestic agricultural production should return to normal, and the desirability of such a move was made clear (or should have been made clear) to farmers through those more or less automatic market forces which are reflected in price and profit changes. If farmers had taken the hint at that time and if there had been any considerable movement out of farming into other lines of work, it is quite likely that the "farm problem" would not have arisen to plague the entire society for two decades—even possible that other related maladjustments would not have arisen to plunge the entire economy into depression 10 years later.

To be sure, a minor move of this sort did occur. Many farmers found themselves unable, now that their products were selling at lower prices, to raise enough money to meet the interest on mortgages and the taxes on real estate, both of which remained at or even above their wartime peak. Debtors, as we long since discovered,[1] always suffer when prices fall, and a considerable proportion of our farmers were heavily in debt at the close of the war because they had purchased land and equipment at inflated wartime prices. After the 1920 drop in prices, mortgage foreclosures became common, and debtors were forced out of farming whether they liked it or not. This, to be sure, removed these particular farmers from farming, but it did not remove the farms. The creditors who had foreclosed either took over the farms themselves or sold them to new owners, and there was very little net reduction, as there should have been, in agricultural activity.

There were several reasons for the failure of farmers to withdraw from farming. In the first place, although farm costs may not be fixed rigidly in the production of particular products,[2] they are fixed in the field of agriculture as a whole. A farmer can often turn wheat land to corn, or tobacco land to melons; but he cannot turn the productive assets of a farm to the making of shoes or the manufacture of motorcycles. In 1920 and the years immediately following, there was no advantage in shifting from one farm crop to another, since all were depressed. The decision was one between

[1] See Sec. 21 • 6.

[2] See Sec. 29 • 4 for mention of this problem.

farming or going out of business, and in such a decision almost all the farmer's costs were fixed. There were no profitable alternative enterprises into which he could put his productive resources. As a result, the farmer, like any other rational businessman, faced with an unpleasant decision, accepted what seemed to him the lesser of the evils and decided for the time being to neglect his fixed costs and to operate at a loss.

We have already noted another reason for the failure of farmers to shift easily out of farming. This is their labor situation, which differs materially from that in industry. A very large part of the farm labor supply is composed of farmers themselves and of members of their families. Farmers not only could not transfer their land and their machinery to nonfarm uses, but they also could not lay off their labor force easily. To be sure, there was some movement of farmers and of members of their families to the city; but it is always a difficult thing to break off family ties, and workers always tend to be more or less immobile. The tendency was to hang on on the farm, hoping that things would soon be better; and, as large numbers of people took this course, they made it more and more certain that things would not be better.

Contraction is always painful in any business. Failure and bankruptcy are unpleasant things. Yet, like the lancing of a boil, they are often the forerunner of better things to come. Sensible and understanding men face the facts under such circumstances and get it over with. Yet, for the reasons that we have mentioned, contraction is even more painful in agriculture than in most industries. In addition, the farmer was faced with another problem which to a considerable extent excuses him for his continued insistence on refusing to face the facts and move out. This was the continued inexcusably bad advice which he got from the politicians. This bad advice started immediately with the beginning of the problem and continued unabated (and even greatly increased) through both Republican and Democratic administrations right up to the outbreak of the second World War, when the farm problem at least temporarily disappeared. Some political leaders perhaps failed to understand the problem and can be excused; others seem to have feared to take the political consequences which they imagined would follow the giving of unwelcome advice, and thought of the problem primarily in terms of getting votes rather than in terms of ultimate social welfare. Be the cause what it may, it is certain that hardly a single political leader, and none of the policy makers in Washington, followed the sensible course of encouraging farmers to move out at a time when there were still many profitable nonfarm ventures to which they could move, and of planning legislation that would have made such a movement easier and more certain. On the contrary, farmers were advised to "stick it out," and to rely on their government for aid. They were led to believe that, if they hung on just a

little longer, the government would adopt some program of "farm relief" which would settle all their difficulties. One such proposal after another was suggested, but the farm problem continued, as it was bound to do as long as a disproportionately large amount of the nation's man power and equipment remained in agriculture.

Society is under no obligation to support men in fields of activity where they are not needed. Under our system, men are encouraged to leave such fields by low prices, low profits, and low incomes; and year after year farm prices, farm profits, and farm family incomes stayed low. But with the coming of general depression in 1929 the bad situation became even worse.

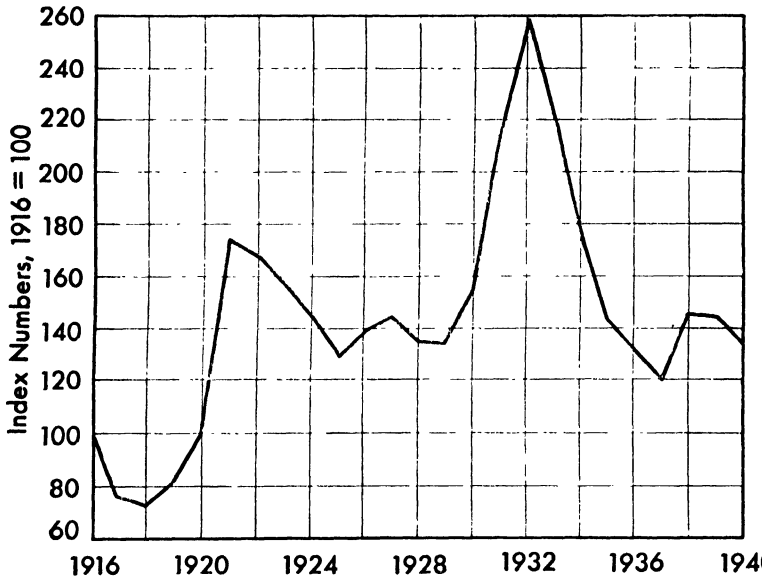


FIG. 29 • 9.—Burden of farm mortgage debt, United States, 1916-1940. This diagram shows not the actual amount of the mortgage debt but the amount of debt in terms of the prices of farm products. It indicates roughly the amount of farm products necessary to pay the existing mortgage debt at any time. Index numbers, 1916 = 100.

Reference to Fig. 29 • 8 will show that the prices of farm products again fell more rapidly and further than the prices of those commodities which farmers buy, and the farmers' buying power shrank to a new low level. This time the problem was not so simple as before. Maladjustment had continued to the point where the whole economy was affected,[1] and there were not the same profitable alternatives to which farmers could turn in other lines. Moreover, the long-continued burden of farm debt was becoming unbearable.

[1] This statement should not be taken as indicating that the failure to adjust agricultural production in the twenties was the sole cause of the depression of the thirties. In the opinion of the author, it was one of the major factors.

Total farm mortgage debt, which had been at a level of about 5 to 6 billion dollars before the first World War, rose to a peak of 10.8 billion in 1922; although it fell slightly in following years, it was down to but 9.6 billion when the crash came in 1929. Moreover the student will realize that a reduction in the total amount of farm mortgage debt does not mean necessarily a reduction in the burden of the debt. If the amount of the debt is cut in half but the prices of farm products fall by more than half, then the difficulty of paying principal and interest has increased rather than diminished. Figure 29·9 shows the burden of farm mortgage debt in the United States from 1916 to 1940; the amount of the debt as compared with the prices of farm products. When the line of Fig. 29·9 rises, it indicates that farmers were having greater difficulty in meeting their mortgages; when it falls, it means that the actual burden of mortgage payments was reduced. In this figure we see the great increase in the burden of farm mortgages immediately after the first World War, with a slow reduction in the subsequent years until the coming of the depression in 1929. For the next 2 or 3 years the burden of this debt rose to such heights that it could not be borne, and farm foreclosures took place on a wholesale scale. Tenancy became more common as owners lost their farms, and a farm situation which had been unhealthy for a decade became desperate.

29·10. Farm Legislation of the Thirties. Farm legislation of one sort or another has been common at least since the granger and populist movements of the late nineteenth century; but the long-continued serious plight of farm people after 1920 led to repeated suggestions that "there ought to be a law"—that the government should take a hand and do something about it. The demand was all the more natural in that political leaders had been promising all sorts of impossible help to the farmers. For the time being, the tendency was to give up individual efforts and depend on government interference. To be sure, the basic problem was still just what it had been before: an unbalanced economy with too many people in agriculture. But the government programs that were established in the thirties relied less on the obvious and direct method of helping people to shift to other lines of work than on keeping them in agriculture and making them less efficient—reducing by various artificial means the amount which they produced. It had long been known that small crops of many farm products yield more total income than do large crops. The government adopted the questionable procedure of trying to force people to pay more money for less goods by restricting output. Such a program is always disadvantageous to the buying public and can be excused on no other ground than that the gain to farmers would more than offset the loss to nonfarmers. Such an assumption is on a priori grounds exceedingly unlikely, and on a posteriori grounds there seems no reason to suppose that it was justified; but perhaps it is fair to say that the

political administrations of the thirties showed no poorer imaginaion than those of the twenties, and that they did have more courage even if less discretion.

The details of the farm legislation of the "recovery program" and the New Deal are too many for inclusion here. The proposals were intended, so their sponsors assure us, to give a fair share of the nation's income to farmers. It was for the time being no longer felt that a man was responsible for his own income, and that it was his job to see to it for himself that his share was fair. On the contrary, the philosophy grew up that it was the government's responsibility to ensure adequate income to all.

For the most part, the measures proposed recognized the fact that farmers are small producers in an unorganized and highly competitive market, while they deal with people, both in buying supplies and in selling produce, who are operating on a larger scale and often with at least some degree of monopoly advantage. Instead of attempting to reduce the monopolistic power of these other groups, efforts were made to give farmers an equally monopolistic position. In fact, not only in agriculture but in the NIRA program for manufacturers and in the labor legislation of the time, the administration turned to a policy of encouraging, fostering, aiding, and almost forcing monopoly. As we know, restriction of output is one of the fundamental tendencies of monopoly, and restriction of output was basic in one guise or another in almost every farm program of the thirties.

The first legislation attempted to handle the matter by buying "surplus" stocks and withdrawing them from the market. The Federal Farm Board began in 1930 to buy certain agricultural staples in rather large quantities. But no provision was made for reducing or limiting production and, what is more natural, if it is known that the government intends to support the price, than that farmers should grow more rather than less? The Agricultural Adjustment Act of 1933 attempted, through what turned out to be unconstitutional means, to remedy this situation. This act was the one afterward criticized for having "paid farmers not to produce." The heart of the act was contained in provisions that allowed the Secretary of Agriculture to pay rent to cooperating farmers who would hold land out of production. The act was "voluntary" in the sense that farmers were not forced to participate; but it can hardly be said in an economic sense that an act is voluntary when we make it profitable for a man to perform it and unprofitable for him not to do so. The act also provided for the dumping of surplus agricultural commodities on foreign markets—a practice common to nonagricultural monopolies and one that benefits foreign consumers at the expense of domestic ones. The cost of the program was to be borne by using the proceeds from "processing taxes," paid by the manufacturer or processor of the commodities. The act provided for payment of "benefits" to farmers on

their "allotments." The government set up for each farmer an allotment based on his earlier production, and then paid him a benefit or bonus on any amount of production up to the amount of his allotment. On any amount above it he received only the market price; and, of course, noncooperating farmers received the market price on their entire output. From the same source, farmers were paid a rent for land held out of its normal use.

The act of 1933 was declared unconstitutional in 1936. Studies seem to indicate that the program under it had been either entirely unsuccessful in reducing output of some of the commodities covered, or, if it had some small effect, it was negligible as compared with those year-to-year variations which occur as the result of natural causes.[1] With other commodities there seems to have been some gain in farm income as a result of the act, with tobacco and cotton showing the greatest evidence of such gains. Moreover the administration of the act was criticized, even by its friends, for having "taken on such a character, particularly through its definition of the basis of payment, as places minor emphasis on true agricultural adjustment and major emphasis on disbursing a very large amount of money to the maximum number of farmers."[2] If the economic effects were varied, there seems good reason to believe that the act made a lasting psychological impression on both farmers and politicians. The belief that "there ought to be a law" had been cemented, and the willingness to substitute reliance on the government for self-reliance had increased. The spirit of independence and "rugged individualism," which had marked our entire economic life and especially our agriculture, was materially and perhaps permanently weakened.

It is an old saying that "there is more than one way to skin a cat," and Congress long ago learned that, when it cannot do certain things constitutionally, it can sometimes contrive to do the same thing by pretending to do something else. Thus, for example, when Congress could not forbid the issuance of bank notes by state banks, it laid a tax on such issues which made it unprofitable to make them.[3] Similarly in the present case, when the Agricultural Adjustment Act was declared unconstitutional, the administration contrived to carry out substantially the same program under other guises, such as "soil conservation." To be sure, there was, and had been for some time, a problem of soil conservation, but the adoption of the Soil Conservation Act at this time, and the method of its administration, made it patent that it was a subterfuge for continuing activities that had been forbidden in the ruling on the Agricultural Adjustment Act. The Soil Conservation and Domestic Allotment Act, approved in February, 1936, con-

[1] See NOURSE, E. G., J. S. DAVIS, and J. D. BLACK, "Three Years of Agricultural Adjustment Administration," Brookings Institution, Washington, D. C., 1937.

[2] *Ibid.*

[3] See Sec. 20 · 3.

tained provisions not only for conserving the soil resources of the nation but also for reestablishing the prewar relationship between farm and nonfarm incomes. The Agricultural Marketing Act of 1937 reenacted certain provisions of the Agricultural Adjustment Act and reemphasized the idea of parity incomes. These and several other subsequent acts were directed toward the goal of putting farm incomes back at their prewar level as compared with incomes in other occupations. In this respect they differed from the 1933 program, which had sought parity of prices rather than parity of incomes. Amendments to these acts from time to time, and especially in 1938, were directed toward increasing the government's power of regulation of production and toward adding compulsory features to the act.

These new acts purported to set up a program of soil conservation. To this end, the Secretary of Agriculture was given funds which he might pay to farmers for doing whatever he thought would forward the "purposes of the Act." The Secretary thereupon set up machinery for rewarding farmers who would plant "soil-conserving" rather than "soil-depleting" crops. Perhaps it was not surprising that the Secretary listed as soil-conserving those crops (grasses, legumes, etc.) which did not to any considerable extent enter the market, while listing as "soil-depleting" those crops (cotton, corn, tobacco, potatoes, etc.) which were major sources of farm income. The new act actually attempted to make it profitable again for farmers to limit the production of certain key crops and to turn their land into other supposedly soil-conserving uses. The policy of the administration was again a policy of aiding the farmers through planned scarcity—through making it possible for them to adopt the same kinds of repressive and restrictive policies that had long been used to such social disadvantage by monopolists in other fields.

The recent farm legislation, especially as amended in 1938, does not rely on processing taxes but provides that benefit payments are to be made, if at all, from funds appropriated annually by Congress. Thus Congress retains some control of the administration of the program. Voluntary quotas are established for the acreage of various crops, but provision was made for compulsory marketing allotments when ratified by a two-thirds vote of the farmers concerned. This provision, often pointed out by the advocates of the legislation as being "democratic," is, in fact, just as democratic as would be a provision that allowed two-thirds of the manufacturers of any other article to enter into combination in restraint of trade. Surely the consumer, who was directly affected by the outcome, had no voice in this "democratic process." Provision was also made for the government to give crop loans on basic farm crops when their prices fell below stated percentages of the "parity price," taking the crops themselves as security for the loan. Since the loans made on this basis were often in excess of the actual market price of the commodities, the loan turned out to be a sale in excess of the market price.

29·11. Appraisal of Farm Legislation. The various legislative acts intended to solve the farm problem, which followed each other with rapidity in the period between 1930 and the outbreak of the second World War, had as their major objective the restoration of farm incomes to their prewar parity as compared with nonfarm incomes. To some extent at any rate, they seem to have succeeded in this objective, although the crop reductions that came as the result of a long period of serious droughts seem to have had even more effect, and it is difficult to separate the two causes. Certain it is that

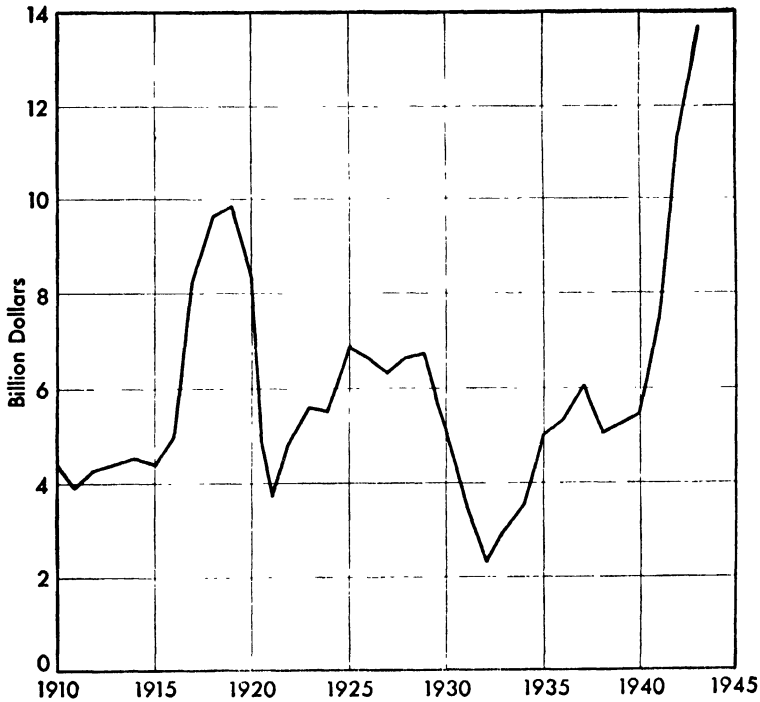


FIG. 29·10.—Net income to persons on farms from farming, including government payments, 1910-1943. (*Statistical Abstract of the United States*, 1944-1945, *p.* 647.)

farmers' incomes, which had been very low indeed in 1932 and 1933, rose considerably thereafter. Figure 29·10 shows the changes in the net income that farmers received from their operations between 1910 and 1943. But this legislation was subject to at least two serious criticisms. In the first place, it attempted, not to reduce the number of farmers, but to establish in the field of agriculture the same types of limitation of output that have been used by monopolies—to adopt a program of scarcity and, in the worst cases, even to destroy products already wholly or partly produced. In the second place, where it did take land out of use, it made no attempt to take the worst

land out of use and leave the best in. Each farmer was given an allotment lower than his former production. Thus, every wheat farmer in a given area might be given a quota equal to 85 per cent of his average production of wheat in the preceding 2 or 3 years. Under these circumstances, each of them would be expected to withdraw 15 per cent of his acreage from wheat production. From a social standpoint, if it were desirable to reduce wheat production 15 per cent, it would be far better to eliminate entirely that 15 per cent of the production which was produced at highest cost, either because it was produced on the poorest land or because it was produced by inefficient farmers.

Economists have been prone also to question the entire idea of "parity prices" and "parity income" on which recent farm legislation has been based. No doubt "parity" furnishes a good political slogan and a rallying cry for farm groups; but its economic basis is questionable. The central idea of the parity programs has been that farmers should be as well off relative to other people as they were during the "base period" of 1910-1914. (Tobacco was given a base of 1919-1929, as being more favorable to tobacco growers.) The validity of this idea can be challenged on several grounds. First, how does one know that the 1910-1914 relationship was the right one? If we took a 1932 base, we would find that farmers have been relatively very well off almost all the time in the past 30 years.[1] Any year can be selected as a base period, and there is at least some ground for suspicion that advocates of farm relief have insisted on using the 1910-1914 base period solely to make a better political story for themselves long after all other agencies of the government have discarded it.

Even if we leave the difficult and technical (and, be it admitted, largely subjective) statistical problem of selecting a base period out of the picture, we have the question whether the government should attempt by artificial means to maintain the price of any commodity or group of commodities at any fixed level relative to other prices, or whether the government should attempt by artificial means to maintain the income of any group of people at a fixed level relative to the incomes of other groups. Economic conditions are not static. There are constant changes in consumer tastes and producer costs. These changes should be reflected in the market place, through changes in prices and incomes, if we are to get those production changes which consumers desire. If the farm relief program is a charity program, intending to subsidize poor people, it should be frankly and openly described as such. If it is intended as a program for agricultural adjustment and

[1] This the student can discover easily for himself by reference to Fig. 29 • 8. Taking 1932 as a base in that figure would have the effect of shifting the broken line vertically up or the solid line vertically down until the two lines coincided in 1932. Under these circumstances, it is clear that throughout the entire diagram the broken line would lie above the solid line, and it would appear that the prices of things that farmers sell were higher than the prices of things that farmers buy relative to the situation in this new base period.

readjustment, then it is fair to suggest that the prices of farm products and the incomes of farm people should be allowed to fluctuate according to the changes in the underlying situation. If it is a charity program for the poor, it should be applied only to poor farmers, and poor people of other occupational groups should share in it. Otherwise it is certain to impede exactly those readjustments which it pretends to promote.

In all fairness, it should be said that almost every occupational group sets up organizations to bring pressure on legislative bodies, urging them to adopt programs for the selfish advantage of the petitioners. The farmers are not unusual in this respect. Although such groups usually put on a show of basing their program on economic principles and of seeking general social advantage, the student should early learn to be skeptical of all such efforts. There is, as a matter of fact, very little actual economic basis for most of the arguments that farmers and their friends have used for the farm program; but neither is there much actual economic basis for the arguments used by organized labor groups in support of labor legislation, or of organized manufacturers in support of the tariff. The motives of these pressure groups are seldom what they say they are, and the "social benefit" which they invariably shout about is usually mere window dressing. Their slogans—"parity income," "collective bargaining," "full dinner pail," "protect against coolie labor," etc.—seldom mean much when analyzed carefully, and what little they do mean is seldom defensible on social grounds.

Questions

1. Section 29·1 speaks of increased efficiency in agricultural production, and of the fact that this increased efficiency has released many people to do other work. This is obviously true; but it is also true that some work that was once performed on farms has been shifted and is now carried on by nonfarm groups. In these cases, the farmer is not doing the work more efficiently; he is not doing it at all. We might illustrate by the shift of butter making from the farm to the creamery. What are some other illustrations?

2. Name several specific changes in agriculture that illustrate the industrial revolution in that field.

3. Why are farms, on the average, small-scale industries? In what other ways could we measure the size of an industry in addition to measuring it by the number of workers?

4. What is the relationship between the size of the population and the demand for farm products? Do you think that the relationship is closer for farm products than for manufactured products?

5. It is sometimes said that advertising does not pay in agriculture as it does in industry, and the statement is based on the argument that *the demand for agricultural products is inelastic, and therefore advertising could bring at best only small in-*

creases in quantities sold. This argument involves a *non sequitur* which should be obvious to the careful student who remembers what we mean by elasticity of demand. Discuss.

6. "The farm products that are put on the market today are much more nearly like those of three or four generations ago than would be true of most manufactured goods." What, on the other hand, are some of the ways in which today's farm products differ from those of great-grandfather's day?

7. What are the peculiarities of the supply of agricultural products?

8. In Sec. 29 • 4 appears a long statement in italics describing the effect of changes in demand or supply when both demand and supply are inelastic. Make the correct parallel statement as it should be if it were to start, "*If either the demand or the supply is inelastic, . . .* "

9. What, as you understand it, are the facts about fixed costs in agriculture as compared with fixed costs in other parts of the economy?

10. Why is it easier for a manufacturer to lay off workers than it is for the farmer?

11. What are supposed to be the advantages of farm ownership as opposed to farm tenancy:
    a. From the standpoint of the farmer?
    b. From the standpoint of society?

12. Why has tenancy tended to increase in the United States?

13. What were some of the measures used by the government during our early history to stimulate the settlement of the public domain?

14. Why did farm land values rise during the period of the first World War? In answering this question, recall what you know about the capitalization of net income (see Sec. 26 • 9).

15. Why were farmers hurt more than many other groups of people by the falling prices of 1920 and after?

16. Give some of the reasons for the failure of farmers to move out of agriculture when the farm situation became bad in the early 1920's.

17. In what ways were the activities under the various agricultural acts similar to activities carried on by monopolists in nonagricultural fields?

18. Parity prices and parity incomes are defined in great detail in the statutes. Without looking up these technicalities, what do you understand as the basic idea of parity?

19. What are the major objections that have been raised against the concept of parity?

PART SIX. Distribution

30. THE NATURE OF THE DISTRIBUTION PROBLEM

30 • 1. Distribution Defined. When the economist uses the term "distribution" in its technical sense, he refers to what the layman is more apt to call "the distribution of wealth." The problem is not that of the physical distribution of goods through the marketing system, carrying them from the point of original production to the consumer. Such activities add place utility to goods and, since they add utility, we consider them to be a part of production[1] rather than a part of distribution. The problem is, rather, one of deciding how the total product of society will be divided among all the people of the society. Who has claims to the product, and how large are the claims?

Most of the people of society are engaged in production. A considerable part of their waking time is spent in producing various goods and services that people want. During the course of a year, a tremendous volume of goods and services is produced, and our problem is to find out how those goods and services are then apportioned. Who has rights to them, and how does he get them; especially, what are the relative sizes of the rights of different persons and groups of persons?

The great stock of goods and services to be distributed we call the social dividend" or the "national income" or the "distribuendum." If the people of a nation are industrious and able, if they have good natural resources and equipment to work with, if the weather is propitious, and if there are no interruptions of war or other calamities, then the national income will be large. There will be a great volume of good things available for distribution, and it will be possible for everyone to have a large portion. If, on the other hand, the people of a society are shiftless or incompetent, if their natural endowment and equipment are poor, if the weather is unfavorable, or if social efforts are thrown into channels of destruction rather than construction, then the national income will be small. Only a small volume

[1] See Secs. 10 • 1 and 10 • 2.

of good things will be available for distribution, and it will be impossible for everyone to have a large portion.

To this extent, then, distribution depends on productivity. Men cannot share what they do not have, and in the main they cannot have desirable things unless they produce them. But the problem is not quite so simple. Even if we have a large pie to divide amongst us, it may turn out that most of us get small pieces, while a few get more than they can possibly digest. On the other hand, if there is but a small pie for many of us, still one or two of us may manage to get very sizable slices at the expense of others who get far less than their proportionate share. Obviously, the same thing is true in the distribution of the economic goods of society. We must first of all be interested in the size of the pie itself—in the size of the national income. Forces that tend to increase the size of the national income make it possible for the individual shares to be larger, and make it certain that the average share will be larger. But then we must interest ourselves, if we are to understand the problem at all, in the way the pie is divided—the equality or lack of equality in the sizes of the pieces. The study of distribution is a study of this latter problem, just as the study of production is a study of the former. The two are intimately woven together in the economic life of any nation.

30·2. Personal and Functional Distribution. So far we have spoken of the problems of distribution as though we were to ascertain the facts for single individuals: What is the size of each man's income, and what are the forces that determine it? In a nation of 140 million people it would take too long to examine each case, although it would be instructive if the student would consider the cases of say a dozen people in his home town who are well known to him—people at different levels of income—and try to decide as well as he can why their incomes differ. He would discover immediately that many of the causes of differences in income are well known to him. They would include, among many other things, innate ability, persistence, ambition, honesty, thrift, industry; for some people, unfortunately, they would also include guile, callousness, deceit, or just plain luck.

We shall not try to consider here the forces that determine the sizes of individual incomes. Such a study we call the study of "personal distribution." We have, however, already discussed[1] the results in a general way, noting the numbers of personal incomes of various sizes. We have discovered that the social dividend is not divided equally but that, on the contrary, some of the portions are a thousand times greater than others. In this chapter and in these immediately following we shall turn our attention to the division of the social dividend among the major groups of productive factors: land, labor, capital, and management. We shall thus consider the distribution of

[1] See Chap. 4.

the product among the main groups of those who produced it. We shall be classifying people, as we did when we studied production, according to the function that they perform in the economic system, and we shall see how much of the product goes to each of these functional groups, and why. This approach is called that of "functional distribution" as contrasted with the study of individual incomes, or personal distribution.

30·3. Distributive Shares as Costs. The problems to which we now turn in the field of distribution are not, however, entirely new. We have spent a good deal of time, especially in Chap. 17 but also elsewhere, in studying costs of production and their economic significance. We have talked of fixed costs and variable costs, of joint costs and overhead costs, etc. We have tried to discover something about the forces that influence any businessman's costs of production. Now we should notice that, in general, one man's costs are other men's incomes. The farmer spends money for fertilizer, seed, labor, machinery, gasoline, etc. In the farmer's eyes these are costs. But they appear as incomes to the fertilizer company, the seed house, the hired hand, the machinery store, and the filling-station proprietor. Two men look at the same expenditure. One calls it a cost; the other calls it an income. This is not unusual, of course. One's point of view often determines the way things appear to one. The housewife goes to the store and buys a dozen eggs. Was the transaction a purchase of eggs or a sale of eggs? Obviously, it was both. The person who saw the eggs going out called it a sale; the person who saw them coming in called it a purchase. Similarly when one man borrows, another lends. Borrowing and lending are not two transactions. They are the same transaction viewed from two different points of view. Similarly in our present case, we can say that in our type of society costs and incomes are not two different things, but one thing viewed in two different ways. The man who sees the money or goods going out talks in terms of cost; the person who sees the money or the goods coming in talks in terms of income. When, in our earlier study, we discussed the forces that affect costs, we were at the same time discussing the forces that affect incomes, and our problem is now the relatively simple one of viewing these same forces again but from a new point of view.

30·4. Demand for Productive Factors. As has just been said, we are about to consider the ways in which the productive factors, land, labor, capital, and management, share in the total product. We are about to discover, as well as we can, how much of the social dividend is allotted to each of these functional groups. There are a number of ways to approach this problem, but perhaps it will be simplest from the standpoint of the student if we begin with a method already familiar to us. We will see as we go along that one can explain a considerable part of the phenomena of distribution in terms of the supply and demand analysis studied heretofore. Under certain

circumstances we have seen that the prices of consumption goods are determined by supply and demand.[1] Now we shall see that the amounts paid for various factors of production (the costs of these factors, or the incomes to these factors) also depend to a considerable extent on the supply of them and on the demand for them.

But the demand for factors of production is different from the demand for consumption goods. Men demand consumption goods for their utility, and we learned in Chap. 6 that, if the utilities of the goods are given, the demands for them are given. But utility is the ability to satisfy human wants—an ability that every good must have before we call it a good. Therefore, while we are dealing with consumption goods our problem is easy. When we turn to the factors of production, we cannot talk about utility in the same direct sense as before. A steam shovel does not satisfy wants directly; it has no direct utility. Men want it, not for itself, but for its product. That is why we call it a "factor of production"—because it is used, not for direct satisfaction of wants, but as an aid in producing other things which will satisfy wants.

Productive factors are valued, not for themselves, but for their products. Consequently whatever utility they can be said to have is not their own ability to satisfy wants, but an ability to produce other things which will. The "utility" of the cookstove is not a direct one. It is, rather, a reflection of the cherry pies and strawberry shortcakes and sirloin steaks smothered with mushrooms and brown gravy that can be had with the aid of the cookstove and not without it. The demand for factors of production is a *derived demand,* depending on the demand for the product. If men no longer cared for cherry pies and strawberry shortcakes and for the other things made with cookstoves, then the stoves would have no value whatever. In the same way, if the demand were to disappear for the products of any other factor of production, the value of the factor itself would disappear. In this way we see that the value of factors of production is, even if indirectly, tied back to human wants and that the demand for factors of production reflects a demand for the products that men hope to obtain from them.

30·5. Distribution and Marginal Productivity. Since factors of production are desired only for their products, it would be natural for us to expect that they would be paid at least roughly in proportion to their productivity. If a factor was expected to be very productive indeed, businessmen would be willing to pay a large amount for it, and the income of the factor would be large. If, on the other hand, it was anticipated that a factor would produce very little, no one would be willing to pay very much for the use of it, and its income would be low. To be sure, our earlier study should have led us to suspect that the payment would depend, not on the total productivity, but on

[1] See Chap. 15.

the marginal productivity. We know, for example, that a consumption good may have very great total utility, yet, like the air, its economic value may be nil. We learned in Chap. 5 that marginal utility is more important than total utility in determining the value of consumption goods. Now when we turn to factors of production, which are wanted for their product rather than for their utility, we would naturally expect that the value of the factor would depend to a considerable extent on its marginal productivity. In the chapters immediately following we shall discover that this is, in the main, the case. There are, to be sure, many factors to be considered, but *in our type of society it seems to be true that the most important single factor influencing the amounts of income received by the various factors of production is the marginal productivity of those factors.* If an entrepreneur believes that he can add significantly to his product by adding a unit of any factor of production, he will be willing to pay something to get the added unit; and he will be willing to pay, if he has to, anything up to the amount that this unit of the factor will add to the value of the product. This principle will be illustrated in more specific terms in the following chapters.

30 · 6. Distribution and Ethics. Every system of distribution must be judged on the basis of how satisfactorily it divides the product of society. Do the various people in the society get what they deserve—what they ought to get—what is fair and just and reasonable? Obviously these questions are not questions merely as to how the product is divided, but they include decisions as to how it ought to be divided. These are decisions in the field of ethics or morals.[1] They cannot be answered on any scientific basis, since they are outside the field of science. This does not mean, of course, that the questions are unimportant. They are among the most important questions that face any society. The fact that a particular problem is or is not in the field of science has nothing whatever to do with its importance in human affairs.

The people of different societies differ sharply in their opinions as to what is a fair and reasonable system of distribution. Even within any given society there are decided differences of opinion, not on how the distribuendum *is* divided, but on how it *ought to be*. For example, some men would say that it is fair to divide the product among people in proportion to the respective contributions that they have made to it. Such a scheme would give large incomes to those people who produced much and small incomes to those who were unproductive. Other men have held that the fair thing is to divide the product among people in proportion to their respective needs, giving large incomes to those who have large needs and small incomes to those who have small needs. Either of these suggestions, taken alone, is one that we all tend instinctively to accept as being fair, yet a moment's reflec-

[1] See Chap. 2.

tion will show that they are radically different systems. The shares that each person would receive would be very different under the two systems.

Moreover, we must not forget that if we were to change the system of distributing the product, we might find that our decision had changed the amount of the product to be distributed. In our society one's income is his major incentive to work. If there is no relationship between the size of his income and the amount of his work, he has no reason to work. If he can get just as much income by loafing as he can by working, he is apt to loaf. Then the amount of the product is diminished. Society stimulates production either by rewards or by punishments—by promises or by threats. A slave society operates on the latter, but in the main our society operates on the former. As a consequence, we are forced to remember that the decision as to what would be the best way to divide a given product cannot be divorced from the question of how much product there will be to divide. One cannot decide in the abstract, what wages, for example, are fair, without considering also what will be the ultimate effects of the decision.

30·7. Payments to Property. Along with the other ethical questions raised in connection with problems of distribution, we have to notice the fact that some payments are made to men in return for their own productivity, while other payments are made to men in return for the productivity of their property. This we can see immediately if we hark back to things that we discussed when first we took up the study of production. At that time[1] we saw that the factors of production can conveniently be divided into four major categories: land, labor, capital, and management—the resources furnished by nature, the human resources, the machinery and equipment, and the combining of those other factors into an efficient working unit. Labor and management are human factors and, as we have said, they might well be lumped together as a single factor of production, as some authors in fact do handle them. But the other two—land and capital—are inanimate, and no one pays them anything. No one ever makes a payment to land; he makes his payment to the person who owns the land. No one makes a payment to a motor truck; he makes his payment to the owner of the truck. Men in these cases are paid, not for their own productivity, but for the productivity of things that they own. The situation may be somewhat clearer if we tabulate the various factors of production and the human agencies involved with each; and, while we are at it, we shall tabulate also the name that we give to the payment made for the use of each of the factors. Table 30·1 gives the relevant material. Since we have defined in earlier chapters the words in the first column of the table (land, labor, etc.), we can use the other columns as definitions of the words that appear in them. Thus, we define a landlord as one who owns land, and rent as any payment made for the use

[1] See Chap. 10.

of land. The layman may say that he "rented" an automobile. But an automobile is not a natural resource; it is capital. Any payment made for the use of capital is defined as a payment of interest, and in the technical sense we would call the payment interest, not rent.

Now as we look at the table we see that laborers and managers are paid wages or profits in return for their own efforts and their own productivity; but landlords and capitalists are paid for the productivity of their property—their land or their capital—rather than for their own productivity. Those who consider questions of justice and morals have often raised the question as to how we can justify paying men who have not themselves been productive. They may admit, at least for purposes of argument, that if a man has worked hard and produced something by his own labor, he has some moral claim to the enjoyment and disposition of what he has produced—in other words, that he has a just and reasonable claim to ownership.[1] But, how does

Table 30·1. Factors of Production and Their Shares in Distribution

Factor of production	Human agent involved	Share in distribution
Land	Landlord	Rent
Labor	Laborer	Wages
Capital	Capitalist	Interest
Management	Entrepreneur	Profits

one get any right to goods that were produced by natural resources or by capital?

The second of these questions is in some ways the simpler, and we shall comment on it first. The natural answer to make to the question, as far as capital is concerned, is that men did make the capital, and the capital is productive. If we do not pay men to make the capital, or for the use of the capital that they have made, they will not make it. Payments for the use of capital are not robbing other factors, taking away what they might otherwise have had. On the contrary, the use of the capital has increased the output, and it would be quite possible to pay laborers *more* in a society that used capital than could be paid without it, and still have plenty left over to pay the capitalist—to encourage him to sacrifice his present wants and to save for the production of capital.[2] We pay for capital because it pays us to do it and

[1] The student will note that no assertion is here made that a worker *does* have any inherent right to what he has produced. The matter cannot be proved one way or the other. But most people have thought that production is the most valid claim to ownership that there is; some people, of course, do not believe in ownership anyway, on any ground.

[2] For a more detailed discussion of this matter, see Chap. 13.

because if we did not pay for it we would not get the capital, and the capital is productive.

To be sure, it has long since been pointed out that we would probably get some capital even if we did not pay people for it, just as we would get some water if we did not dig wells and some wheat if we relied on that which grew wild in nature; but, in practice, most men would not put in the time, effort, and money necessary to produce factories and machinery unless they could look ahead with reasonable assurance to getting some return on their investment.

With land the problem is a harder one, and the answer must be given with a good deal less assurance. What right do men have to be paid for the productivity of natural resources, which they did not produce? Let us divide the answer into two parts. First, it ought to be pointed out that it may well have been a mistake for our ancestors to build up a system of private property in land. Perhaps it would have been fairer and better all the way around if natural resources had been held in common by everyone. Even if that were true, it would not at all make it certain that it would be fair or wise to go back to a system of common ownership now. As we have pointed out at several places in this text, the social scientist cannot experiment in the same way that the laboratory scientist can. He cannot try the thing out one way, see how it works, and, finding that it works poorly, go back and try again. He has to make up his mind what to do, and then always go ahead from where he is. There is no possibility of going back in society in the sense that we can again put things back as they would have been if we had not once started out.

Let us consider the case of some man who owns land today—perhaps a farmer with a quarter section of fertile wheat land in Jackson Township, McPherson County, Kans. How does this farmer get any right to the product of his land? The answer is, at least in part, that, although he did not make the land, he did buy it. He gave up in return for it valuable things that he had produced. The land represents in a very real sense his labor and sacrifice and productivity. Perhaps, when the first white settler went into that territory and wanted to get possession of the northwest quarter of section 2 of that township, or of any other piece of land, it would have been better had society refused to allow private ownership. Such speculations as to ancient history would weigh but lightly with the present owner if it were now to be suggested that he give up the land. It does represent his own sacrifice. We have the problem of vested interests, about which we have spoken before. When society adopts a system, be it good or bad, people adjust themselves to it and, if the system is later changed, these people suffer a loss from the change.

Our first answer, then, to the argument that private ownership of land is undesirable, could be that, even if the original decision was unwise, there would now be grave injustices if we were to change it. There is also a second part to the answer. When men went to McPherson County, Kans., to take up wheat land, society was by no means reluctant to have the land taken up. It was not thought that the process was antisocial. On the contrary, society encouraged it in the belief that there were social advantages which more than offset the disadvantages. It was thought that men are more productive when they are working for themselves—that the land would be cleared faster and cared for better if men could see a personal advantage from it. One of the difficulties with tenant farming today, as mentioned in the preceding chapter, is that tenant farmers tend to use up the fertility of the soil without replacing it; that they tend to follow cultural practices which yield the most product with the least expense immediately, but which wear out the soil and leave society poorer in the long run. Section 3·4 recounts some of the difficulties experienced in the early Plymouth Colony when the settlers tried for a time to get along without private ownership of land. These settlers, after some experience with it, voluntarily, purposely, and knowingly decided to adopt a system of private property in land as being socially advantageous. Therefore, when the question is raised as to how men get rights to receive income from land that was originally a gift of nature, we may reply with either or both of two answers:

1. Even if it was a mistake to have allowed it at first, there would be serious injustice in upsetting the system now.

2. The system may well be socially advantageous anyway and, if we did not have it, we might voluntarily decide to institute it.

30·8. Productivity as a Claim to Ownership. Before we leave this problem entirely, we should give brief attention to one other aspect of it which is commonly misunderstood. There are those apologists for our present order who assert that our capitalist system tends to distribute the product at least roughly in proportion to productivity and that consequently men tend to get what they deserve—that the system is inherently right or just. The following chapters make no such claim. Although it will appear that income is distributed in our society at least roughly in proportion to marginal productivity, no suggestion will be made that this is what ought to be. The beginning student needs some warning on this matter.

In the first place, a man's productivity is often determined by things beyond his control. Is there reason why the man who is lucky and gets a large product should be rewarded for his good fortune? Is there reason for punishing the man who is unlucky and who, as a result, gets but a small product? Insofar as differences in product are the results of differences in things that men can control—their own industry, foresight, and prudence—

then we might possibly accept at least in part the suggestion that they should be rewarded for productivity. But, insofar as differences in men's productivity are the results of forces that they did not themselves control, just to that extent the ethical claim vanishes.

Moreover, a man may himself be very productive because other men have set him free to be so. They have done the routine jobs so that he can have time for the creative ones. To this extent the very productive man always depends on his fellows and shares with them the responsibility for his productivity. Likewise, his productivity may well be dependent on his tools or the land on which he works. Thus, his productivity may depend on land or capital as well as on himself. Productivity in modern society is a joint affair, toward which many people contribute and, although we may set up some bookkeeping mechanism by which we pretend to account for the separate productivity of men, it is actually as impossible to assign accurately productivity to the various men and other agents involved as it is to separate the effect of costs that are jointly incurred.[1] It is literally true in our society that no man knows what he produces. If the student thinks that the preceding statement is an overstatement and brings up the case of the "self-sufficient farmer" who raises his own vegetables and therefore knows at least that he produced them, it is fair to ask how many carrots did the farmer produce, and how many of them were produced by the man who made the plow and the rake, and how many by the company that mixed the fertilizer, and how many by the men who worked on the railroad that brought these things to the farmer. In a very real sense we are productive as a society rather than as individuals. Our total product is far greater than our separate products could be. This, after all, is the advantage of division of labor; and it means that no man can hold a valid claim to having produced anything unaided and by himself.

Even if we were to get around these difficulties by agreeing on some sort of accounting (presumably for our purposes an accounting that imputed productivity on a marginal basis, since that seems to be the way our society works) showing how much various individuals produced, there would still be no valid basis for arguing that we ought to distribute the product in proportion to this productivity. For we must remember that, when the economist says that our system tends to distribute the product in proportion to what men produce, he is referring to men's productivity *to their employers* rather than their productivity *to society*. For example, it may pay the unscrupulous grocer to put sand in the sugar and water in the milk; and in this case he will have to pay the men who dig the sand and the men who dilute the milk. If they work hard and are unusually "productive," he will pay them more; if they are lazy and adulterate only a few gallons of milk a day,

[1] See Sec. 17 • 7.

he will pay them less or fire them altogether because of their inefficiency. He pays them well if they are productive to him—productive in the sense of doing what he wants done. Often, fortunately, what the employer wants done is something that is socially advantageous. Most employers are hiring men to do things that the general public desires. But there is no certainty whatever that we have set up a system that can be justified from the social standpoint when we set up a system under which private individuals pay other people to do those things that are to the employer's personal advantage. The ethical claim may still bear some weight, and will still bear some weight, insofar as one believes that the economic activities of men are directed toward socially desirable ends. To the extent that economic activities are devoted to socially undesirable ends, it is disadvantageous to pay men in proportion to their "productivity" to their employers.

30 • 9. Supply of Productive Factors. In Sec. 30 • 4 it was suggested that distribution may be explained in terms of the supply of and the demand for the factors of production. Some attention was given to the nature of the demand for those factors. Let us now turn to a brief consideration of the characteristics of the supply of the factors of production.

If we consider the society as a whole, the quantity of land available is practically fixed. It does not vary appreciably with the price offered for it. In the long run, the quantity available may change as a result of new discoveries, the wearing out of old land, or the discovery of new methods of using land. Even in the short run, of course, the amount of land available to a particular entrepreneur can be varied almost at will. In addition, the economic usefulness or productivity of the land depends to a considerable extent on the amount and the quality of the other factors that are used with it.

The quantity of labor is somewhat more easily varied, depending, as we have seen on the birth rate, the death rate, and migration. Likewise, there is considerable variation possible in the number of persons in a society who will work as compared with the number unemployed. The number of hours of work a day and the number of days of work a year can be altered materially. Without question, for example, it would be possible to get twice the number of hours of work a year out of the present population of the United States that we now get. There is, thus, a very considerable reserve supply of labor in any society at any time, and the quantity actually offered on the market does depend in great degree on the rate of remuneration offered.

Perhaps capital is more readily responsive to changes in rates of remuneration than any of the other factors. To be sure, the total quantity of capital available is approximately fixed at any moment; but it can be increased or decreased rather quickly if circumstances seem to warrant. Moreover, capital seems to be more mobile than the other factors—it moves more readily

from place to place. If there is a deficit of capital in a region, it can be readily brought in from other areas if the rate of remuneration makes it worth while. Land is difficult or impossible to move. Labor moves only with difficulty. But the quantity of capital available in any area depends largely on what is offered for it.

To be sure, in any practical economic situation we are more interested in the characteristics of supply as they present themselves to individual businessmen than as they appear socially. Although the total amount of land available to society is approximately fixed, the amount available to any individual producer is for all practical purposes limitless. He can get all that he can possibly use. In his eyes and from his own standpoint all the factors of production can be varied in quantity without limit. How many men he can hire depends primarily on the opportunities that men can find for employment elsewhere. How much land or capital he can control depends on alternative opportunities for the employment of land and capital in other lines of production. Our producer will have to offer at least enough to attract the factors of production from the next most favorable employment. As alternative opportunities become more and more attractive, it becomes harder and harder for the producer to get the productive factors he needs. His costs of employing them rise. The economist calls such costs "opportunity costs" or "alternative costs," and such costs, measured by the amount that the productive factor could obtain in other employments, are among the most important factors in determining the incomes that productive factors will receive.

This concept is so simple that it hardly needs illustration. If a given workman finds that he can command wages of $20 a week if he sweeps streets, $25 a week if he digs ditches, or $28 a week if he mows lawns, and if in all other respects the different kinds of work seem equally attractive to him, a new employer who wishes to attract this workman as a hod carrier will have to pay at least $28 a week to get him. The new employer (assuming that the jobs are equally attractive in all other respects) will find that he has to pay at least as much as the workman can get in his next best opportunity. We would say that there was an opportunity cost of $28, which has to be covered to attract this workman. In computing the amount of income that any productive factor can command in any line of work, we must keep in mind that there is always a minimum if the factor has alternative opportunities.

30 · 10. Distribution in a Noncompetitive Society. As we go through the following chapters, we shall note the tendency for factors of production to be remunerated roughly in proportion to their marginal productivity *insofar as they are employed in competitive industry.* We shall there illustrate conditions as they would be in such competitive industries and shall assume to

a considerable extent that competition does exist. Now we have already learned[1] that competition is by no means universal in our economic system. Although it is without question a stronger force than monopoly, the latter is strong enough so that we can by no means neglect it. We should, therefore, inquire here as to the effect of monopoly on distribution, just as we have inquired earlier as to its effect on production and on exchange.

In the first place, it should be pointed out that, even if there were a very considerable degree of monopoly in production, these monopolists might compete against each other for the factors of production. It would not be unexpected if they competed, for example, for labor and raw materials. A monopolist usually has a monopoly in selling, not in buying.[2] Now that we are considering the returns to factors of production, it is reasonable for us to point out that monopoly plays a much smaller part in the market for the factors of production than it does in the market for finished products.

Yet, we cannot leave the matter there. Even if a group of monopolists competed among themselves for the factors of production, the payments to these factors would not be the same as those made by a number of competitors competing for the same factors. The question as to whether the producers sell in a competitive or in a monopolistic market is an important one, even if the effect is indirect and presumably smaller than it would be if there were no competition in the purchases of the factors of production themselves. The reason is that monopolists choose, as we have seen, a different (and smaller) level of output than would be most profitable for competitors, and hence their total demand for the factors of production is different (and smaller).

We can illustrate with a simple numerical example such as those that we have used before. Let us suppose that a producer is varying the number of men which he employs and that as he adds men the amount of his product increases. Let us imagine that, when he hires 10 men, he gets 1000 units of product; when he hires 11 men, he gets 1050 units of product. The marginal product is 50 units of output. How much can the producer afford to pay for this eleventh man who is to increase the output by 50 units? That depends on the price of the product and also on whether the industry is monopolistic or not. Let us suppose first that the business is competitive and that the product sells for $10 per unit. Under these circumstances, the value of the marginal product is 50 units at $10 each or $500. If the employer can get the laborer for $300 or $400 or $450, it will be advantageous for him to do it. In fact it will be worth his while to pay the worker anything up to $500 rather than not to get him at all. He will, if he has to, pay anything up to the full value of the marginal product.

[1] See Chaps. 16 and 25.

[2] In fact, we noted in Sec. 16 • 3 that unified control in buying is technically called "monopsony" rather than "monopoly."

If we leave everything else in the problem unchanged and call the man a monopolist, what difference will it make? The answer is that, if the man is a monopolist, the expansion of his output will depress the market price. This is the fundamental characteristic that distinguishes monopoly from competition. Therefore let us alter our example just enough to make it fit our definition of monopoly. Let us assume that, when we hire the eleventh man and increase our output from 1000 to 1050 units, the price of the product drops from $10 to $9.99 apiece. Now what can the monopolist afford to pay the worker? We might say, as before, that the marginal output is 50 units and that these 50 units are worth at least $9.99 apiece, or $499.50. But the monopolist cannot afford to pay so much as this. On the contrary, if he sells 1000 units at $10 each, his total income will be $10,000. If he hires another man and produces 1050 units that he sells at $9.99 apiece, his total income will be $10,489.50. Hiring the additional man will increase his income by $489.50, and surely the monopolist cannot afford to pay him more than that. The competitor may offer the eleventh man up to $500, but the monopolist cannot afford to go above $489.50. The reason is easy to comprehend. The extra 50 units brought in $9.99 apiece, or $499.50 of increased income. But, in putting them on the market, the price of all the other 1000 units was reduced 1 cent each, or a loss of $10 on the original 1000 units. The gain of $499.50 and the loss of $10 give the net gain $489.50—the maximum that the monopolist can offer.

Some students distinguish these two ideas by saying that the competitive producer can offer up to the *value of the marginal product*, while the monopolistic producer can offer only up to the *marginal value product*.[1] The meanings of the two terms are apparent from our illustration. The value of the marginal product is found by computing the amount of the marginal product (50 units in the illustration) and multiplying by the price ($10 in the illustration). The marginal value product is found by subtracting the total revenue from the smaller output from the total revenue with the larger output. As can be seen from our numerical illustration, if the price of the

[1] Some students use the term "marginal revenue" where we have just used the phrase "marginal value product." The two are not identical. The marginal revenue is the amount by which the total revenue is increased when one more unit of output is sold. The marginal value product is the amount by which the total revenue is increased when one more unit of the factor of production is used. If we were to find the marginal revenue in our example, we would have to know how the total income changed when 1 more unit (not 50 more units) of product was put on the market. If we use the approximation here which we have used heretofore in estimating marginal revenue, we would say that, since the total revenue increased by $489.50 when 50 additional units were produced, presumably the total revenue would increase one-fiftieth as much had we added 1 more unit. This would give us a marginal revenue of $9.79 for our example. This is only approximate, but if we compute it on this customary basis we can say that the value of the marginal product is the product of the selling price and the marginal product, while the marginal value product is the product of the marginal revenue and the marginal product.

product is decreased when the output is increased, the marginal value product will be smaller than the value of the marginal product.[1] The monopolist will either employ smaller quantities of the factors of production than would be employed under competition, or will employ the same quantities only at lower rates of remuneration. Thus what we say in the following chapters will have to be qualified in that direction insofar as our society is a monopolistic one.

30 · 11. Size of the Social Dividend. We noted in Sec. 30 · 1 that incomes depend at least in part on the size of the social dividend or distribuendum. The larger the social dividend, the more there is available for distribution. Before we start to take up, as we shall in the following chapters, the division of the national income among the particular factors of production, we should give a brief space to a description of the changes in the size of that income itself.

The past century and a half have seen unusually rapid changes in production techniques—changes referred to as the Industrial Revolution—although we noted in the preceding chapter that these changes have appeared in agriculture as well as in industry. The development of human knowledge has made it possible to increase man's output many fold. Of course, that means that the national income has been growing in absolute amount. Unfortunately, this growth is difficult to measure accurately. If the output of each product were to double, it would be easy to say that the total output had doubled. But changes are by no means so uniform. In the past 150 years the output of automobiles has increased from nothing at all, so we cannot compute any percentage increase. There is nothing on which to base our percentage. On the other hand, the output of hoop skirts and moustache cups and top hats has decreased. How can we throw all these things together and determine whether production has increased or decreased? Usually we do try to measure it in terms of value. How many dollars' worth of product have we manufactured? To be sure, we have learned that the dollar itself is a poor measure when passage of time is involved in a problem, as it is here.[2] Consequently, we should "correct" any such figures for changes in the general price level. If we follow this procedure, we get the data of Table 30 · 2, which are shown graphically in Fig. 30 · 1.

This figure makes apparent at once the fact that the total production has been increasing. There have been, to be sure, temporary interruptions; but the general trend has been upward. Most significant of the interruptions was, obviously, the one connected with the depression of the 1930's. The smooth curved line drawn on the diagram shows not what the production actually was (that, of course, is shown by the irregular jagged line) but what the

[1] For a more precise discussion of this matter see the Mathematical Appendix, Note 15, p. 915.

[2] See Chap. 21.

production would have been if it had increased regularly and uniformly at the constant rate of 3½ per cent a year. The student will see that the curved trend "fits" the data pretty well except for the period of the depression. In

Table 30·2. Total Production of Goods in the United States, 1863–1943, Measured in Terms of 1935 Dollars. Production Figures Are in Billions of 1935 Dollars*

Year	Total production	Year	Total production	Year	Total production
1863	$3.5	1890	$16.7	1917	$45.0
1864	3.9	1891	19.5	1918	44.6
1865	4.6	1892	18.3	1919	43.8
1866	5.5	1893	17.3	1920	47.2
1867	5.9	1894	17.0	1921	39.9
1868	6.2	1895	20.2	1922	49.4
1869	6.7	1896	20.0	1923	55.7
1870	7.0	1897	21.4	1924	55.6
1871	6.8	1898	23.2	1925	58.4
1872	7.9	1899	23.7	1926	59.6
1873	7.9	1900	23.9	1927	59.8
1874	7.6	1901	24 1	1928	63.6
1875	8.5	1902	28 0	1929	66.3
1876	8.5	1903	27.5	1930	57.4
1877	9.4	1904	28.4	1931	49.4
1878	10.0	1905	31.5	1932	40.7
1879	10.7	1906	33.4	1933	45.2
1880	12.2	1907	32.2	1934	51.4
1881	11.1	1908	29.4	1935	55.2
1882	13.4	1909	33.4	1936	62.7
1883	13.3	1910	34.5	1937	66.3
1884	13.9	1911	33.8	1938	65.4
1885	13.7	1912	38.4	1939	73.5
1886	14.5	1913	37.8	1940	79.0
1887	14.7	1914	37.5	1941	88.6
1888	15.9	1915	40.9	1942	99.5
1889	17.1	1916	43.7	1943	116.0
				1944	123.0

* Data for 1863–1937 taken from "Structure of the American Economy," p 377, National Resources Board, 1939 Data for 1938–1944 are rough approximations based on figures of national income in the "Statistical Abstract of the United States, 1944–1945, pp. 403 and 416–417.

other words, it would be fair to say, as describing the general situation, that the physical output of goods in the United States has tended to increase roughly at the rate of 3½ per cent per annum, compounded annually.

Yet, this is not in itself sufficient to tell us whether the people are better

off or worse off. If we double the goods available for consumption but treble the number of consumers, people will on the average be worse off rather than better. We recall from our study of population,[1] that the number of people in the United States increased materially during the period covered by Fig. 30 • 1. In fact, reference to the data shows that in 1863 the population was 34 million, and in 1944 it was 138 million. Thus the population had not only doubled, but redoubled. Figures on value of output have little significance until corrected for changes in the value of the dollar; even then, they have little significance until they have been put

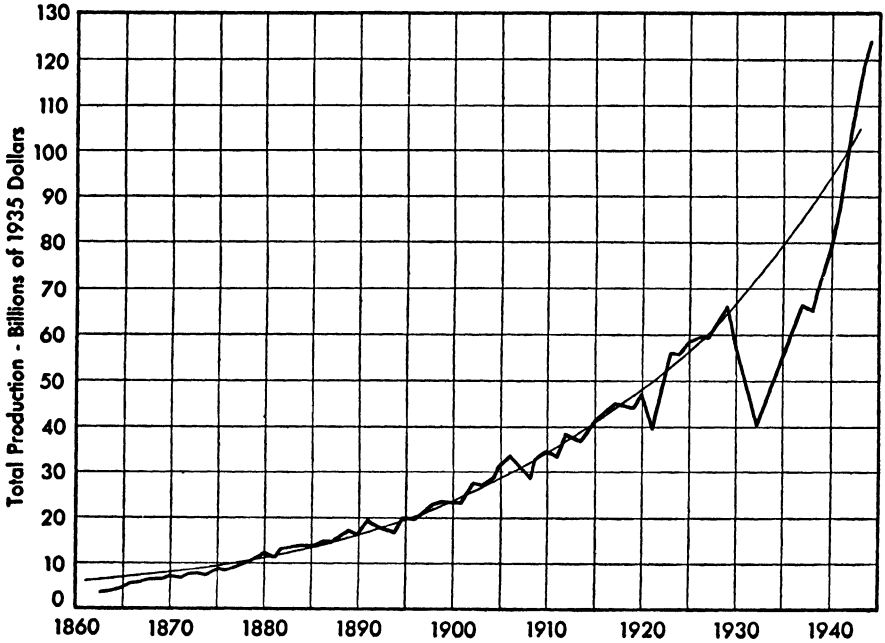


FIG. 30 • 1.—Total production in the United States, 1863-1944, in 1935 dollars, with trend line.

on a per capita basis. If the student cares, he may compute the per capita production in the United States by dividing each figure in Table 30 • 2 by the appropriate figure in Table 12 • 1. He will then find that the production per capita has increased as shown in Fig. 30 • 2. Here we see that the trend has been rising. In fact, the production of goods per capita has increased at the rate of about $6 per year. This may seem very small until we realize how small per capita incomes are anyway. If we think about it as covering a lifetime of 70 years, we see that it would mean an increase of $420, which is very large indeed compared with the sizes of the figures themselves. Even if we think in terms of a generation of 30 years, we can say that the tendency

[1] See Chap. 12, and especially Table 12 • 1.

has been for per capita output to increase in the amount of approximately $180 per generation.

As we said in our discussion of personal incomes in Chap. 4, the general

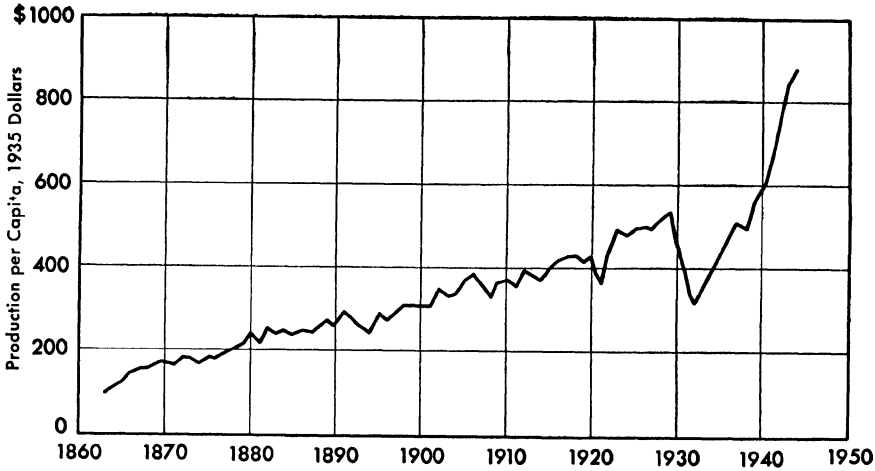


FIG. 30·2.—Per capita production in the United States, 1863-1944, in 1935 dollars.

level of income is higher in the United States than anywhere else in the world. This is even more apparent from the figures in Table 30·3, which show the per capita incomes in several different countries in the years 1929 to 1937. In this table the incomes for all countries have been put on a per

Table 30·3. Per Capita Incomes in the United States and Selected Foreign Countries, 1929–1937*

Country	1929	1930	1931	1932	1933	1934	1935	1936	1937
United States	$654	$588	$485	$374	$356	$407	$441	$508	$537
Australia	533	424	273	215	277	351	365	409	434
Austria	152	155	147	125	123	151	143	155	157
Bulgaria	70	59	53	47	59	73	77		
Canada	538	466	358	260	256	306	315	343	377
Chile	134	123	100	64	81	119	63	69	77
Denmark	281	282	249	177	186	236	239	249	
Estonia	82	75	62	53	58	62	66	76	
Finland	127	123	94	61	80	112	114	125	
France	233	229	214	193	239	288	271	276	204
Germany	279	257	208	163	215	313	353	389	421
Hungary	115	103	87	77	92	120	127	138	
Japan	87	80	75	43	44	52	52	54	
Norway	208	210	178	123	143	173	178	200	
United Kingdom	417	410	338	254	316	403	418	457	500
U.S.S.R.	46	50	75	70	73	83	96	120	136

* "Studies in Enterprise and Social Progress," p. 261, National Industrial Conference Board, New York, 1939.

capita basis and converted into dollars so that they will be as nearly comparable as possible. It is not to be supposed that the figures are exact; yet they do give a good idea of the general situation. The data of Table 30·3 have been computed on a different basis and by different methods from those used for Fig. 30·2, and consequently the results cannot be compared directly.

Perhaps the most important thing to remember, then, as we set out upon our study of distribution, is that the total social dividend has been increasing and that the rate of increase has exceeded the rate of increase in the population, so that year after year there is more and more for people to consume. This still leaves the question as to how it is divided, and this question we shall consider in the following chapters.

Questions

1. Define distribution.
2. "Most of the people of society are engaged in production." Would this include housewives? Lawyers? Ministers? College students? Babes in arms?
3. What factors determine the total size of the distribuendum in a nation?
4. How would you explain the fact that the typical school teacher or minister receives a smaller income than the typical corporation executive? How would you explain the fact that in many institutions the football coach gets a larger income than the head of the department of mathematics?
5. How would you explain the fabulous incomes of movie stars and of big-league baseball players?
6. "The demand for productive factors is a derived rather than a direct demand." Explain.
7. In the dispute as to whether it is better to distribute income in proportion to productivity or in proportion to need, which side do you take and why? Or, do you prefer some third basis?
8. Explain how the system used to divide the social dividend may in itself influence the size of the social dividend.
9. How do men justify paying people for the use of capital? Does the argument seem to you valid?
10. How does the justification of payment for the use of land differ from that of payment for the use of capital?
11. Do you suppose that in a completely communistic society, where all land belonged to the government, any payment would be made to anyone for the use of the best land? If not, what difficulties would arise?
12. Define interest, as used in economics.

13. What problems arise in connection with the idea that distribution on the basis of marginal productivity is just in that it gives each man what he has personally contributed to the distribuendum?

14. Why is it to be expected that incomes would be lower in a society based on monopoly than in one based on competition?

15. Distinguish value of marginal product and marginal value product.

16. In a competitive industry the output is 1500 units when 25 men are employed, and 1600 units when 26 men are employed. The selling price of the product is $2 per unit. What is the value of the marginal product?

17. In question 16, if the industry is operated by a monopolist and if, when the output increases from 1500 to 1600 units, the price of the product drops from $2 to $1.95, what is the marginal value product? What is the marginal revenue?

18. Using the data of Tables 30 • 2 and 12 • 1, compute the production per capita in 1935 dollars for each year, and make a graph similar to Fig. 30 • 2.

19. Explain the meaning of the statement that the figures in Table 30 • 2 are in 1935 dollars. What does it mean to say that the value of production in 1902 was 28 billion 1935 dollars? Explain why it is that corrected dollars are used in such a table rather than the actual values of goods produced each year.

20. Explain why the data of Fig. 30 • 2 are of more importance for our present purposes than those of Fig. 30 • 1.

21. Comment on the data of Table 30 • 3. For example, which countries showed the greatest decreases in income during the depression? What countries, if any, had income levels that surprised you, either because they were higher or lower than you expected as compared with the other countries in the table. Can you explain why the incomes in these cases were high or low?

31. ECONOMIC RENT

31 · 1. Rent Defined. In common speech "rent" is likely to mean any payment made for the temporary use of a durable good that is to be returned to the owner. The intent is that the good itself will be returned without damage, in as good condition as it was when rented, and that the rent itself will pay merely for the enjoyment or use of the good in the meantime.

Economists have long found it convenient in their own problems to define the word more narrowly, using it to mean, as Alfred Marshall has expressed it, "the income derived from the ownership of land and other free gifts of nature." Since we have defined land as the free gifts of nature, we can say that rent, as the economist uses the term, means any payment for the use of land.

This definition appears, however, to be much simpler than it is in practice. Natural resources are often so inextricably combined with human improvements that one is hard put to it to tell where the gifts of nature (the land) stop, and where the improvements of man (capital) start. If a farmer owns a piece of "land," which he lets someone else use, charging him "rent" for the use of it, it is likely that the payment covers not only the "original and indestructible powers of the soil" (to use a famous expression used by David Ricardo, one of the early English writers on rent) but also payment for such improvements as drainage, fencing, the removal of stones, and the clearing off of trees. In part the "land" that the renter receives is a gift of nature, but it is a gift of nature which has been altered and improved by the labor of man. The economist in his technical moments would insist that part of the payment is rent and that part of it is a return on the capital improvements which he would call "interest."

To be sure, we are less certain than men used to be that the original powers of the soil are indestructible. Erosion, leaching, overcropping—these and other things tend to decrease or destroy its original productive powers. Yet the economist still tries, as far as he can, to distinguish between that productivity which is a gift of nature and that which is the result of man's

efforts. It should be admitted at the start, however, that, although this distinction is useful for analytical purposes, it is nevertheless arbitrary and to some extent unreal, as are all useful classifications. The division of the factors of production into land, labor, capital, and possibly management is not a hard and fast, clear-cut obvious distinction. There are many difficult borderline cases that are hard to classify, and there are many, many problems where the differences between different kinds of labor or between different kinds of capital are far more important than the difference between labor and capital. The distinctions are particularly arbitrary in dividing labor and management, and in dividing land and capital. Yet, as we have said, the distinctions serve a pedagogical purpose, and we shall continue to use them.

31 · 2. Supply of Land. It is not necessary to go back here for a detailed discussion of the peculiarities of the supply of natural resources. The matter was taken up in Secs. 11 · 2 and 30 · 9. Suffice it here to repeat that land, being a gift of nature, is limited in quantity. This is the major reason for distinguishing it from capital. Some authors have insisted that no such distinction should be made, while others have insisted just as strongly that the distinction is basic to an understanding of major economic issues. The thing that we really need to keep in mind is the fact that factors of production differ greatly in the degree to which the quantity available can be and will be adjusted to market demands. The differences are those in degree rather than in kind. Some kinds of capital are relatively fixed over long periods of time and are in this respect more like land. Other kinds have a very short life and can be replaced on short notice. Land cannot be increased or decreased at all by men. Some writers, to be sure, would make a statement somewhat less extreme, and say that the quantity of land cannot be altered "significantly" by men. They would have in mind the fact that some land has been reclaimed from the ocean, etc. But, even in these cases, if we want to insist on the fine points, we would have to remember that draining land does not make new land. It merely combines land with capital. Every society has to take its land as a given factor and adapt the other factors to it.

From the standpoint of the individual entrepreneur, of course, the quantity of land is variable. If he wants more land and will bid enough for it, he can get it. When we consider the problems of individual businessmen or firms, the distinction between land and capital tends to disappear. When we consider social problems, we find that land differs on the supply side from any of the other factors of production; and this is the only fundamental way in which it does differ from the other factors.

31 · 3. Differences in the Quality of Land. We have talked so far as though land were homogeneous—a single uniform factor, the quantity of which is fixed. Yet everyone knows that this is by no means true. There are

great differences in lands, varying from completely sterile rock outcrops or sand to the most fertile black, fine, sandy loams. Lands differ, to be sure, in many respects—in color, chemical and physical composition, geological origin, etc. Each of these differences is important in its place, some to the geologist, some to the agronomist, some to the chemist, etc. But the economist, as we are already aware, is interested in land for the part that it plays in production. He cares about these other differences only as they affect productivity. When he talks about different "qualities" of land, he really means "differences in the productivity" of land. Two pieces of land may be very different geologically and chemically, but if they are equal in productivity the economist would call them land of the same quality.

Differences in productivity can be measured in two ways. The more direct and obvious is to measure the difference in the quantity of physical product under equal conditions. For example, if we have two pieces of land of equal size and put the same amount of labor, fertilizer and tillage on them, handling them in exactly the same way, we may find that one of them yields 100 bushels of corn and the other yields 83 bushels. These are the amounts of physical product. One piece of land is 17 bushels more productive than the other. If corn is worth 80 cents per bushel, the value of the product on the first piece is $80 and the value of the second is $66.40. Measured in this way, the first piece of land is $13.60 more productive than the second. The *physical product* of a piece of land is the actual quantity of product produced on it: 100 or 80 bushels of corn in our illustration. The *value product* of a piece of land is the value of the product produced on it: $80 and $66.40 in our illustration. Even though physical products are computed more directly and seem simpler, the economist is usually interested in the value products. There are two major reasons for this preference. One is that we often wish to compare pieces of land growing different products—one growing corn and the other producing onions. Which is greater, a bushel of corn or a bushel of onions? In the economic sphere, the values are the significant comparisons. Our other reason for preferring to work with value products is that businessmen seldom care about the total or gross product. They are interested, as we know, in the net product—in the difference between the product and the cost of obtaining it. If we work with physical inputs and outputs, we may find that our farmer got 100 bushels of corn by using 10 hours of labor, 100 pounds of fertilizer, etc. If we work with inputs and outputs in value terms, we can say that the farmer got $80 worth of corn by using $65 worth of factors of production. This allows us to ascertain directly that the farmer had a net income of $15 on the enterprise.

31 · 4. Rent at the Extensive Margin. Now that we have defined the principal terms that we shall need to use, let us start with a simple case. Suppose we have a newly settled country in which the population is still very small

and land is plentiful. There are, let us assume, hundreds and thousands of acres of land lying idle, land just as good as any that is in use. If under these circumstances Mr. Kinsey wants to use 5 acres of land and Mr. Sedgewick offers to let him use 5 acres for a rent of $5, Kinsey will naturally refuse to pay the rent. Why pay $5 to Sedgewick when there is idle land of just as high quality lying there to be had for the asking? When land is plentiful—when there is an abundance of land of the best grade to be had for the asking—no one is able to collect rent on any piece of land.

Of course, there are really many grades or qualities of land in any society. Some of it is far more productive than other. Let us suppose, now, that there are five grades of land in a society. In order to distinguish them we shall call them A grade, B grade, C grade, D grade, and E grade land. These pieces of land differ in productivity. This means, as we have seen, that we get different amounts of product from them if we give them the same treatment. Let us suppose, then, that we put equal amounts of labor and capital on each of these five pieces of land. To be specific, let us suppose that we put $60 worth of labor and capital on each. And, with these equal applications of the other productive factors, we get the following yields of potatoes:

Land	Bushels of Potatoes
A grade	75
B grade	60
C grade	50
D grade	30
E grade	15

We see that the different pieces yield different amounts of physical product, which is all that we meant in the first place when we said that they differed in quality.

If the population in our society has grown large enough so that all the A grade land is taken up and all the B grade land likewise, and if we are even using part, but not all, of the C grade land, what will be the situation with regard to rent? Kinsey now wants to take up a piece of land. To be sure, he can get some of the idle C grade land for nothing. But Sedgewick offers him the use of a piece of A grade land. Obviously, if he can get it for nothing Kinsey will prefer to use the A grade land. He can then get 75 bushels of potatoes for his $60, while if he uses C grade land he will get only 50 bushels for the same expenditure. In fact, if Sedgewick were now to offer Kinsey the use of A grade land and were to charge him a rent of 20 bushels of potatoes, it would pay Kinsey to use A grade land and pay the rent rather than to use C grade land. If he uses A grade land, he will get 75 bushels of potatoes and give up 20 bushels as rent, leaving a net return of 55 bushels for his $60 of expenditure. If, on the other hand, he takes up the free

C grade land, he will spend his $60 and get back only 50 bushels. The 55 bushels on A grade land that are left even after paying the rent are better than the 50 bushels that he could get on C grade land without rent.

Under circumstances of this kind, C grade land would be "no-rent" land, or "free" land. But the owners of A grade land could collect a rent. It is evident that they could charge a rent of 20 bushels, as we have just seen, or, following the same reasoning, that they could charge 22, or 24, or 24.9 bushels, or anything up to 25 bushels—the difference between the physical product on A grade and on C grade land. Similarly, the owners of B grade land could charge up to 10 bushels rent.

Rents are often, in fact, collected as we have here suggested, by taking part of the product itself. Such rents are called "share rents," since the owner of the land shares in the crop with the tenant. Often, however, rent is paid in money, and it will be worth our while to expand our problem to take account of the monetary aspects. If we continue with our numerical case, we may assume that it just barely pays to use C grade land if no rent is charged for it. Such land is called "marginal land"—land so poor that it barely pays to use it if no rent is charged. It would be reasonable to ask the student what the price of potatoes must be if it barely pays to use C grade land without rent. The answer is that the price of potatoes must be $1.20 per bushel, and we find this answer as follows: If it barely pays to use C grade land without rent when we are putting $60 worth of the other factors into it, it must be that we are just barely getting back our costs. But these costs are $60, and we are getting 50 bushels of potatoes on the C grade land. If 50 bushels of potatoes are worth $60, then each bushel is worth $1.20. Hence the price of potatoes is implicit in our statement that C grade land is marginal land.[1] If the price of potatoes is $1.20 and if we know the physical products, we can easily compute the value products, since they are simply the values of the physical products at whatever the price happens to be. In our present case, taking a price of $1.20 per bushel for potatoes, we can rewrite our table.

Grade	Bushels	Value
A	75	$90
B	60	72
C	50	60
D	30	36
E	15	18

[1] The student should test his understanding of this point by seeing if he can tell what the price of potatoes is if each of the other grades of land is marginal. For example, if B grade land is marginal, the price of potatoes is $1 per bushel, since the 60 bushels produced on that land must just be returning the $60.

Following the same reasoning as before, Kinsey will just break even if he goes on C grade land. He will spend $60 and get $60 back. (This is why we call this land marginal land.) He will be willing to pay Sedgewick a rent of anything up to $30 to use A grade land rather than use C grade land. If he pays Sedgewick $20 rent, his situation will be as follows:

Paid for labor and capital	$60	
Paid for rent	20	
Value of product	90	
Surplus left over		$10

If he uses C grade land for nothing, his situation will be as follows:

Paid for labor and capital	$60	
Value of product	60	
Surplus left over		Nothing

Under these conditions there will be a rent on A grade land of $30 and on B grade land of $12. These are *money rents* or *cash rents* as distinct from the *share rents* that we computed before. The student will see at once that the cash rent is merely the value of the share rent at the existing market price for the product. The share rent on A grade land was 25 bushels of potatoes, and at $1.20 per bushel these potatoes would be worth $30, which is the cash rent.

One or two questions are certain to arise in the student's mind at this point. First he will ask, why would Kinsey use C grade land at all if he has to spend $60 on it and gets only $60 back? Will he work for nothing? The answer is that he will not work for nothing, and he will not have to. We have said that he would get back enough to pay for his labor and capital. He gets back $60 to cover these other factors. He merely gets nothing to pay for the land.

Another question is this: To be sure we can see that Kinsey would be willing to pay up to $30 rent for A grade land *if he has to* rather than use C grade land, but why does he have to? The answer is that we have assumed that our society has grown large enough so that we cannot get all the potatoes we need from A grade land. It is necessary for someone—Kinsey or someone else—to use some of the C grade land if we are to get the crop we need. In fact, consumers have bid the price of potatoes up to $1.20 per bushel in an effort to stimulate production, and at this price it will actually be worth men's while to use C grade land if they can get it without paying rent for it. Kinsey does not have to use C grade land, but some producers will use it. And any of these producers who do use it will find it to their advantage to use A grade land if they can get it for less than $30. That is exactly what we have been pointing out. If Kinsey offers only $20 for A grade land, someone else who is using C grade land will offer $25, since it is better to pay $25 for

A grade land than to use C grade land free. Competition among users of land will make it certain that owners of A grade land will not have to accept much less than $30, and the difference in the amounts of physical product makes it certain that no user will be willing to pay much more than $30. Hence the rents will tend to be as follows:

Grade	Rent
A	$30
B	12
C	0
D	Not used
E	Not used

31 · 5. Submarginal Land. In the example above, we would call D grade and E grade land "submarginal land," land that it does not pay to use at given prices. To be sure, if the price were to rise enough, these lands would come into use. If, for example, the price of potatoes were to rise to $2 per bushel, D grade land would be marginal and C grade land would now pay a rent of $40. The rents on B grade and on A grade land would rise to $60 and $90, respectively.[1] Land that is submarginal at one price may cease to be submarginal at higher prices. We have already defined marginal land as that which at existing prices barely returns the other costs without paying any rent. Land like A grade or B grade land in our original illustration we call "super-marginal," or "superior land," land so productive that, at existing prices, it brings in enough income to pay off all the other factors used on it and still have some balance left to go to its owner. Rent is ordinarily paid for the use of superior land, but no rent is paid for marginal land, and submarginal land is not used.

Just as land that is marginal or submarginal at one price of the product may be superior at another price, so land that is marginal or submarginal for one use may be superior for another. Land that is submarginal for growing cotton may be superior for raising wheat; land that is submarginal for raising lettuce may be superior for grazing beef. Usually the owner or renter of land puts it, insofar as he is able, into the particular product in which it will yield the largest net value product—the largest net return after paying off the costs. Therefore, in computing rents, we compare the value products of pieces of land in their best uses.

It is plain that when we say that a piece of land is marginal we always mean "at present prices." Whether the land is marginal or not depends on the price. We have to know the price before we can tell whether the land is marginal, superior, or submarginal. This we can see all around us

[1] The student should check these results, using the same methods used in the illustration. The amounts of physical product will remain as before, but at the new prices the amounts of the value product will increase.

in practice. During the war, when prices of farm products soared to high levels, it paid farmers to use much land that was not profitable to use at the low prewar prices. The land that had been submarginal became superior at the new price level. And, as we noted in Chap. 29, the acreage of land in cultivation did expand greatly as a result of the higher prices. In the preceding numerical illustration, the price of potatoes was not $1.20 because C grade land was marginal. On the contrary, C grade land was marginal because the price of potatoes was $1.20 and, if the price were to change, it would stop being marginal. At higher prices it would be superior; at lower prices it would be submarginal.

31 · 6. Rent at the Intensive Margin. The reasoning of the two preceding sections has indicated that when men use land of different grades, rent will be paid on the better grades in use. But we may still ask why men use the poorer grades of land at all. Why, if they need more potatoes, do they not raise them on the best land? It is a matter of common observation that yields can be increased by applying more fertilizer, tilling oftener, etc., and it should be possible to get the additional potatoes that we need as our society grows larger by tilling the best land more intensively rather than by using poorer and poorer land.

The question is a natural and fair one. But we must remember in connection with it what we learned in Sec. 14 · 3. To be sure we can get more and more product from the best land, at least within limits, by cultivating it more and more intensively; but sooner or later we begin to get smaller marginal and then smaller average outputs. We can illustrate the situation, using our old familiar methods, by assuming that the A grade land, as before yields 75 bushels of potatoes if we apply $60 worth of labor and capital to it. Now what will happen if we apply more labor and capital in an effort to increase the output of the A grade land—if we apply another $60 worth, and a third $60 worth, and a fourth, etc.? We know from our study of production that the results will be of the general nature shown in the ac-

Amount of labor and capital used	Total product, bushels	Marginal product, bushels	Amount of labor and capital used	Total product, bushels	Marginal product, bushels
$60	75	75	$240	240	40
120	140	65	300	260	20
180	200	60	360	270	10

companying table. The added labor and capital do increase the total output, but the marginal output falls.

Assuming that this table shows the facts and that our society has grown

large enough to require 200 bushels of potatoes, what can we deduce from the data? In the first place we can see that we cannot get the needed 200 bushels without applying the third "dose" of labor and capital, the third $60 worth.[1] But this third $60 worth increases the output by but 60 bushels. The farmer will not be willing to produce these most expensive "marginal" 60 bushels unless he can get his expenditure back and unless he can recover the $60 that he will have to spend for labor and capital. Evidently if he is to get $60 for 60 bushels, the price must be $1 a bushel. Again the price of the product is implicit in the data that have been given us. If society expects to get as much as 200 bushels of potatoes under these conditions, it will be necessary to pay $1 a bushel. If the price is lower, the output will be less than 200 bushels, because the farmer will not find it profitable to apply the third dose of labor and capital.

If the price is $1 a bushel and if the farmer gets a total output of 200 bushels, his income will be $200. Yet the table itself tells us that he can pay for all the capital and labor applied with $180. There is a surplus of income over outgo. There are $20 left over after all the other factors have been recompensed. These $20 do not go to labor or capital, which have already been paid their $180. These $20 are left for the landowner—a payment to him for the use of his land, because he had such productive land. If Kinsey of our former illustration were now to ask Sedgewick for the use of this piece of A grade land, Sedgewick could hardly afford to let him have it for nothing. If Sedgewick uses it himself, he can make a net income of $20 over all expenses. Even if Kinsey offers him $15 for the use of it, Sedgewick will prefer to keep it and get $20. In fact, anyone wanting to use this land will find that the owner cannot afford to give it up for less than $20 and the renter cannot afford to pay more than $20 for it, because if he did there would not be enough left to pay for the labor and capital. Under the circumstances that we have set up, the rent of this piece of land will tend to be $20 even if no one uses any poorer grade of land at all.

Rent would arise even if there were no differences in the quality of land —even if all the land in the world were homogeneous—if that land were scarce and if it became necessary to cultivate it so intensively that the marginal products were falling. Rent arises because some capital and labor have to be used less advantageously than other. The disadvantage may arise because the labor and capital are applied on poorer quality land, therefore bringing reduced returns; or it may arise because the labor and capital are applied on the best land but beyond the point of diminishing returns, thus bringing reduced returns exactly comparable to those on poorer land. It is

[1] To simplify the problem we have assumed here that labor and capital have to be added in $60 units. This makes it possible to give a table with but few entries. In practice it would often be true that the use of the factors could be varied in much smaller steps.

not necessary to have differences in the quality of land to have rent in a society. It is merely necessary to have differences in the yield of different applications of labor and capital.

If we care to formalize our problem a little more, we can see how the rent arises in still another way. We note from the table that the first application of labor and capital brought 75 bushels, the next brought 65, and the third (which just paid for itself) brought 60. This last application was the "marginal" one, since it just paid for itself. The 60 bushels cost $60 and, at $1 a bushel, brought back a return of $60. But the first application brought back a return of 75 bushels worth $75. In this case, the farmer spent $60 and got $75. There was a net gain or surplus of $15. The second application of $60 brought back $65, a net gain of $5. The third application of $60 brought back $60, no net gain at all. But the gains of $15 on the first and $5 on the second applications amount to a total of $20, which is the rent on this piece of land. The rent is again a difference, just as it was when we went out on to the poorer land. Then the rent was a difference between the value product on the given land and the value product on the worst land that it paid to use. Here the rent is the difference between the value products on the earlier and more efficient applications and the value product on the marginal or "last" application which barely pays for itself.

When we compute the rent on a piece of land by comparing it with the poorest land in use, we say that we have computed the rent from the "extensive margin." The extensive margin is the poorest land that it pays to use even if no rent is charged, or what we have called "marginal land." When we compute the rent, however, by using the best land more and more intensively and noting the difference between the productivity of the earlier and of the marginal applications of labor and capital on this good land, we say that we have computed the rent from the "intensive margin." On the worst land it barely pays to make the first applications of labor and capital. Even on the best land the farmer applies more and more labor and capital until finally the last application barely pays its way.[1]

31·7. Intensive and Extensive Approaches Combined. In practice, of course, the people of society do not have to make a choice between producing more and more on the best land on the one hand, or using poorer and poorer land on the other hand. They do both. As they need more potatoes to take

[1] We fall here into a lax terminology which confuses some students. It is not intended to suggest that one of the applications of labor and capital is "last" in point of time, and that it is less efficient because it is last. If the farmer is applying 1000 pounds of fertilizer, we do not think of some particular pound as being the "last" pound. If he eliminated any one of the pounds, the effect would be the same. If we were more careful in our speech, we would say that as the farmer adds more and more fertilizer the output of potatoes increases, but at a decreasing rate, and the farmer continues to add fertilizer until the rate of increase in the value product equals the rate of increase in the cost.

care of a growing population, they cultivate the best land more intensively and, in addition, bring into use poorer and poorer land.[1] They run into diminishing returns as they use the best land more intensively; but they also get reduced returns as they use the poorer land. The combined results might be like those of Table 31·1, in which we combine our two earlier problems by assuming that (1) there are five grades or classes of land and (2) as we use any one of the grades of land more and more intensively, we increase the total product but diminish the marginal product. Table 31·2 shows the marginal products computed from the data of Table 31·1. As before, it is assumed that capital and labor are added in $60 doses.

Table 31·1. Total Products on Various Grades of Land at Various Levels of Intensity, in Bushels

Amount of labor and capital applied	Grade of land				
	A	B	C	D	E
$60	80	70	60	50	40
120	150	130	110	90	70
180	210	180	150	120	90
240	260	220	180	140	100

Table 31·2. Marginal Products on Various Grades of Land at Various Levels of Intensity, in Bushels (Based on data of Table 31·1)

Amount of labor and capital applied	Grade of land				
	A	B	C	D	E
$60	80	70	60	50	40
120	70	60	50	40	30
180	60	50	40	30	20
240	50	40	30	20	10

As the student will have guessed from his previous experience, the marginal figures are more likely to be useful in economic problems than the total figures.

Let us assume that the population of our society has grown to the point where it is necessary to get a total of 340 bushels of potatoes. There are various ways of getting them. We could raise 260 bushels on the best land, and the rest on B land. We could put one application of labor and capital on each grade

[1] At least they have to do these things to get more potatoes unless they develop new techniques of production or new strains of potatoes which will make it possible for them to get the greater product with the same capital and labor and land that they used before. In other words, they have to cultivate more intensively and more extensively barring changes in the state of the arts.

of land, which would give us a total of 300 bushels, and we could get the other 40 bushels and more by putting a second application on A grade land. If we look at Table 31 • 2, we see that, if we put each application of labor and capital where it will yield most, we shall put our first application on A grade land and get 80 bushels of potatoes. Our next application could be either a second application on A grade or a first application on B grade land. Either would yield 70 bushels. As a matter of fact, we make both of these applications, since this will give us a total from all three applications of only 220 bushels. We can now continue with a first application on C grade land, or a second application on B grade land, or a third application on A grade land. Any one of these will bring 60 bushels of product. If we make two of these three applications—any two of them, it makes no difference which—we shall have the total of 340 bushels that we wanted. And we shall have produced the 340 bushels as efficiently as could be done, with the least possible outlay of labor and capital.

If we produce in this most efficient possible way, we shall be putting on two doses of labor and capital which bring back returns of only 60 bushels. In order to get men to spend $60 for labor and capital under conditions where they will get back only 60 bushels of product, the price of the product will have to be $1 per bushel. At this price, the farmer will get back enough to pay him for his labor and for his capital, but he will have nothing left to pay as rent—nothing for the use of the land.

How much rent will there now be on the various pieces of land? On A grade land we will get back either 210 bushels by spending $180 for capital and labor, or 150 bushels by spending $120 for capital and labor. (This will depend on whether either of the last two applications was put on A grade land. Possibly one was put on B grade and one on C grade land.) But no matter which turns out to be true, whether we get $210 worth of product by spending $180 or $150 worth of product by spending $120, we shall have a surplus of $30 left over after paying off the labor and capital. Under these conditions, there will tend to be a rent of $30 on A grade land. And on B grade land we shall be putting on either one or two doses of labor and capital. We shall either spend $120 and get 130 bushels worth $130, or spend $60 and get 70 bushels worth $70. Either way there will be a surplus of $10 for rent after paying off everything owed for labor and capital. The rent on B grade land will be $10. Finally, if we use C grade land at all, we will get back 60 bushels of product by spending $60 for labor and capital. There will be nothing left for the payment of rent on the land. Grade C land is marginal, or no-rent land; and D and E grades are submarginal. As long as the price of potatoes stays at $1 per bushel, it will not pay to use these latter two grades even if we can get them for nothing.

We note now that it makes no difference whether we compute the rent from the intensive or the extensive margin. At either margin, just enough

product is grown to pay for the labor and capital costs. Some farmers find that it just barely pays to take up the marginal land, while farmers on better farms find that it pays to cultivate their land more intensively until finally their added output also barely pays for itself. How far it pays to go depends on the price of the product.

The student may find it instructive to work through this problem again with different figures. For example, let us suppose that the society grows still larger and needs even more potatoes than before. In order to get the added potatoes, it will be necessary to use poorer land or to push the better land farther beyond the point of diminishing returns. In either case, the marginal costs will rise, and consumers will have to pay more for potatoes if they are to get the quantity now desired. Under such circumstances, suppose that the price of potatoes does rise to $1.50 per bushel. The student will find that E grade land is now marginal and that all the other land is superior. E grade land will pay no rent, D grade land will pay a rent of $15, C grade will pay a rent of $45, B grade land will pay a rent of $90. We cannot tell from our former tables just how much rent will be paid on A grade land, because the tables do not go far enough to show. If Table 30 · 2 were extended one step by adding another figure at the bottom of the column for A grade land and if the new marginal product when we apply $300 of labor and capital were 40 bushels (making the corresponding new total figures in Table 30 · 1 equal 300 bushels), then the rent on A grade land would be $150. These figures should be checked by the student until he is sure that he knows where they came from.

31 · 8. Ricardian Law of Rent. It will be noticed by reference to the preceding examples that, as populations grow larger and more product is needed, there is a tendency for rents on land to rise. This is true unless, to be sure, changes in the state of the arts make it possible to get the added product without added expense. In a new country where land is plentiful, rents are low or nonexistent. In older countries, burdened with a large population, yields per man tend to be low and rents high. Thus the growth of population tends to enrich the landowners at the expense of the workers and owners of capital. This tendency we see all around us in actual practice.

A natural result of this tendency has been to call especial attention to problems of land tenure and land rents in the older societies. Early in the nineteenth century these problems were particularly acute in England, and a retired stockbroker, David Ricardo, made them the subject of particular study. His reasoning was similar to that just presented and, although he was perhaps not the first man to understand these tendencies, he was the first to bring them to popular attention. As a result, the formal statement of the principle involved is usually called, in his honor, the "Ricardian Law of Rent" and is stated in some such form as the following: *The rent on any piece of land measures the difference between the value of the product obtained on it by the most efficient use of labor and capital, and the value of*

the product that could be obtained by applying the same labor and capital at either the intensive or the extensive margin. If we illustrate this principle by reference again to the numerical example just used, let us ask what is the rent on A grade land when the price of the product is $1 per bushel. The Ricardian law tells us that the rent will be the difference between the value of the product obtained from A grade land by the most effective use of labor and capital (and when the price of potatoes is $1 per bushel the best use gives a product of 210 bushels worth $210) and the value of the product that would be obtained by putting this same amount of labor and capital ($180 worth of them) at either the intensive or the extensive margin. But, at either margin, an expenditure of $60 brings back $60 worth of product, so at either margin this labor and capital, three doses of it, would return $180 worth of product. The actual value product is $210, and the value product at the margin would be $180. The difference, $30, is the rent on A grade land.

31 · 9. Rent Is Price-determined. Older writers on economics emphasized the fact, as they put it, that "rent is price-determined, not price-determining." They argued that the farmer's payments for land were not a part of the cost of the product like his payments for labor and capital—that rent was not a part of cost. Their argument was simple and lucid. They pointed out that one cannot find the rent until after the price is given. If we are merely handed Tables 30 · 1 and 30 · 2 and asked how much rent there will be on A-grade land, we cannot answer; but, if we are told that the price of the product is $1 per bushel, we can state that the rent will be $30. Similarly, if we had been told that the price was $1.50 per bushel, we could say that the rent was $150. The rent depends on the price. As the price goes higher and higher, it pays to take over poorer and poorer land and to cultivate the better land more and more intensively, thus raising the rents on the better grades because the price rose. It is not fair to say that the price rose because the rents were high. On the contrary, the only reason that rents are high is because potatoes sell for enough to make it profitable to use poor land or to use good land very intensively.

A famous American economist of a past generation[1] commonly stressed for his classes the importance of this principle, stating that, if one did not understand it, he was certain to be led into circuitous reasoning, and illustrating with the following bit of doggerel:

The price of pig
Is something big
Because the corn, you'll understand
Is high priced too
Because it grew
Upon the high-priced farming land.

[1] Professor H. J. Davenport of Cornell University.

If you'd know why
The land is high,
Remember this: Its price is big
Because it pays
Thereon to raise
That high-priced corn for the high-priced pig.

Here in the first verse the author says that products (pigs) are expensive because the factors of production with which they were produced (land) are expensive. In the second verse he turns around to say that factors of production (land) are valuable because they yield a valuable product (pigs). First, the pig gets its value from the land, then the land from the pig. Davenport, like many of his contemporaries, insisted that the line of causation ran only one way: that factors of production have a value only because their products are valuable. If consumers no longer wanted the products, then the factors that made the products would be valueless. If consumers suddenly stopped wanting lemonade, then lemon squeezers would lose their value (unless, to be sure, they could be transferred to some other use and make some other product that consumers still wanted).

There is an important element of truth in this line of argument. It helps us to see that human wants underlie all economic activity—that consumption, not production, is the aim. From the social standpoint, the argument is entirely valid. Society pays no more for potatoes because it has good land. The existence of the good land does not add to the cost. In a real sense the price is fixed at the margin, where there is no rent, and hence rent is not a part of the price. If I have an unusually productive piece of land that yields a very high rent and if something happens to reduce the productivity of the land and eliminate the rent, society will not get potatoes at lower cost than before. From the social standpoint (the one in which most economists are interested most of the time), rents are not part of costs. On the other hand, the individual farmer who rents a piece of land and grows products on it can correctly say that for him the rent is part of the cost. He must now realize enough on his products to pay for the capital and labor used and also enough to pay the rent. Rent is to him a cost just like any of the other outlays that he has to make. It is only when the social aspects of the problem are considered that we are justified in saying that rent is price-determined, not price-determining—that rent is not part of the cost of commodities.

31 · 10. Some Other Kinds of Rent. We have been discussing what can be described as "economic rent"—that return which is paid for the use of natural resources. In practice, if a man rents a farm and agrees to pay $500 per year as a "rent," the payment covers not only the use of the land but the use of buildings and other capital equipment and improvements. Part of the payment is economic rent and part of it is interest. Yet the farmer will say that he is paying $500 rent. This sort of rent we call "contract rent," since it

is the type commonly called for in actual contracts for the use of land and equipment. Contract rent usually involves payments for other factors in addition to land; yet it is contract rent to which the layman ordinarily refers when he says that he is paying a rent of so much per year.

If we go back to the numerical example of Sec. 31 • 7, we discover that the rent on A grade land, when the price was $1 per bushel, amounted to $30. If the owner of this piece of land, Sedgewick, rents it to Kinsey, he will charge $30 for the use of the land. Here we have an actual payment of money for the use of land (or it might, of course, be a payment of a part of the product.) But suppose Sedgewick decides to use the land himself. He will spend $180 for labor and capital and get 210 bushels of potatoes, which he will sell for $210. This will leave him a surplus of $30, which he will probably call his "profit." Yet we know that it is really a return that he gets because of the productivi·y of his land—because he owns a piece of land that is more productive than the worst land that it pays to use. The economic theorist would still speak of this return as rent, even if the farmer himself calls it profit. No payment is made to anyone for the use of the land, yet there is a return that depends on its productivity. In these latter cases, where the owner uses his own land and gets a surplus return because of the productivity of the land, we say that there is "implicit rent," as contrasted with "explicit rent," which arises when one person makes an actual payment to another for the use of land. Implicit rent arises when men use their own super-marginal land; explicit rent is paid when one person uses super-marginal land belonging to another. The cash rents and share rents about which we spoke earlier in the chapter are explicit rents.

31 • 11. Rents and Land Values. How much is A grade land worth when potatoes sell for $1 per bushel, if we use again Tables 31 • 1 and 31 • 2? We know already that, under these circumstances, there will be a net income —a rent—of $30 per year. But we also learned some time ago[1] that when property is desired for the income that it produces, which is clearly the case with land, its value depends on the amount of the net earnings and on the interest rate than can be earned on other investments—what we called the "going rate of interest." A grade land in our illustration yields a net income of $30 per year. If the going rate of interest is 5 per cent, we can capitalize the rent of $30 at 5 per cent and discover that the piece of land will be worth approximately $600. The value of any piece of land is roughly equal to the capitalized net rent. Marginal land yields no rent and has no value. It is worth nothing. Submarginal land also is valueless. But superior land has a value that depends on its earning power as compared with earning possibilities in other directions.

To be sure, it should be emphasized that, when we speak of the value of land as the capitalized net rent, we have to remember that men value

[1] See Sec. 26 • 9.

land not according to what it is earning and what the going rate of interest is, but according to what they think it will earn and what they think the going rate of interest will be. Values of durable goods depend on men's *expectations*. Perhaps there was a bad drought this year, and the land yielded nothing. That does not mean that it has no value. If men expect it to yield a given product, if they expect future prices to be such that it will pay to cultivate at a given level of intensity, and if they expect earnings in other directions to be at some given percentage rate, then they will set a value on the land that corresponds more or less closely with the capitalization of these expected figures. If we keep this element of expectancy in mind, we can say that land values do seem to reflect rather accurately capitalized net rents.

31 · 12. Rent of Urban Land. So far all our illustrations have had to do with potatoes. The student will realize that we might just as well have substituted wheat, corn, or ginseng roots; at any rate our illustrations seem to indicate that there is to be some physical product taken from the soil and that the rent will depend on the value product corresponding to this physical product. Such physical products are easy to see in agriculture. In fact, the student could easily extend the illustration to cover mining. He knows that coal mines, for example, differ in quality. Some are so poor that it barely pays to work them if one can get them rent free. Others are rich enough so that it pays to use them more intensively, putting in more labor and capital. Here, too, as we go deeper and deeper into the earth and as we try to supervise the work of more and more men and equipment, we get diminishing returns. There are extensive and intensive margins in coal mining as there are in potato raising.

How about those cases where land is used for factories or for retail stores? Here is it reasonable to talk about different "products" from different pieces of land? Can we talk about different "qualities" of land in the sense that the different lands differ in quantity of product? To be sure we can. If a prospective merchant is looking for a place to build a retail store, he knows that all sites are not equally attractive. Some will yield large amounts of business, while others will yield little or none. There will be differences in the "value products" here just as in the raising of potatoes. Perhaps some sites are to be had for the asking, at no rent whatever. Other sites are more productive, and the merchant will prefer to use them even if he does have to pay some rent as long as it is not greater than the differences in the value product. Some of the most valuable land in the world is land of this kind—not agricultural land, but land used for trading and merchandising. In some cases preferred sites in New York City or in Chicago actually sell at prices of several hundred dollars per square foot. This is not because these pieces of land have a good view or because they yield large crops of potatoes, but because they yield large crops of customers for businessmen.

As in agriculture, these best sites tend to be used very intensively. The

owners or renters find it worth while to apply a great deal of capital and labor to them. They build 20-story, or 40-story, or 80-story buildings, piling more and more investment into the land. And as the buildings go higher and higher, the point of diminishing returns is soon passed. Each extra story gets more expensive as foundations have to be strengthened and as floor areas are reduced by setbacks and by increased elevator space. Sooner or later the time comes when it does not pay to add more stories. One has reached the intensive margin. On the less productive sites, this point is reached soon. As we go out into the suburbs, the heights of building decrease, until we reach the area of 1- and 2-story buildings. One does not build skyscrapers in Storrs or Hanover or Amherst or Podunk, not because it could not be done, but because it does not pay to "cultivate" such land so intensively. All the problems of rent arise on urban land in the same way that they arise on potato land.

31 · 13. The Unearned Increment on Land. As noted in the preceding section, some land is very, very valuable indeed despite the fact that it is not particularly fertile. We learned in Sec. 11 · 3 that there are qualities other than fertility which give value to land, and these cases of extreme value are usually associated with what we called "location" or accessibility. As is always true with land, the productivity of these pieces of land, and consequently their value, is not something given them by their owners. It is a gift of nature, which men have appropriated. Many people believe that in the ordinary case there are social advantages from allowing men to own land. In the United States we have especially emphasized these advantages and tried to encourage private ownership. But, in the extreme cases, where land is worth thousands and even millions of dollars per acre, there is much natural criticism of the fact that private individuals are allowed to use, for their own benefit, the tremendous returns. Some of the largest fortunes in the United States are based on the ownership of land. In most cases the land was bought originally at very low prices. Then population grew in the area, and the demand for land increased. The land became more valuable, not because of anything that the owner had done, but because of the growth of population. Society, not the landowner, is responsible for the increment in value; yet the landowner reaps the reward.

Such increases in value of property, brought about by changes in social conditions over which the owner of the property has no control, are called "unearned increment"—an increase in value which the owner enjoys without having earned it. Let us give one specific illustration, taken from the city of Chicago. The figures are those of a Chicago real-estate expert,[1] and we give

[1] The expert was Mr. F. R. Chandler, and the data are published in the "Eighth Biennial Report of the Bureau of Labor Statistics of Illinois (1894)," pp. 276-279, and are here copied from L. C. Marshall, C. W. Wright, and J. A. Field, "Materials for the Study of Elementary Economics," p. 637, University of Chicago Press, Chicago, 1913.

here only extracts to show the value of a particular plot of land at 5-year intervals, although the original data show figures by single years. The plot, said to be the most valuable quarter acre in Chicago, is located at the southwest corner of State and Madison Streets. Its estimated value at 5-year intervals from 1830, when Chicago was first settled, to 1890 appear in Table 31·3. In 1894, the value is said to have been 1¼ million dollars.

Here is a small plot of a quarter acre of land which rose in value from $20 to $1,250,000 in 64 years, a single lifetime. The rise in value was not the result of the labors of the owners, but because the city of Chicago grew up around it. The people of the region, not the owner of the land, were responsible for the increase; yet the owner got the benefit. To be sure, the value of the land fluctuated as there were recurrent booms and depressions, but the general trend was upward as population grew.

Because in such extreme cases individuals profit tremendously without any contribution to merit their gain, it has often been proposed that society

Table 31·3. Value of a Quarter Acre of Land at the Corner of State and Madison Streets in Chicago, Taken at 5-year Intervals, 1830–1890

Year	Value	Year	Value	Year	Value
1830	$ 20	1850	$17,500	1870	$120,000
1835	5000	1855	40,000	1875	92,500
1840	1500	1860	28,000	1880	130,000
1845	5000	1865	45,000	1885	275,000
				1890	900,000

should adopt some scheme to confiscate at least these major unearned increments. Most widely known of the proposals was one advocated by an American economist and politician, Henry George, who believed that the government should be supported by a tax that was to take the entire economic rent of land. Since he thought that this would raise enough revenue to support the government, he favored it as the only tax, or as the *single tax,* as it came to be known. Although this particular proposal has never gone far in the United States, the student will realize that today an attempt is made to take at least a large part of such unearned increments through the income tax, which is laid, among other things, on increases in the value of assets, including land.

31·14. Quasi Rents. Early writers, including Ricardo and his followers, supposed that land was unique among the factors of production in being the only one subject to diminishing returns. We know today, of course, that diminishing productivity is common to all factors of production and that, if a man owns a piece of machinery that is more productive than any other machinery, it will yield a return exactly like rent. To be sure, if this machine

can be easily reproduced, we do not have the same problem that arises with land—the fixed quantity. Of course, there are differences in quality among the units of every factor of production—different qualities of labor and different qualities of capital. Thus, just as we talk of different grades of land, we could talk of different grades of bricklayers or different grades of motor cranes. High-quality units of any factor of production ordinarily receive a higher return than do units of lower quality, and we have a situation exactly analagous to rent.

Although, then, the economist uses the term "rent" to refer to returns for the use of land, he recognizes today the fact that there are analagous returns for other factors of production. The owners of the better grades of capital receive a differential return, which continues as long as the less efficient grades continue in use. The length of this period depends, of course, on how long it takes to produce new capital and may be influenced by patent restriction, etc. These returns, which arise in connection with the use of the better grades of durable goods when several different qualities are in simultaneous use, are called "quasi rents." Economists have tended to lay more stress on them as they have come to realize that the distinction between land and capital is not so absolute and distinct as was once supposed.

31 · 15. Summary. Rent is a payment for the use of natural resources. Men are willing to pay for natural resources only when they are scarce. The better grades of land become scarce, as demand for the products of land increases, because of the operation of diminishing returns. As population increases and rising prices make it profitable, producers tend to move on to poorer grades of land and to use the better grades more intensively. In our illustrations we have simplified the problem by considering distinct classes or grades of land, while in practice there is a slow and continuous gradation from one grade of land to the next. Whether we go on to the poorer land or produce more intensively on the best land, we get diminished returns. Under such circumstances, the man who sees that he will have to accept low returns, either on poor land or at the intensive margin, is willing to offer something for the privilege of using the better grades of land. It is here that rent arises. The rent tends to take up the surplus—the extra value product over and above that necessary to recompense labor and capital. Always there is some submarginal land, too poor to use at existing prices, and some marginal land that pays no rent. The better grades or superior land yield enough more to make rent possible. The existence of these rents does not increase costs to society as a whole, yet the individual businessman rightly thinks of them as a part of his costs of production. Rents arise on land in all kinds of uses, and the value of the land tends to equal the capitalized value of the anticipated rent from it. The desirability of allowing rents to accrue to private individuals has long been questioned, and there are strong arguments on

either side. Many of the largest fortunes in this country have sprung from unearned increments on land, and various proposals have been made for confiscating these unearned increments.

Questions

1. Suppose we have two pieces of land. Piece *A* is very productive of cotton but rather unproductive of wheat. Piece *B* is very productive of wheat but rather unproductive of cotton. How would you decide which was the more productive?

2. Explain what is meant by physical product and value product.

3. Define marginal land; superior land; submarginal land.

4. Suppose that a farmer has settled on a piece of land so unproductive that he would not now settle it at present prices if he had the choice to make over again but, since he has his home there and is established, he actually does continue to farm. Would you call his farm marginal, submarginal, or superior?

5. Suppose that a man is operating C grade land. As he varies the amount of labor and capital applied, he gets the variations in total output shown in the accompanying table.

Capital and labor applied.....	$100	$200	$300	$400	$500
Bushels of total output.....	500	700	800	850	875

If the product sells for $2 per bushel, what will be the rent on this piece of land?

6. In question 5, if the price of the product is 50 cents per bushel, what will be the rent on the land?

7. At what price of the product would this be marginal land?

8. If the rent on this land is actually $500, what is the price of the product?

9. The accompanying table shows the amounts of total product obtained on each of five pieces of land if we applied various amounts of labor and capital.

Amount of labor and capital	Grade of land			
	A	B	C	D
$20	100	50	45	35
40	150	95	80	60
60	195	130	105	75
80	230	155	120	80

If the product sells for $0.80 per unit, what is the rent on each piece of land?

10. In question 9, what is the rent on each piece of land if the product sells for $0.40 per unit?

11. In question 9, what is the rent on each grade of land if C grade land is marginal?

12. What was the price of the product in question 11?

13. What land would be used and what would be the amounts of rent in question 9 if the price of the product were $2 per unit?

14. Explain what is meant by the statement that "rent is price-determined."

15. "Factors of production derive their values from the value of their products, not vice versa." Explain.

16. What is implicit rent?

17. What is the value of A grade land in question 10, if the going rate of interest is 5 per cent?

18. Explain why it is foolish to build tall buildings in country towns.

19. Explain what is meant by quasi rent.

20. Make up an original problem of your own, based on an original table similar to that given in question 9. Be sure that the table evidences both differing productivity on different grades of land, and diminishing marginal productivity on each grade. Determine what price of the product would make each grade of land marginal, and what the rent would then be on each of the other grades.

32. WAGES

32 • 1. Wages as the Return to Labor. As we noted in Sec. 30 • 7, the term "wages" is used by the economist to mean a payment or return for labor—for human effort. Yet labor is itself difficult to define. In what units is it to be measured? Possibly in hours of time, yet we all realize that 2 hours of labor by one man and 2 hours of labor by another man may not be equal. Nor can we, like the physicist, measure it in foot-pounds. Some sorts of work are physically arduous and exacting and are evidenced by the displacement of physical objects in such a way that the measurements of the physicist can be applied; other kinds may be equally exhausting and difficult without the performance of any "work" in the sense of the physics laboratory. How can we compare the work done by a hod carrier with that done by a student who is mastering a difficult passage in Latin or Greek, or the work done by a man sawing wood with that done by a lawyer arguing an important case before a court or a minister comforting a bereaved mother? There are no simple units of measure that can be applied to labor. In the view of the laborer himself work includes "not solely the exertion itself, but all feelings of a disagreeable kind, all bodily inconvenience or mental annoyance, connected with the employment of one's thoughts, or muscles, or both, in a particular occupation."[1]

Realizing the difficulty of measuring labor by the hour, since an hour of one man's labor obviously is not equal to an hour of another's, some authors have suggested that we measure it by its product. They would argue that two people producing the same thing have given the same volume or quantity of labor, while if *A* produces twice as much as *B*, *A* has given twice the volume of labor. But this suggestion is one that we positively and obviously cannot accept. Suppose, for example, that we have two men of equal physical and mental ability and equal training. We set one of them to work with his bare hands and the other to work with a spade, both of them digging a ditch. Does the man who works with his bare hands give

[1] MILL, JOHN STUART, "Principle of Political Economy," Book 1, Chap. 1, Sec. 1.

a smaller volume of labor merely because his output is small? The differences in product obviously do not reflect differences in the labor, but in the capital. If we add more and more men to a given plot of land and experience diminishing returns, does the reduced marginal product indicate that a smaller quantity of labor has been added? By no means. If we add more and more *equal* quantities of labor under these circumstances, we get continually smaller and smaller added volumes of product. We cannot measure the labor by the product. In fact, we are forced to admit that we have no good measure that we can apply to all the different kinds of labor; as a consequence, our data in the field of wages are rough and approximate even beyond most of those with which we deal in economics.

The problem is even more difficult because, as we pointed out in Sec. 12 · 1, physical and mental exertion are not always classified as work. A man playing a game of chess may exert himself mentally quite as much as does a lawyer in the preparation of his case; students playing touch football may exert as much physical energy as the same number of carpenters at work. The human being is not a laborer, in the economic sense, however, unless he is engaged in production. Men exert physical and mental energy in direct consumption, but only when they become factors of production do they become laborers. Effort exerted for the pleasure of it, as an end in itself, is not labor; but effort exerted as a means to an end—as an aid in the production of goods or services for the satisfaction of wants—is labor. Thus we depart again from the simple definition of the physicist, not because his definition is wrong but because it is not useful for our present purposes even though it is useful in his problems. Much that the physicist regards as work has nothing whatever to do with wages or with earning a living; and much of the real exertion of life which is directed toward the earning of a living and which does influence the labor market would not be called work in a problem in physics.

The student will note that the economist classes all productive human effort as labor, whether it is physical or mental effort, and regardless of the degree of skill required or the rate of pay received in return. In common speech, the word "labor" often refers to unskilled manual labor and carries with it an implication of suffering, hardship, and poverty. In this popular sense, a study of wages would be merely a study of the incomes of the poor. Under the arbitrary definitions of the economist, however, no such distinction is intended. Both rich and poor may receive wages, and wages often are, for both rich and poor, distinct from income. Wages are but a part of income, with rent and interest being other parts. Even a poor family may receive part of its income from interest on a small bank balance, or from rent of a room in the house, etc. In such cases the family's wages would be but part of its income.

32 · 2. Kinds of Income. In Sec. 4 · 1 we learned the distinction between

money income, real income, and psychic income. The student whose recollection of these concepts has become hazy should read that section again at this time, since everything said there about incomes in general could be repeated here about wages in particular. Money wages, especially the wages of skilled and professional workers, tend to move more slowly than the prices of consumer goods. As a result, in those periods of the business cycle when prices are rising, the real wages of these classes of people tend to fall. The term "psychic wages" is seldom used but, like psychic income, would refer to the satisfactions obtained in consuming the goods purchased with wages.

In everyday use there is a difference, at least by connotation, between wages and salaries. It is hard to put one's finger on the precise difference between the two concepts, but certainly the term "wages" usually refers to an income which is lower, less regular, and less secure and which arises from an occupation less acceptable socially than is true of those incomes which men call salaries. Salaries and wages differ in degree rather than in kind, and the economist does not distinguish between them at all. All incomes that reward productive human exertion are by him classed as wages.

In considering wages, the student is warned to remember that different sets of wage data are often not comparable. The commonest form in which wage information is presented is as *wage rates*—so much per unit of time. Thus a man may earn $1 an hour while another earns 50 cents an hour. The former wage rate is twice the latter. But, if the man receiving the higher rate works only 15 hours a week, while the man receiving the lower rate works 40 hours a week, the man with the high rate receives an income of $15 at the end of the week and the man with the low rate receives an income of $20. High rates of pay do not always mean high *earnings*. Men have to be quite as much interested in actual earnings as in wage rates. Short hours, seasonal unemployment, industrial accidents and disease, and many other forces may reduce the actual earnings even when wage rates are high. The wise man, considering which job to take, considers not only the wage rate and the annual earnings, but the *lifetime earnings* in his occupation. Different occupations differ tremendously in the average length of working life. In one line of work a man is "old" at thirty or thirty-five, while in another line a man of this age is barely making his start. The age at which people start work varies in different lines, as does the age at which the typical working life ends. We have but to ask at what average age men start as physicians and as newsboys to see extreme cases—or at what age professional boxers become too old to work as compared with college professors. One's total income throughout his productive period depends on his rate of wages, the regularity of his employment, and the length of his working life.

In recent years labor leaders and others have often talked in terms of what they call "take-home pay." This is the amount of money that the employee actually receives and "takes home with him." It excludes such things as deductions for income-tax withholding, union dues and hospitalization. Thus if, in a given week, a workman earns $45 and if his employer deducts $1 for union dues, $1.50 for hospitalization, and $2 as withholding tax, the workman receives $40.50 take-home pay. The student should realize in such cases, however, that the deductions ($4.50 in our illustrative case) are just as much a part of the employee's income as the remainder that he actually receives. These deductions are expenditures, and there is no more reason to overlook them in computing a man's income than there is for overlooking his expenditures for such other things as groceries, house rent, and clothing.

Many economists separate one particular kind of human effort from all others, and call it "management." The recompense of this particular type of labor they call "profits." We shall see later (Sec. 33·16) that there is no particular advantage in this separation and that the phrase "wages of management," which is becoming more and more popular among economists in recent years, describes the facts more accurately. If a man is paid to manage a business, he receives wages just as truly as anyone else, and the return which is properly called "profits" is not really a return for human effort at all. This point, however, must be postponed for later treatment.

32·3. Peculiar Nature of Labor. In some respects labor is so much like other factors of production as to make any distinction seem arbitrary. To a considerable extent, for example, capital and labor can be substituted one for another. A particular job can be done by using either many men and little machinery or few men and more machines. In fact, often capital can be substituted for some kinds of labor more easily than some kinds of labor can be substituted for other kinds. It has been remarked that there is more similarity between the work of an accountant and that of an adding machine than there is between the work of an accountant and that of a jockey. When some kinds of capital and labor are more nearly alike than different classes of labor, the segregation of labor in a separate class seems questionable.

Labor does have, however, some peculiarities which make the distinction seem worth while. There are certain sorts of human problems connected with labor that are not present with the other factors of production. Men cannot bring themselves to act toward other men as they willingly act toward plows and steam shovels or even toward horses and cattle—or slaves. Especially, it is noteworthy that labor and the laborer are inseparable. Capital and the capitalist can be separated. The capitalist can live in peace and security in New York or Chicago while his capital works for him in the fever-ridden tropics or in the snow fields of Alaska. The landlord, similarly can be separated from his land. But labor and the laborer are inseparable.

The laborer himself is forced to undergo the hardships of the economic struggle and to experience all those "feelings of a disagreeable kind, all bodily inconvenience or mental annoyance, connected with the employment of one's thoughts, or muscles, or both, in a particular occupation," which Mill mentioned. This direct and necessary contact of the laborer as a human being with the processes of production raises many human problems which are not present, or not present in the same degree, in connection with the use of land and capital.

32 · 4. Supply of Labor. The value of labor, like the value of other things, depends in the short run on the forces of supply and demand. Congress may declare, as it did in the Clayton Antitrust Act of 1914, "that the labor of a human being is not a commodity or article of commerce"; but all the lawyers in the world cannot alter the fact that from the economic standpoint labor is a commodity and an article of commerce—one of the most important ones—which is bought and sold in the market places of the world. As we set out on our problem of studying the forces that determine wages—the price of labor—it is but natural that we should pay some initial attention to the supply of and the demand for labor.

The total quantity of labor available for use in any society depends, as we saw in Chap. 12, on migration and the natural reproduction rate. These, in turn, are at least to some extent influenced by known factors. Thus, migrations depend to some extent on the relative levels of living in different parts of the world, and (what is largely the same thing) on the relative quantities of natural resources, capital, and population in different regions. The natural reproduction rate depends also on standards of living and on progress in science and medicine.

These factors, however, cannot affect the immediate or short-run supply situation. Over a considerable period of years they are of utmost importance, but short-run supply deals with the variations in the amount of labor offered in the labor market which occur at a given time in response to changes in the price of labor (wages). Some authors have given the impression that, at any given moment, the quantity of labor offered on the market is fixed. Yet this is by no means the case. There are always many people in society who are not working, but who could be drawn into the labor market if it were made worth their while. Thus, there are retired people who could go back to work, and young people in school and college who could give up or postpone their education in order to take jobs. There are wives who could leave the home to work in industrial plants, and people who could give up vacations to continue work. And, by no means least, most of those who are working could work longer hours if it became necessary. There is in every society a huge store of reserve labor which can be tapped in time of need—which is, indeed, tapped in time of war or other emergency. On

the other hand, except under the spur of patriotic war fervor or other emergency, labor tends to be relatively immobile and the short-run supply correspondingly inelastic.

Many authors have suggested that, in backward and undeveloped societies, there is an inverse relationship between the wage rate and the quantity of labor offered on the market; that, when the wage rate goes up, less labor is offered and, when the wage rate goes down more labor is offered. The explanation is that these "backward" people have not developed the limitless wants that characterize "civilized" men. Their wants are few and simple and fixed. If they receive low wages, they have to work many hours to get money enough to satisfy their wants; if their hourly earnings are higher, they can get sufficient income for their needs in a shorter time. They work until they have the necessary sum and then stop. It is possible that, when we consider the entire labor market, some such phenomenon exists even in our own society. There seems to be some evidence that, when wages are high, there is less tendency for wives and children to work, and for the total quantity of labor offered on the market to be smaller than when wages are low and wives are forced to take in washing while children peddle papers. Perhaps the most careful study ever made of this situation in the United States[1] indicates that the elasticity of the short-run supply of labor lies in the neighborhood of -0.15. This indicates that, if real wages rise by 1 per cent, there is a tendency for the quantity of labor offered to fall by about 15/100 of 1 per cent.[2]

The businessman, however, is not particularly interested in the nature of the supply of labor in general, but, rather, in the supply of particular types of labor, and within his own particular industry. Just as we long since noted[3] that the demand for food may be very inelastic while the demand for a particular kind of food is very elastic, so we can now understand that, even when higher wage rates and higher earnings that occur uniformly throughout industry may reduce the quantity of labor offered in the entire society, nevertheless it may well be true that higher wages for a particular kind of work or in a particular industry may increase the volume of that type of labor offered or the quantity offered within that industry. Thus, if wages of carpenters rise while wages in other lines remain unchanged, more people are likely to learn the carpenter's trade, and some who already knew the trade but were following other occupations are apt now to turn back to carpentering. The amount of carpenter's labor offered on the market will

[1] DOUGLAS, PAUL H., "The Theory of Wages," pp. 288-289, The Macmillan Company, New York, 1934.

[2] Douglas' study is not directly in terms of money wage rates and quantity of labor offered, but in terms of real annual earnings and number of persons employed per 1000 standardized population.

[3] See Sec. 7•6.

tend to rise, while the quantity of other kinds of labor offered tends to fall. Similarly, if wages of electricians rise in the automobile industry, electricians will tend to be attracted to it from other industries, and the higher wages will bring on to this market a larger quantity of labor. It seems to be true for any particular kind of work or for any particular industry that higher wages stimulate an increase in the quantity of labor offered and that the supply curve of labor, like the supply curves discussed in Chap. 14, is a curve with positive slope.

32·5. Noncompeting Groups of Workers. Although it is common for people to think of the entire population as making up the potential reserve labor supply, it must be remembered that only a small part of this population has the natural ability and training for most particular kinds of work. When a businessman goes out on the market to hire men, he usually is looking for men with particular aptitudes and skills. He may, to be sure, be looking for entirely unskilled labor. On the other hand, he is likely to be looking for a cabinetmaker, a blacksmith, a millwright, a painter, a plasterer, an automobile mechanic, a printer, a plumber, a metal worker, a tailor, a furrier, a welder, or a motion-picture projectionist, a photographer, a watchmaker, an upholsterer, a harness maker, a barber, an elevator operator, a bartender, a stevedore, a lawyer, or a railroad engineer—or some one of thousands of other skilled or semiskilled men to fill some post that cannot be filled satisfactorily by any Tom, Dick, or Harry. In such cases, the number of possible acceptable applicants is limited, not to the entire population, but to those people who have the necessary experience, training, and natural ability. Usually the watchmaker does not know enough to handle the lawyer's job, just as the ordinary lawyer does not know enough to handle the watchmaker's job. These men do not compete for each other's jobs. They compete only within their own groups.

To be sure, as the types of work done become more and more nearly alike and as the skill and training required become less and less, the possibility of competing becomes greater. Men can to some extent cross occupational lines. Especially is it true that men in skilled occupations often can, if they wish, compete for jobs that require less skill, while men in unskilled occupations usually cannot compete for those jobs requiring more skill. To be specific, many a surveyor could carry a hod, yet few hod carriers can run a transit; many a lawyer can operate his typewriter, but few typists are qualified for admission to the bar; many an automobile mechanic can shine shoes, but few bootblacks can adjust faulty hydraulic brakes. We might express this by saying that workers find it easier to compete downhill than uphill—that they find it easier to compete for jobs that require less skill than for those which require more skill. This means, of course, that the total potential supply of qualified workers available for any particular job depends

to a considerable extent on the amount of skill and training required for that job, and usually the less the skill and training required, the more people are qualified to compete for the job—the greater the potential labor supply for the job.

Thus, we see that labor is not a single homogeneous factor of production but is divided into a number of noncompeting classes. Within each class, the wages depend on the marginal productivity, and the marginal productivity depends on the number of workers in the class. It is natural, therefore, for us to be interested in learning what determines the numbers of workers within the various classes. In some cases, it is limited by the natural or innate ability required. For example, certain kinds of work cannot be done by color-blind men, and other types can be done only by persons of great manual dexterity. On the other hand, the number of persons in a given group often depends on the amount and difficulty of the training and preparation required. Few people are willing, even when able, to undergo the arduous training necessary to become a physician. Few people are able, even if willing, to hit 50 home runs per year in the major leagues. Many a man willingly moves into a low-paid classification rather than undergo more schooling or training. Moreover, some occupations stand higher in social esteem than others, and more people are attracted into these fields, getting at least part of their remuneration in the form of honor.

32 · 6. Demand for Labor. The demand for labor, like the demand for any other factor of production, is a derived demand, reflecting the demand for labor's products.[1] If labor promises to yield a large product,[2] employers will be anxious to use it and will be willing, if necessary, to pay handsomely for the use of it. But if it is expected that labor's product will be small, employers cannot be expected to pay much for it. Possibly the situation can be best understood by means of a simple numerical illustration. If it is thought that John Jones is productive enough so that, if he works for us, he can add $100 to the value of our products, we will be glad to hire him and pay him, if necessary, $50, $75, or even $98. If Ralph Gilman is so unproductive that he is expected to add but $10 to the value of our product, we cannot afford to offer him $50, or $25, or $15. We will not take him on at all unless we can get him for $10 or less. What we can afford to pay will depend on his productivity. If the price of labor (the wage rate) rises, we can afford to employ only the most productive workers; if wages fall, we can afford to employ some who are less productive. Thus we see that there will be an inverse or negative relationship between the wage rate and the number employed,

[1] See Sec. 30 • 4.

[2] As the student will understand from our earlier discussion, the employer is interested in the size of the value product rather than in the size of the physical product. To be sure, the two are closely related. See Sec. 31 • 3.

other things equal. At higher wage rates, *ceteris paribus,* we will employ fewer persons. In this respect, the demand for labor is similar to the demands for other goods, and we can think of the wage rate as being determined by the supply of and the demand for labor.

32·7. Marginal Productivity and Wages. It is now time to stop our more or less abstract discussion of the wage problem, and turn to concrete cases, even if at first they are hypothetical. Let us suppose that the figures of Table 32·1 show the total amounts of output (and the corresponding amounts of marginal output) which a manufacturer will get when he employs various numbers of men. These figures are like those studied in Chap. 14 and, to be consistent with the principles discussed there, they evidence, at least after the early stages, diminished marginal productivity. The price of this product will be determined by supply and demand, and let us assume that the product does sell for $2 per unit. Then the marginal value

Table 32·1. Hypothetical Case Illustrating Wage Determination

No. of men employed	Total output	Marginal output	No. of men employed	Total output	Marginal output
1	20	20	7	147	12
2	45	25	8	157	10
3	75	30	9	165	8
4	100	25	10	170	5
5	120	20	11	172	2
6	135	15	12	173	1

products can be found by multiplying each figure in the last column of the table by 2. If, now, 10 men are competing for jobs in this plant, what will the wage rate be? If the employer has 9 men on the pay roll and is considering the employment of a tenth, and if he knows the facts shown in the table, he will realize that the total value of goods produced will rise, when the tenth man is employed, from $330 to $340, an increase of $10. This can be discovered from either the second or the third column in the table. The employer cannot, under such circumstances, afford to offer more than $10 for this tenth man. If the wage rate is $11, the tenth man will be refused employment. Then, to be sure, if he cannot find work elsewhere, he is apt to offer to work here for less than $11—to accept wages that the employer can afford to pay rather than to remain unemployed.

We have often emphasized the fact that the marginal output when we employ 10 men is not the output of some particular one of the men who is labeled as being "Number 10." It is not the output of the "last man employed" or anything of that kind. It is the amount by which the total output would be reduced if we dropped any one of the 10 men, not some particular

one who is the tenth one. If our employer now adds the new employee at $10 or less, it will pay him to lay off any employee whatever who is receiving more than $10. If one of the former employees has been getting wages of $11, the employer now sees that he can lay him off, save the $11 of wages, and lose but $10 worth of product. Under these circumstances it is important to note that the wages of all the workers will tend to be reduced to the value of the marginal product as a maximum.

But, while we can see that the employer cannot afford to pay more than the value of the marginal product (in our numerical illustration he cannot afford to pay more than $10 if he is to hire 10 men), we may well ask if he has to pay this much. What is the minimum that he will pay? The answer is that he will tend to pay as little as he can. He will not stop, if he can help it, with a wage of $10 but will cut to $9, $8, or $3. If he cuts below the value of the marginal product, so that a man who is productive enough to add $10 worth of product is getting wages of but, say, $7, it will be to the advantage of other employers to hire the worker away by raising his wages. If I am paying him $7, you will offer him $8, since you would gladly get the $10 worth of product if you can get it by spending $8. And some other employer will offer him $9 or $9.75. In fact, if there is competition in the labor market, *the wages of any particular kind of labor tend to equal the value of the marginal product of that labor.* To be sure, the labor market is not completely competitive. On the demand side, employers have tended to some extent to combine in the hiring of labor (even where they have not combined in hiring, we noted in Sec. 30 · 10 that, if the employer is a monopolist in selling his products, there is a reduction in the amount of labor that he is willing to take), and on the supply side the development of labor unions has to a considerable extent substituted monopoly for competition in certain segments of the economy. We shall soon see, however, that even these developments do not seriously alter the general principle given in italics above.

Let us go back to Table 32 · 1 for the moment. We noted that, if 10 men seek employment in this plant, the employer will be unwilling to pay more than $10 as wages. By the same type of reasoning, the student will discover that the wage rate will tend to be about $16 if 9 men are seeking work, about $10 if 10 men are seeking work, about $4 if 11 men are seeking work, etc. In fact, it will be immediately evident that, if there is diminishing marginal productivity and if wages tend to equal the value of the marginal product, *the greater the number of men seeking work, the lower will be the wage rate.* If we combine this with what we said in Sec. 32 · 5 to the effect that many more people can compete for unskilled jobs than for skilled ones, it is easy to understand why wages in unskilled occupations are low. So many people can do such sorts of work that the marginal product and the value

of the marginal product are low; in those tasks that require large amounts of innate skill or long periods of specialized training, the number of persons available for employment is small and consequently the marginal product and the value of the marginal product are high. We do not pay low wages to ditch diggers because ditch digging is unimportant, but because so many, many people can do it. We do not pay fabulous salaries to the best professional baseball players because it is so vital that their work be done, but because so very, very few people can do the work. Wages for a particular kind of work reflect, not how valuable that sort of work is *in toto,* but how valuable is the contribution that would be made if another man were attracted to the field or how serious the loss would be if one of the men in the field withdrew.

32·8. Wages in a Noncompetitive Market: Labor Unions and Marginal Productivity. Some people, not understanding the economist's statement that wages tend to equal the value of the marginal product, have argued that this can not be so and that it is disproved in practice by the admitted success that labor unions have sometimes had in raising wages. These people have said that the economist makes it appear that wages are set by forces beyond our control—by the value of the marginal product—and that, if such impersonal forces really did set wages, there would be nothing that unions could do in the wage field at all.

This misunderstanding can be most easily cleared up by referring again to Table 32·1. Suppose, as before, that 10 men are seeking work in this plant. We have seen that the wage will tend to be approximately $10 per man. Suppose, now, that the 10 men in the plant were to form a union and that they could get a closed-shop agreement binding the employer to hire none but union members. If, now, the union were to hold out for a collective bargaining agreement that would set wages at $20 per man, it would be better for the employer to sign the agreement than not to do business at all. To be sure, if he must pay wages of $20, he will not find it worth while to employ all the 10 men. He can lay off one man and lose but $10 worth of product while saving $20 of wages. In fact, he can lay off another worker and lose but $16 worth of product while saving another $20 of wages. *The union can raise the wage rate above its competitive level only at the expense of some degree of unemployment.* But it can, nevertheless, raise the wage rate, and from its own selfish standpoint it would in this case pay it to do so. When the wage rate was $10, ten of the union's members were employed. The total wages received were $100. Now that wages have been raised to $20, eight of the union's members are employed and the total wages received are $160. From a purely selfish standpoint it would, under such circumstances, pay the union to force up wages and to support its idle members by means of its own unemployment benefits. To be sure, the practice will be harmful to society as a whole. Consumers will be getting a smaller quantity of output

(157 units in place of 170, as we can see from Table 32·1) and yet the cost of this smaller output will be even greater than before. Such monopolistic limitation of output is common with those who have the power to put it into force—manufacturers, public utilities, labor unions, etc.—because it is to the advantage of those who have the power to use it; but such monopolistic limitation is always to the disadvantage of the general public.

The point that should be made here is that, when the economist says that wages tend to equal the value of the marginal product, he is not by any means suggesting that wages are set by forces beyond human control. On the contrary, anything that affects the value of the marginal product will tend to affect wages. Labor unions have long opposed immigration, and we can easily see how this would fit into the marginal productivity picture, how heavy immigration increases the number of people competing for jobs and reduces the marginal product and the wage rates.[1] Labor unions have advocated the closed shop and limitation of apprentices on the same grounds. But there is no reason why the marginal productivity need be raised by such restrictive measures. Labor unions have advocated health measures and training programs, which could make the individual worker more productive and raise his marginal productivity without reducing the total product. Far from denying that unions can affect wages, the marginal-productivity theory tells unions or others just what forces can be used in raising wages.

32·9. Elasticity of the Demand for Labor. Although it happened to be true in the numerical illustration of the preceding section, based on Table 32·1, that a union could raise the total wage receipts by raising the wage rate and reducing the number of people employed, this would by no means always be the case. In fact, reference to Sec. 7·4 indicates that the result of raising wages will depend on the elasticity of the demand for labor. If the demand is inelastic, then higher wages will reduce but slightly the number of persons at work, and the total wage receipts will be higher, as in our example of Sec. 32·8. If the demand for labor is elastic, higher wages will reduce very greatly the number of workers, and the total wage receipts will decline.

The elasticity of the demand for labor, like the elasticity of the demand for anything else, depends to a considerable degree on the availability and efficacy of substitutes. As noted earlier, to some extent one factor of production can be substituted for another. A farmer wishing to increase the size of his wheat crop can do it by using more land with the same labor and capital, more labor with the same land and capital, or more capital with the same land and labor. Cement may be mixed by a man with a shovel or a hoe, work-

[1] The value of the marginal product is decreased in two ways: (1) the physical marginal product falls and (2) as the total physical product rises, the price of the physical product falls. The marginal value product is the product of these two.

ing with little capital, or it may be mixed in an expensive power mixer using much capital and little man power. To the extent that capital can be substituted for man power, the demand for labor becomes more elastic, and the power of unions to raise the total wage bill tends to disappear. It would still be true, to be sure, that the raising of wages by the union would be possible. On the other hand, it could now be done only at the expense of a larger volume of unemployment. Those workers who still had jobs would benefit, but workers as a whole, as well as the general public, would suffer. Paul Douglas has estimated[1] that the elasticity of the demand for labor in general is somewhere in the neighborhood of 3. This indicates that an increase of 1 per cent in wages can be expected to reduce by about 3 per cent the quantity of labor demanded. If this estimate is even approximately accurate, it indicates that any advantage that accrues from raising wages above their competitive level accrues solely to people who still continue to work, rather than to members of the "working class" in general.

To be sure, if there is some reason, based on statistical investigation, to suppose that the demand for labor in general is elastic, one will recognize on a priori grounds that the demand for a particular kind of labor or for labor in a particular business will tend even more toward elasticity. This raises a serious question as to whether or not the efforts of President Hoover in 1929 and 1930, and of President Roosevelt in later years, to maintain wage rates during the depression were not exactly the opposite of the proper procedure. In the first place, had wages been reduced, it would have been profitable for businessmen to employ larger numbers of people, and fewer would have been thrown on relief rolls. Most people would probably agree that at such times it is better to have more people working at low wages than to have higher wages for the few who are employed but more people who are not working at all. A reduction in the wage rate would certainly have reduced the volume of unemployment. If the demand for labor is elastic, as Douglas and some other investigators believe, then the reduction in wage rates would not only have increased the number of persons working; it would also have increased the total amount being paid out as wages. Thus, those who talked in terms of maintaining the "buying power" of the workers by "holding the line" on salaries should have realized that there is at least a considerable probability that the total buying power of wages would have been higher, and certainly the number of people unemployed would have been smaller, had wage cuts been made promptly toward the beginning of the depression.

32·10. Long-run Level of Wages. If wages tend to equal the value of the marginal product of labor, what forces, in the long run, determine

[1] DOUGLAS, *op. cit.*, p. 488.

whether that value will be high or low, and thus whether wages will be high or low? Economists of a past generation thought that the answer was solely one of the number of workers. They believed that the total wage bill was at least roughly fixed, and that the amount of it going to individual workers depended on how many workers there were to share it, just as the size of a quotient depends, when the dividend is fixed, on the size of the divisor. This idea, which was called the "wages-fund theory," rested on the assumption that there was just so much work to do, and just so much available with which to pay labor. Today this theory finds little support. We now realize that there is not a fixed quantity of labor demanded, but that the quantity demanded varies inversely with the wage rate. Moreover, we realize that there is not a fixed sum available for payment, but that the amount available for payment depends on the productivity of the labor. This productivity depends on the number of men seeking employment, on their training and innate ability, on the kind of equipment and natural resources with which they work, etc.

The preceding statement raises still other questions about the long-run general level of wages. What, in the long run, determines the number of men seeking employment and the kind of equipment and natural resources with which they work? The followers of Malthus[1] believed that population tended ever to increase, and that the increase in population in the face of diminishing productivity forced the marginal product, and with it the wage rate, lower and lower, until it reached the bare subsistence level. David Ricardo, the man who stated the Law of Rent, helped to popularize this so-called "Iron Law of Wages," which stated that in the long run wages tend to be ground down to the point where workers can barely subsist on them. For example, Ricardo said, "Diminish the cost of subsistence of men, by diminishing the natural price of the food and clothing by which life is sustained, and wages will ultimately fall, notwithstanding that the demand for labourers may very greatly increase."[2] The argument was that, if wages were considerably above the subsistence level, men would marry early and raise large families, and the population would increase. Diminished productivity as a result of the increased population would reduce wages until they did reach the subsistence level. Ricardo and his followers were often called the "pessimists," and their view that men were bound in the long run to sink into poverty gave to the economics of their day the name "the dismal science."

Modern economists, as we learned in Chap. 12, give much more weight than did their predecessors to the "preventive checks." They realize that wages

[1] See Sec. 12 • 15.

[2] RICARDO, DAVID, "Principles of Political Economy and Taxation," Chap. 30, published 1817.

may be based, even in the long run, on a standard of living far above subsistence if men hold strongly enough to that standard.[1] Standards of living differ greatly in different parts of the world. People with low standards tend to multiply until they force their wages down close to the actual subsistence level. Other people, with higher standards, limit their population to that which can support higher wages. It is for this reason that many economists favor free trade in goods but not "free trade in people," or unrestricted immigration. Trade in goods tends to benefit both parties to the trade.[2] But, if people from overpopulated parts of the world are able to move to places of less dense population in order to share in the higher productivity of those parts, then the people of the sparsely settled regions cannot enjoy the benefits that they have earned through their own higher standards of living and their own sacrifice.

As we have said several times, the productivity of men depends to a very great degree on the other factors of production with which they work—the land and the capital. It is difficult for men, no matter how strong of body and mind, to wrest a good living from the desert, or even from fertile territory if they have no equipment to help them. The land resources of a society are almost wholly fixed, but the capital equipment can be varied. The society that is frugal, saving, and foresighted—that sacrifices in some measure its present possibilities for satisfaction in order that it may have equipment to aid in future production—that society will ordinarily find its men more productive. The relative proportions of labor and capital in a society determine in large degree the productivity of labor and the wage rate in that society. Although in the preceding two or three paragraphs we have talked in terms of limiting the size of the population, the really important thing is not the absolute amount of the population, but its amount relative to the land and capital resources with which it works. In order for the general long-run level of wages to be high in any society, it is necessary for the capital resources to increase more rapidly than the labor resources, or else for the people of the society to be sufficiently inventive so that they can discover newer and better processes, which will enable men to produce more

[1] We should not overlook the fact that the "pessimists" themselves recognized this possibility, even if they did not consider it likely. To quote several scattered sentences from Ricardo's fifth chapter, on Wages, "The natural price of labour is that price which is necessary to enable the labourers, one with another, to subsist and to perpetuate their race, without either increase or diminution. . . . It is not to be understood that the natural price of labour, estimated even in food and necessaries, is absolutely fixed and constant. . . . It essentially depends on the habits and customs of the people. . . . Many of the conveniences now enjoyed in an English cottage would have been thought luxuries at an earlier period of our history. . . . The friends of humanity cannot but wish that in all countries the labouring classes should have a taste for comforts and enjoyments, and that they should be stimulated by all legal means in their exertions to procure them. There cannot be a better security against a superabundant population."

[2] See Chap. 10.

output with the same capital. Capital accumulation, invention and ingenuity, and restriction of population are the three major things that stand between us and the working of the Iron Law of Wages, the three things that make it possible to maintain in any country a level of wages well above bare subsistence.

32 · 11. Importance of Wages among the Distributive Shares. It tends generally to be true that those people whose incomes are low receive all or almost all of their income in the form of wages, as a direct return for their

Table 32 · 2. National Income and Total Wages and Salaries (in Billions of Dollars) United States, 1929–1944, with the Percentage of Total Income That Was Paid in the Form of Wages and Salaries*

Year	Total national income	Total wages and salaries	Percentage as wages and salaries
1929	$83.3	$52 6	63
1930	68.9	47.6	69
1931	54.5	40 0	73
1932	40.0	31.0	78
1933	42.3	28.7	68
1934	49.5	32.6	66
1935	55.7	35.6	64
1936	64 9	40.0	62
1937	71 5	45.0	63
1938	64 2	41.2	64
1939	70.8	44.2	63
1940	77.6	48.6	63
1941	96.9	60.8	63
1942	122 2	80.8	66
1943	149.4	103 1	69
1944	160.7	112.8	70

* "Statistical Abstract of the United States, 1944-1945," Table 439, p. 403.

current exertions. At the higher income levels, on the contrary, we find larger and larger proportions of the income in the form of rent on land or interest on capital. Keeping this in mind, let us examine the general situation for income earners as a whole.

Table 32 · 2 shows the total national income of the people of the United States, and the amount and percentage paid out in the form of wages and salaries. Roughly similar figures are shown for earlier years and for the manufacturing industries alone in Table 32 · 3. Both sets of data indicate that just under two-thirds of all income in the United States is received

in the form of wages (including salaries) and that wages are roughly twice as important in the aggregate as all other forms of income put together.

Table 32·4 shows for 1941 the percentage of total income of people at various income levels which came from wages and salaries. One notes at once that in the low-income brackets substantially the entire income is in

Table 32·3. Percentage of Total Value of Manufactured Products Represented by Wages and Salaries, United States, 1889–1929*

Year	Percentage	Year	Percentage
1889	59	1919	61
1899	57	1921	67
1904	59	1923	61
1909	60	1925	59
1914	65	1927	59
		1929	56

* Article on Wages by Vladimir Woytinsky, "Encyclopedia of the Social Sciences," vol 15, p. 312, The Macmillan Company New York, 1937.

the form of wages and salaries, while at the astronomical incomes at the other end of the table over 98 per cent comes from other sources and less than 2 per cent from wages and salaries. When we recall, however, that a very large proportion of the people of the country are in the lower income brackets,[1] it is easy to understand that most people get all or substantially

Table 32·4. Percentage of Net Income from Salaries and Wages, United States Income Tax Returns, 1941, according to Size of Net Income*

Net income, thousand	Percentage	Net income, thousand dollars	Percentage
Under $3†	97.9	$50–$100	29.3
Under 5†	74.1	100– 150	22.2
5–10	50	150– 300	16.6
10–25	41.6	300– 500	10
25–50	35 2	500– 1000	4.2
		1000 and over	1.5

* "Statistical Abstract of the United States, 1944–1945," Table 287, p 272

† Incomes under $3000 are from optional Form 1040A which gives data on gross income only Data for incomes under $5000 are estimated.

all their income in the form of wages. To say that wages make up roughly two-thirds of the total income, while correct, is but one side of the picture. Surely 95 per cent of American families receive their major income in the form of wages and salaries, and probably four out of every five American families get little or no income in any other form.

[1] See Sec. 4·2.

32 · 12. Summary. Wages, the return for human effort, are difficult to measure at all, especially so when we put them in their most significant form of real wages or psychic wages. Yet they make up by far the largest part of the national income. Wages depend, like other prices, on supply and demand. The supply of labor depends in the long run on migration and natural reproduction, and in the short run the total quantity offered may vary inversely with the wage rate. For particular types of labor, however, or for labor in particular industries and firms, the short-run supply curve is positive and the short-run demand curve negative.

It is impossible to get far in talking about the wage rate as though labor were a single homogeneous factor of production selling at one uniform rate. On the contrary, workers are split into many noncompeting groups. Within any group the wages at any time tend to equal the value of the marginal product of that group and, as a result of diminishing marginal productivity, larger quantities of labor will be employed only at lower rates. The effect of changing wage rates on the total amount of wage payments depends on the elasticity of the demand for labor. Since this seems to be elastic, it would seem that higher wage rates tend to reduce the total amount paid to all workers. In the long run, the level of wages in any community depends on the rate at which capital accumulates as compared with the rate at which population grows and on the inventive genius of the people.

Questions

1. What difficulties are there in measuring the amount of labor?
2. What is meant by real wages? Do you suppose that real wages in the United States are higher or lower today than they were at the time of the Revolutionary War? On what basis do you reach your conclusion?
3. Would you expect wage rates to be higher or lower in industries where employment is irregular than in industries where employment is regular?
4. Since the population is fixed at any given moment, how can the quantity of labor offered on the market vary with variations in the wage rate?
5. Under what circumstances might it be true that smaller amounts of labor would be offered at high wage rates than at low ones?
6. What is meant by noncompeting groups? Are they completely or only largely noncompeting?
7. Explain the statement that it is easier for workers to compete downhill than uphill.
8. If, in connection with Table 32 · 1, seven men are seeking employment, what will be the wage rate?
9. What would happen if the wage rate were higher than your answer for question 8?

10. What would happen if the wage rate were lower than your answer for question 8?

11. "The greater the number of men seeking work, the lower will be the wage rate." This statement is true only under certain conditions. At what part, for example, of Table 32 • 1 would it not be true? Why?

12. It is noted in Sec. 32 • 8 that a union would find it advantageous to raise wages from the competitive rate of $10 to a collective bargaining rate of $20. How high would they raise the wages under conditions of Table 32 • 1 if they were actuated solely by selfish reasons and if they had a complete monopoly? Why?

13. Why is it that an employer who already is employing 9 men under the conditions of Table 32 • 1, and who now hires a tenth man, insists on paying them all at the rate of $10 a piece instead of just the last one? Cannot the others successfully point out that they are more productive, and that the employer cannot afford to lose their services? For example, if he tries to reduce the wages of the fifth man to $10, cannot this man claim that he adds $40 worth of product and refuse to accept the reduced wage? Explain.

14. State the wages fund theory of wages. Criticize it.

15. State the Iron Law of Wages. Comment on it.

16. Does Table 32 • 2 seem to you to show any significant trend? Any real tendency for wages and salaries to make up a larger or smaller share as time passes?

33. INTEREST AND PROFITS

33 • 1. Interest Defined. There is perhaps more difference in point of view among economists over the definition of the word "interest" than over any of the other terms that we have so far considered in this book. Some would define it as any return made for the use of capital; some would put it in more exclusively monetary terms and define it as a payment for the use of money or of loanable funds or of credit; some restrict it even more, distinguishing so to speak between different kinds of credit and money, and define interest as a premium that is paid men in order to get them to sacrifice liquidity in their investments—to accept a less liquid in place of a more liquid investment. Each of these definitions has its place. Each of them assists the economist in treating particular sorts of problems. But for our purposes here, in elementary work, we shall take the first and most general definition which states that *interest is any payment made for the use of capital.* In this respect interest is like rent. Both are payments made, not for the purchase of the thing itself, but for the temporary use of the thing. Both of them have to do, therefore, with durable goods, which are used over a period of time. Rents, however, are payments for those durable goods which cannot be reproduced—which exist in fixed quantities—while interest is paid for durable goods made by men, which can be reproduced if it is desirable to do so.

Most of the incomes of society arise either from the sale of goods and services themselves, or from the sale of the temporary and immediate use of goods and services. Wages are incomes of the former kind, while rent and interest are incomes of the latter.

The student is certain to have heard of interest payments mainly in connection with monetary transactions and, to be sure, almost all interest payments do take on a monetary form. This is so obvious that it is common for authors to define interest as a payment made for the use of money. Since money is a form of capital,[1] it is true, of course, that payments for the use

[1] See Sec. 13 • 1.

of money are interest; but, under our definition, payments for the use of any other capital are also interest. When men borrow money, one may well ask how they can afford to pay any interest on it, since "money is sterile" and produces nothing. The fact is that borrowed funds are ordinarily invested in machinery and equipment—in other forms of capital—and this capital is not sterile. Capital is productive, and the interest can be paid out of the product.

Perhaps it will pay to illustrate a simple payment of interest in a case in which no money is involved. Suppose that there is a crude society which has not yet developed any monetary system. The people rely on direct barter. In such a society a man, finding that he has no wheat to plant, goes to his neighbor to borrow seed. The neighbor has enough wheat on hand not only for his own seed, but for food. He has been intending to use his surplus for food. If, now, he lends a peck of seed to his destitute neighbor, he will have to sacrifice his own chance for present consumption, and he may well refuse to do this unless the neighbor promises to pay him back a peck and a half at harvesttime. In this case, the wheat is capital, being used for production rather than for direct satisfaction. And the wheat is productive. A peck of seed will yield many pecks of wheat at harvest. The borrower can afford, if necessary, to pay back more than he borrows, because the borrowing will be productive. In our simple example, the extra half peck of wheat that was repaid was interest on the loan, just as truly and characteristically interest as any money payment. In fact, it is certain that the phenomenon that we call interest would be just as characteristic of borrowing transactions in a nonmonetary as in a monetary society, although it is doubtless true that borrowing transactions themselves would not be so numerous. Interest is a phenomenon associated with borrowing and with the postponement of satisfactions, and is not essentially or necessarily monetary in form.

33·2. Form of Interest. In our society, on the other hand, almost all borrowing transactions take place through the medium of money, and we shall use monetary illustrations in our further explanation. Usually, when a man borrows money, he finds that he must repay more than he originally borrowed. For example, he may borrow $100 today and find that he must repay $105 a year from today. The difference between the amount borrowed and the amount repaid is the interest. We can put this in simple mathematical form if we let B represent the amount borrowed, R the amount repaid, and I the amount of interest. Then

$$I = R - B$$

In our example the amount borrowed was $100 and the amount repaid was $105, so the amount of interest was $5. In our earlier and nonmonetary example the amount borrowed was one peck and the amount repaid was a peck and a half, so the amount of interest was a half peck of wheat.

Usually, however, we are more interested in what we call the "rate of interest" than we are in the "amount of interest." In computing the rate of interest, as every seventh-grade pupil knows, we find a percentage. We find what per cent the amount of interest was of the amount borrowed. This we do by dividing the amount of interest by the amount borrowed and multiplying the quotient by 100. Thus, if we let i represent the rate of interest (as compared with I, which represents the amount of interest), we have the definition:

$$i = \frac{(R - B)\,100}{B} = \frac{100I}{B}$$

In our numerical example, the amount of interest was \$5 and the amount originally borrowed was \$100. Our formula tells us that the rate of interest was 500/100 or 5 per cent.[1] In our earlier nonmonetary example the amount of interest was a half peck and the amount originally borrowed was one peck. Our formula tells us that the rate of interest on that transaction was 50 per cent.

Of course, either the amount of interest or the rate of interest can be computed for any period of time desired. We might find that one paid \$20 of interest for the use of a sum over a period of 10 years; or we might find that a sum was borrowed for only 2 days. In order to get comparability, it is customary to quote interest always, unless otherwise specified, on an annual basis. Even if a man borrows for but 3 months, we usually find the corresponding annual rate of interest. The student will remember from grammar-school days that the finding of this corresponding rate involves the making of assumptions as to whether the interest is simple or compound. The difference between the two is marked, especially when the actual credit period is much longer or much shorter than a year; for our present purposes, we need not investigate the mathematical niceties of the problem. We shall confine our attention for the time being to the strictly economic aspects of it.[2]

33 · 3. Why Do We Have to Pay Interest? We have said that it is possible for a borrower to pay interest because capital is productive. If he borrows a peck of wheat, it will enable him to grow several pecks; if he borrows \$1000, he can invest it in equipment with which he can produce goods valued at over \$1000. Even though capital is productive and therefore able to earn an

[1] This example should make clear why we use the constant figure 100 in the numerator of our formula. If we computed the rate of interest by dividing the amount of interest (\$5) by the amount originally borrowed (\$100), our result would be .05, or five one-hundredths. Five *per cent* means "five out of a hundred," and five *one-hundredths* means "five out of a hundred." Five per cent and five one-hundredths are the same thing. When we wish to talk in percentage terms, we merely multiply the corresponding decimal by 100, and this is the purpose of the multiplication by 100 in our formula—to give an answer in percentage rather than in decimals.

[2] The simple algebra involved in finding equivalent annual rates is covered, for those interested, in Note 16 of the Mathematical Appendix, p. 916.

income which we call interest, why is it that the borrower has to pay it? The fact that a man can do something is not a sufficient explanation of the fact that he does do it.

The reason for payments of interest was covered in some detail in Sec. 13·2. There we learned that, because of time preference, men do not make capital at all unless they anticipate a reward for doing so. Unlike land, capital is made by men and does not exist at all unless men find it to their advantage to produce it. What capital we do have would wear out and disappear if men were not constantly encouraged to replace and maintain it. In order to make capital, men always have to sacrifice current consumption which would otherwise be possible, and they are not willing to make this sacrifice unless they anticipate even greater satisfaction in the future. It takes perhaps $1.05 worth of next year's satisfaction to equal in our estimation $1 worth of present satisfaction. Thus do we discount the future. Some men have so little foresight (we may alternatively say that they have so much time preference, or that they discount the future so heavily) that they willingly sacrifice almost anything in the future to obtain what seem to prudent men rather unimportant present advantages. Thus cases have been known of people who bought more and more goods on the installment plan until they had promised to pay installments greater than their total income; other cases have been known where men have offered to sell themselves into slavery for life in return for present satisfactions. Everywhere it has been found necessary to protect men against their own weaknesses in this direction—against the magnitude of their own time preference. Yet this time preference must be overcome if capital is to be created, and the payment of interest is necessary if we are to get the advantages that flow from the use of capital in production.

To be sure, statements concerning time preference can easily be exaggerated. Men differ greatly in this regard, and there are certainly cases of men who prefer future wants to present ones. They sacrifice even important present wants, skimping along in a miserly manner for the time being in order to accumulate a great store of wealth against the future. Such cases, however, are thought of as unusual and abnormal—even pathological. To be sure, a normal man does not consume everything in the present and overlook the future entirely. Complete disregard of the future is not necessary for time preference. It is merely necessary that of two *equal* satisfactions one prefer the present one to the future one; or that one insist on a *greater* satisfaction in the future if he is to give up a *smaller* satisfaction now. Actually what must be true to substantiate the idea of time preference is not that a man makes no provision whatever for the future, but merely that if he were given his entire life's income now he would not consume it at a uniform rate throughout his life (assuming that he can forsee the future without uncertainty) but would consume it at a decreasing rate, concentrating somewhat more of it

always in the present than could be allotted on the average for the future.

33 · 4. Gross and Pure Interest. Just as actual contract rent often includes payments that are not by strict use of our definition rent at all,[1] so payments that are called "interest" often include payments for other things than the actual use of reproducible durable goods. *Pure interest,* as distinct from *contract interest,* has no elements in it except the use of the capital, which is to be returned undiminished, undamaged, and unimpaired. Contract interest often includes payments for depreciation and replacement, for the cost of administration, or to cover the contingency that the loan will never be repaid at all. Pure interest assumes that there is no uncertainty with regard to repayment—that repayment is certain and that the capital will come back in its original condition. Likewise, contract interest often includes something to offset changes in the value of money. For example, during the period 1900-1913, the general price level rose in the United States at an average rate of about 1.8 per cent each year. If a man had lent money in 1900 and if the contract had called for "interest" of 1.8 per cent, the lender would in fact have received no interest at all. To be sure, he would have received a repayment in 1913 considerably greater than the amount actually lent, but the repayment would have been worth no more than the original amount. There would have been no gain in purchasing power whatever. On the other hand, during the period 1884-1896, the general price level fell in the United States at an average rate of about 2.5 per cent each year. A loan made for 10 years, starting in 1884, and calling for "interest" at the rate of 5 per cent, would actually have been returning 5 per cent measured in money, but a pure interest return of 7-5/8 per cent.[2] Thus, pure interest may be either larger or smaller than contract interest. Pure interest is the equivalent of a perpetual net income, divested of all elements of operating cost, upkeep, depreciation, or uncertainty. These latter elements are not properly payments for the *use* of capital, but for the production or the maintenance of capital. They are part of the cost of production of the capital, which must, of course, be paid; yet, if the owner of capital were to be repaid merely his cost of production, he would not care to produce capital at all.

33 · 5. Important Market Rates of Interest. Because actual market rates of interest are contract rates, including various charges for risk and for maintenance, etc., in addition to pure interest, we find not one single uniform contract rate of interest on the market, but several differing rates. Some of these are of great importance in financial circles while some of them have only

[1] See Sec. 31 · 10.

[2] In one year prices fall 2.5 per cent, or the value of money rises 2.5 per cent. Thus, if no interest were paid whatever, the value of the money repaid would increase by 2.5 per cent to 1.025 times its original value. But the contract calls for an increase of another 5 per cent of this amount, or 105 per cent of 102.5 per cent. Thus the actual increase is $(1.025)(1.05) = 1.07625$, or an increase of 7.628, or 7⅝ per cent.

minor and specialized uses. The following paragraphs describe very briefly some of the more important contract rates on the market.

1. The *call rate* is the rate on loans that have no fixed future date of maturity, but can be "called" at any time by the lender. The borrower obligates himself to repay whenever the lender requests it. These loans are made largely by commercial banks from surplus funds on hand, and the borrowers are typically security dealers and brokers. In fact, the major demand arises in connection with transactions on the New York Stock Exchange, and the New York City banks are the major sources of supply. Because there is no fixed date of maturity, so that the borrower is in a position to repay or the lender in a position to demand repayment at any time on 24 hours' notice, this type of loan tends to be very sensitive to day-by-day changes in the market situation. Although the call rate is ordinarily rather low as compared with other interest rates, in times of financial stringency it also tends to run to the other extreme. At present (1946) and for several years past, the stock exchange call loan renewal rate has remained at 1 per cent; yet in the Panic of 1907 this rate rose temporarily to 125 per cent.

2. *Federal Reserve Bank rediscount rates,* as we learned in Secs. 19·11 and 20·9, affect private borrowers only indirectly. It is the rate at which the Federal Reserve Banks rediscount commercial paper for their members and may be different at the different Federal Reserve Banks. We have learned that this rate is important in that it is one of the means by which the government's fiscal authorities attempt to control the credit policies of the commercial banks, and thereby to influence the trend of business. The rediscount rate is usually a low rate, since there is relatively little risk involved in lending money on paper that is secured by both the signature of the original borrower and the original lending bank. In the late summer of 1946, the rediscount rate was a uniform 1 per cent for all Reserve Banks, but a rate of around 2 to 4 per cent would perhaps be more typical. This rate is usually somewhat above the rate on bankers' acceptances, but below the rate that commercial banks charge their own customers. The rediscount rate is an "administered" rate in the sense that it does not fluctuate in response to the supply of and the demand for funds, but is set by the government independently of these forces at whatever level seems likely to accomplish the fiscal purpose of the authorities.

3. *Commercial loan rates* are the rates at which commercial banks lend to their own customers. As we have seen, they are usually discount rather than interest transactions, but it is easy to shift from one to the other if desired. Commercial loan rates usually differ in different parts of the country, being lowest in the northeast and growing higher as one goes south and west. They tend to be set to a considerable extent by custom and hence are rather rigid, changing but slowly with the passage of time. In many localities this rate

remains at 6 per cent, year in and year out. This rigidity seems to be more pronounced in rural than in urban centers. Rates in recent years have been much lower than was formerly true. For 1945, the average commercial loan rate charged customers by banks in 19 leading cities in the United States was but 2.39 per cent. In New York City this rate was 1.99 per cent, while in 11 Southern and Western cities it was 3.26 per cent. This rate is perhaps the most important of all interest rates to the businessman, since he usually has direct contact with loans of the type covered here.

4. *Savings bank interest* is the rate paid on savings accounts or on "time deposits" by commercial banks. This rate is usually rather low, in the neighborhood of 1 to 4 per cent. In recent years the rates have tended to run toward the lower end of this range. This rate naturally varies as between banks but is subject to general regulation by the fiscal authorities.

5. The *prime commercial paper rate* is the rate charged for commercial paper running for 3 to 6 months and sold in the open market. Usually only large and well-known corporations are able to float loans directly in the open market, and the rate would naturally vary somewhat according to the borrower's reputation. In 1946, this rate remained pretty steadily in the neighborhood of three-quarters of 1 per cent, although rates of 4 to 5 per cent were formerly considered normal.

6. The *bankers' acceptance rate* is the interest rate on acceptances (drafts that have been accepted) drawn against banks. These acceptances usually run for from 30 to 90 days and, since they are obligations of both the bank and the original drawer, their rates are usually low. A rate of about 4 per cent was considered normal before the war, but in 1946 the rate was half of 1 per cent or even lower.

7. The *interest rate on government borrowing* depends on the reputation of the borrowing government and on the length of the loan. If the borrowing government is weak or if it has repudiated earlier issues and still has a bad reputation, it may have to offer a very high rate of interest indeed if it is to attract funds. Thus, Chilean government bonds were selling on the New York Stock Exchange in July, 1946, at prices to yield over 40 per cent interest. On the other hand, bonds of the stronger governments are considered to be about as safe loans as can be obtained, especially if they do not run for too long a period. United States short-term Treasury bills were selling in New York in July, 1946, at prices to yield a third of 1 per cent. Usually, even with a strong government, long-term loans carry higher rates than short-term ones. This is perhaps in part due to the fact that investors realize that a government that is strong today may be weak before a long-term loan matures. No loan is ever completely safe. In July, 1946, U. S. Treasury bills with 3 months or less to run were selling at prices to yield about a third of 1 per cent. At the same time the government's certificates of indebtedness which had 9

to 12 months to run were yielding about eight-tenths of 1 per cent and 3 to 5-year Treasury notes were yielding about 1 to 1¼ per cent. United States government bonds with over 15 years to run were yielding between 2 and 2¼ per cent.

8. The *interest rates on corporate bonds* vary tremendously depending on the stability, reputation, and expected earning power of the issuing corporation and on the length of the credit period. The average rate is somewhat higher than that on the better government bonds. In July, 1946, the average yield on 40 bond issues of industrial companies was 2.59 per cent; on 40 bond issues of railroads 2.85 per cent; and on 40 public-utility company issues 2.70 per cent.[1] Perhaps rates of 4 to 7 per cent would be thought of as more nearly normal.

9. *Interest rates on real estate and on farm mortgages* depend largely on the location of the property pledged as security. In ordinary times these rates run in the neighborhood of 5 to 9 per cent. These rates are important, of course, to homeowners who borrow to buy or build their homes, but they are also important to those who rent their homes, since the credit cost must be passed on by the landlord in his rent.

10. *Rates on consumption loans* as distinguished from production or self-liquidating loans[2] are usually very high. This is largely because of the high costs and the large element of risk involved. Regulated small-loan companies usually charge about 40 per cent on their loans, while pawnbroker loans often run somewhat higher, perhaps up to 50 or 60 per cent. Unregulated "loan sharks" often charge what amounts to 300 or 400 per cent. Some of the rates are specifically allowed by law because of the large element of risk involved and the heavy losses incurred on bad debts; but the highest ones are usually extralegal and contain a high cost of outlawry along with the other costs. Even installment credit, by means of which consumers buy furniture, radios, refrigerators, etc., at their local furniture store or at a mail-order house, often involves costs of 20 to 30 per cent; when automobiles are bought on credit, as most of them are, the credit and other allied costs usually run to 20 or 30 per cent also. *Consumer credit is almost always very expensive* and should be investigated carefully and understood thoroughly before it is incurred.

33·6. Present and Future Values. The interest rate is a number that indicates a comparison between present and future values. If we say that the interest rate is 5 per cent, we mean that $100 today and $105 a year from today are of equal value. We have already run across this phenomenon when we capitalized the value of future incomes.[3] If we know how men value

[1] Data on current yields in this and the preceding section are from the *New York Times* business section for July 24, 1946, and from the July, 1946, *Federal Reserve Bulletin*.

[2] See Sec. 19 • 1.

[3] See Secs. 26 • 9 and 31 • 11.

the future as compared with the present, that is, if we know the interest rate, we can compute the present value of all the future income by the process which we called "capitalization." Other future incomes can be converted to their present equivalence by similar means, always using the interest rate in the process. The interest rate used, however, must be the actual rate at which men discount the future, and not some arbitrary, formal, or customary rate. The rates of interest quoted on securities and business documents of various sorts are not always by any means the actual interest. If this seems confusing, let us take an example. Suppose that we have a $100 bond which promises to pay 5 per cent. We would call it a "5 per cent bond," and the holder would receive $5 a year income from it. In order to omit for the time being the complications of maturity dates, let us assume that the bond promises to pay $5 a year forever. Bonds do not always sell at par. If I actually buy this bond for a market price of $50, then I receive $5 each year on an actual investment of $50. The bond calls itself a 5 per cent bond, but I am receiving an actual 10 per cent on my investment. If, on the other hand, I were to buy the bond for $125, then I would receive each year a return of 4 per cent on my actual investment of $125, since $5 is 4 per cent of $125. Again the bond calls itself a 5 per cent bond, but at this price it actually yields but 4 per cent. There is, then, a difference between the stated rate of interest and the actual yield of interest—the effective rate.

In the numerical examples of the preceding paragraph, the arithmetic involved was explained in a way that is apt to obscure the actual cause and effect. The number work makes it appear that the interest rate is 4 per cent because the bond yields $5 per year on an actual investment of $125. It is important for the student to understand that the causal connection runs the other way. We should say that a bond that yields $5 a year has a value of $125 because men discount the future at the rate of 4 per cent per annum. The process is merely that of capitalization, as the student will discover if he applies the capitalization formula of Sec. 26 • 9 to discover the value of property that yields a return of $5 per year when the rate of interest is 4 per cent. But the numerical examples should also have shown the student that he cannot always rely on the rate of interest stated in business documents. It may differ greatly from the actual effective rate of interest.

This difference between the stated and the actual rates is of considerable importance in practical business transactions, especially so in the field of investment. We find in the financial pages of the newspaper that a given 4 per cent $1000 bond, which will mature 20 years from today, is selling at an actual price of $874.49. What is the effective rate of interest? Our problem is complicated this time by the fact that we have a maturity date. In 20 years the full $1000 will be paid to us, despite the fact that we invested only $874.49. During the 20 years, we shall be getting $40 each year as interest, which is

4.57 per cent on our actual investment. But in the bargain we shall be repaid after 20 years the full face value of $1000, which is $125.51 more than we paid. If we consider that this gain of $125.51 is spread evenly over the 20 years, it amounts to an additional $6.28 per year, so we are really getting $40 of interest plus $6.28 increase in principle each year—a gain of $46.28. This amounts to 5.29 per cent effective yield.[1]

It is easy to see from our discussion that, if the actual interest rate changes —if men change their relative evaluation of the present and the future— there will be a consequent change in the value of all goods that promise a future income. There will be a change in the value of all durable property, such as land and capital. If the interest rate falls, we have an indication that men's values of future incomes are higher than before. Thus, for example, when the interest rate is 5 per cent, it takes 105 of next year's dollar to equal 100 of this year's dollars; if the interest rate drops to 4 per cent, it takes but 104 of next year's dollar to equal 100 of this year's. At the low rate of interest, 104 future dollars have as much value as 105 future dollars had at the higher rate. Future dollars have become more valuable and, because men place a higher value on future satisfactions, the interest rate falls. Perhaps it would be better to say that the two statements, "men place higher values on future satisfactions," and "the interest rate has fallen," are two ways of saying the same thing.

Because the interest rate is merely a method of expressing the relationship between present and future values, it should be plain that every change in the

[1] The mathematical problem of finding actual effective yields is a difficult one, and can be done only by approximation. In practice, the investor would use a bond table, which would tell him the actual effective yield. The method used here is very crude and gives considerable error in practice, especially if the sale price of the bond differs materially from its face value. Bond tables would show that the actual effective rate in the example used above is 5 per cent rather than 5.29 per cent. The method will serve, however, to illustrate the nature of the problem and does give a rough approximation of the effective rate. The method can be summarized thus: Subtract the actual sale value of the bond from its face value, to find the amount of premium or discount. If the bond is selling above par, the difference will be negative; if below par, the difference will be positive. Divide this premium or discount by the number of years until the bond matures, to find the amount of gain or loss each year. Add this algebraically to the amount of income that the bond pays each year. This latter value, of course, is found from the face value and the stated rate of interest. One now has the total income per year, made up in part of interest on the face value of the bond and in part of appreciation or depreciation in the amount invested. If this total income per year is divided by the amount actually invested, one discovers the rough approximation of the effective interest rate. To illustrate with a case where a bond is selling above par: Suppose a 4 per cent $1000 bond with 10 years to maturity sells at $1085.84. What is the effective rate? The bond will, as before, pay us $40 a year; but in 10 years it will pay back only $1000 instead of the $1085.84 which we invested. We will lose $85.84 in 10 years, or $8.58 per year. This loss of $8.58 taken with our income of $40 makes a net return of $31.42 per year, which is 2.89 per cent effective yield. Bond tables show that the actual effective yield for this case is 3 per cent.

interest rate brings about, or indicates, a change in the value of durable property. If a given bond was worth $800 yesterday and if the interest rate falls today (indicating that men care more for future income), the value of the bond rises above $800. If we think of the pure interest rate for capital that is maintained perpetually without physical deterioration, the effect of changing interest rates on values is obvious from our capitalization formula. For example, suppose that a given piece of capital equipment yields a return of $60 per year and that this return can be expected to continue indefinitely. If the rate of interest is 4 per cent, our capitalization formula tells us that this property is worth $1500. If the rate of interest rises first to 5 and then to 6 per cent, the capitalized value drops first to $1200 and then to $1000. The values of durable goods vary inversely as the interest rate. If the interest rate rises, values of durable goods fall; if the interest rate falls, values of durable goods rise. In fact, the new value of durable goods bears the same relationship to the former value that the former interest rate bears to the new interest rate. If V and v represent, respectively, the values before and after the change in interest rate, and if I and i represent, respectively, the old and the new interest rates, then

$$\frac{V}{v} = \frac{i}{I}$$

If this still seems abstract and far from the affairs of real life, the student will find little difficulty in understanding why it is true. Suppose that the present interest rate is actually 4 per cent. This means that in men's minds 100 current dollars and 104 next year's dollars are of equal value. Under such circumstances a bond, or any other durable property that promised to pay $40 a year, is worth $1000. If I put $1000 into it and hold it a year, I shall have the $1000 bond plus the $40 interest, and on our assumption these $1040 "next year's dollars" are worth exactly $1000 today, and the bond will be worth just $1000 to me. Now suppose that my attitude toward the future changes and that my interest rate becomes only 2 per cent instead of 4 per cent. In my mind 100 current dollars and 102 next year's dollars are of equal value. I am willing to invest $100 if it will yield a return of $2 each year. The same bond that we mentioned before yields a return of $40 each year and is, therefore, now worth $2000 to me. I would be willing to invest $2000 if I could get 2 per cent on it or if I could get $40 a year. The bond promises $40 a year and is worth $2000 at my new interest rate instead of $1000.

Actual values do, in practice, vary in this way. When interest rates are high the values of bonds are low; when interest rates are low, the values of bonds are high.

33·7. Demand for Loanable Funds. Having discussed at some length the importance of the rate of interest, we now turn to a consideration of how that rate is determined. The rate of interest is a price—the price of loanable

funds, the price for the use of capital, or the price at which we can convince people to postpone their consumption—and, like other prices, depends on conditions of demand and supply. We turn first to the former of these, the demand for loanable funds.

The demand for loanable funds is a schedule of the amounts which prospective borrowers would borrow at various prices or, in our present case, at various interest rates. It depends, naturally, on the insistence of the desire for funds on the part of prospective borrowers, on the uses to which they would put the funds, and on the degree to which they discount the future. The uses to which borrowed funds are put are of two major types. Some funds are borrowed by consumers to enable them to obtain and enjoy desired consumer goods sooner than would otherwise be possible. Other funds are borrowed by producers as an aid in production. In the former case, whether consumers will be willing to borrow or not will depend on the rate of interest demanded. If a consumer discounts the future personally at the rate of 4 per cent, he will not be willing to borrow at 5 per cent.

Consumers differ very decidedly in the rates at which they discount the future and, as the interest rate falls, more and more consumers are led to borrow. Many people, for example, borrow willingly at 3 per cent who would borrow reluctantly or not at all at 10 per cent. As the interest rate falls, the quantity that consumers stand ready to borrow rises, and we find a demand curve of the usual form, with negative slope. Producers, on the other hand, are borrowing because they expect the capital to be productive. They have in mind productive uses to which the borrowed capital will be put. If the capital can be obtained cheaply enough, it will pay to borrow it; if the cost of the capital is going to be even greater than its productivity, it does not pay to borrow. Since different businessmen have in mind different uses and since they anticipate different rates of productivity, it will again be true that, as the interest rate falls, more and more capital will be wanted by producers. Again the demand curve will have negative slope. Since the demands of consumers and of producers both have this characteristic, it is natural that the market demand curve for capital should be negatively inclined—that at high interest rates the quantity of capital demanded by borrowers will be small, and at low interest rates it will be large.

It is also necessary to keep in mind the fact that, as more and more capital is used, its marginal productivity diminishes. Thus, a businessman might be willing to borrow $1000 at an interest rate of 5 per cent, and he might be willing even to borrow a second and a third $1000 at this rate; but as he uses more capital and as the marginal productivity of capital begins to fall, sooner or later he will reach a point where an additional $1000 of borrowing will return too little product to be worth the cost. He will then, of course, stop borrowing. Even this same individual, if the interest rate is now lowered to 4 per

cent, may find that it will pay him to borrow somewhat more. His next $1000 of borrowing may promise him a return of less than 5 per cent, so it did not pay him to borrow it before; if it promises over 4 per cent, it will pay him to borrow it now. Thus, when the interest rate drops, we not only attract new borrowers who had not borrowed at all before, but also increase the borrowings of those who had borrowed to what was their "limit" at the higher interest rate.

The business borrower borrows until stopped by the falling marginal productivity of capital, and the consumer borrows until stopped by the diminishing marginal utility of the goods that are purchased with borrowed funds. In either case, of course, the expected productivity or the expected utility governs, rather than the actual productivity or utility. Men borrow on the basis of their expectations, not on the basis of the facts. The facts are not known. They lie in the future at the time the borrowing occurs.

33 · 8. Supply of Loanable Funds. It is on the supply side that capital differs fundamentally from land. The quantity of land is approximately fixed, but the quantity of capital can be expanded or contracted. If the capitalized value of the income from capital exceeds the cost of producing it, then men will produce it and add to the supply. If, on the other hand, the capitalized value of the income falls short of the cost of production, men will not find it advantageous to add to the capital equipment of society, and the stock of capital will dwindle as old capital wears out. The supply of loanable funds is, of course, a schedule of the quantities that lenders would be willing to lend at various rates of interest.

Loanable funds are supplied in large part by commercial banks, by other financial institutions, and by the government; but in the last analysis this supply depends ultimately on the savings of the people. If we offer a higher rate of interest, some people will be willing to save who were not formerly willing to do so. For example, if Mrs. Turner personally discounts the future at a rate of 5 per cent (so that in her estimation 105 of next year's dollars are exactly equal to 100 of this year's dollars), she will not be willing to forego present consumption if she receives but 3 or 4 per cent as a reward; if the interest rate rises to the point where she can get, say, 6 per cent, she will start to save. She can now get 106 of next year's dollars by sacrificing 100 of this year's, and this looks to her like a good bargain.

When we turn to funds loaned by commercial banks, the situation appears somewhat different, at least on the surface. Some people speak of these banks as "creating" loan funds, and we may well ask ourselves whether or not there is any element of saving involved, or any cost of the capital that is loaned. Many people are misled on this point by a careless use of words. Banks do not create credit in any ordinary sense of the word at all. Businessmen who already have credit go to the bank and use it. Some businessmen have credit

and some do not, of course. When Henry Brown goes to the bank, the bank investigates his credit and, if his credit is good, recognizes the fact and gives him cash (or credit of its own) in return for his personal credit. The bank has not created credit, but has bought or exchanged credit. Credit is the ability to get valuable things in the present in return for a promise to pay a future equivalent. The bank did not create this ability in Brown. He had it already, and the bank recognized the fact. Moreover, even these bank loans depend on saving. If the bank lends money to businessmen who use it for productive purposes (for example, for the purchase of machinery), the businessmen are thereby enabled to attract productive factors away from their

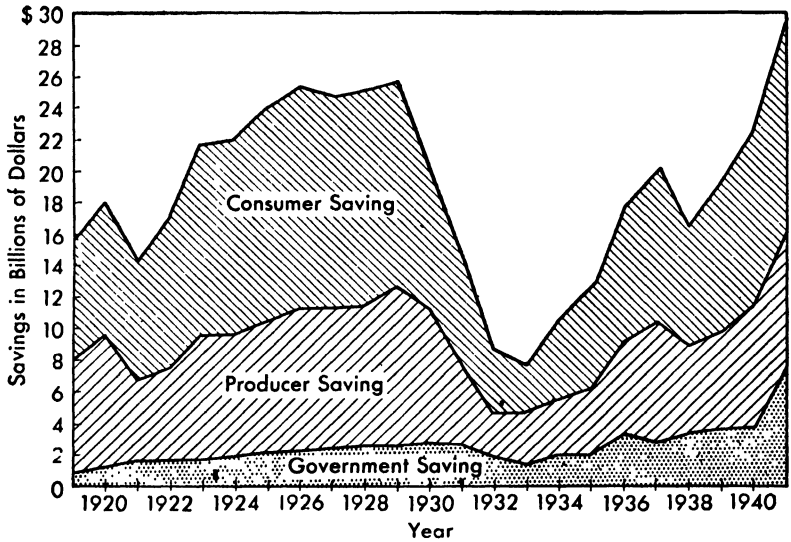


FIG. 33 • 1.—Sources of savings in the United States, 1919-1941, in billions of dollars. (*Federal Reserve Bulletin, September,* 1939, *February,* 1940, *and April,* 1942.)

other uses. People and equipment which would otherwise have been employed in making consumption goods are attracted into the production of machinery, and the general consuming public must, for the time being, get along with less to consume than would otherwise have been the case. The general public has saved to bring the capital into existence.

33 • 9. Volume of Saving. The volume of saving varies tremendously from year to year, being influenced in large measure by the swings of the business cycle. In times of depression people reduce their consumption somewhat, but they still have to eat and, when their incomes drop, they reduce their saving sharply or eliminate it altogether. On the other hand, when incomes rise in prosperous times, men's consumption habits change but slowly, and the added income is to a considerable extent saved. This leads

to very great fluctuations in the volume of saving, as shown in Fig. 33 • 1, which depicts the amounts of saving by various agencies in the United States in the years 1919-1941. These were years of peace, but they contained both prosperity and depression, and both recovery from one war and preparation for another. It will be noted that, at the peak, savings ran to better than 25 billion dollars a year, while in the depth of depression they fell to less than a third as much.

Taken over the period as a whole, we see from Fig. 33 • 1 that the volume of consumer savings is ordinarily somewhat larger than the volume of savings by industry and that savings by governmental agencies are very small as compared with either of them. The relative importance of consumer savings should be kept in mind in reading the next section.

If the figures were available for a continuation of Fig. 33 • 1, we would find that consumer savings increased tremendously during the war. Between 1941, where the data of the chart stop, and 1944, the volume of consumer saving almost trebled, while producer savings remained almost static.[1] The reason for this change is easy to understand. During the war, consumers' incomes rose to unusually high levels, but consumer goods of all sorts, especially consumer durable goods, were scarce and there was little on which to spend the high earnings. As a consequence, many families saved to buy goods after the war. A study of consumer savings made during the first quarter of 1946 showed[2] that personal holdings of such liquid assets as currency, bank deposits, and government bonds had increased from about 90 billion dollars in 1939 to some 130 billion at the end of 1945. It was estimated that 11 per cent of all families were expecting to buy new automobiles out of these savings, at an average price of about $1100 per car, making a total expenditure of between 4 and 5.5 billion dollars. Twenty-eight per cent of all families expected to buy other durable consumer goods at an average expenditure of $320, making a total expenditure in these lines of between 3.2 and 4 billion dollars. And 7 per cent of all families expected to buy or build houses out of their savings at an average expenditure of $5020, amounting to a total expenditure of between 13 and 15.5 billion dollars.

The continued savings of many years make up the productive wealth of the nation. Although estimates of the amount of this wealth are always difficult to make, and although they are especially untrustworthy when made for periods in the distant past, nevertheless the data of Fig. 33 • 2 are suggestive of the general trend in capital accumulation. This figure shows the estimated amount of wealth per capita in the United States at various times

[1] The author finds no data exactly comparable with those of Fig. 33 • 1, but data in the *Federal Reserve Bulletin* for July, 1946, p. 806, indicate that consumer saving probably increased about 2¾ times between 1941 and 1944, while corporate savings increased about 1⅓ times.

[2] See *Federal Reserve Bulletin*, June, 1946, pp. 574*ff*.

since the Revolutionary War, corrected for changes in the value of the dollar. The figures include the value of land and hence are not directly comparable with our other data. They show, however, a remarkable increase in the volume of resources per capita with which our people can work.

33 • 10. Saving and the Inequality of Incomes. In Sec. 4 • 5, the question was raised as to whether or not inequality of incomes is desirable. We should now point out that there is a close connection between the equality or lack of equality in the distribution of income and the amount of saving that takes place in a society. The typical family income is pitifully small, and the

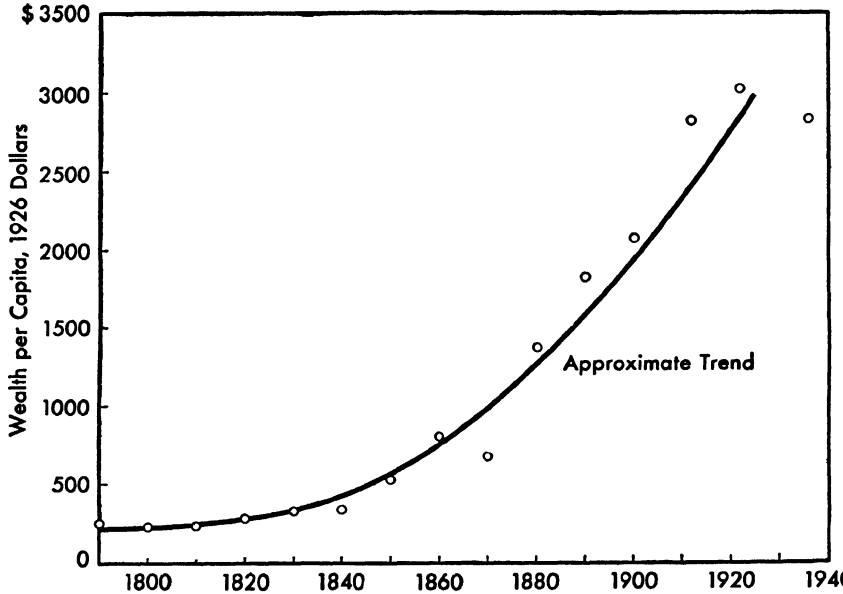


FIG. 33 • 2.—Wealth per capita in 1926 dollars, United States, 1790-1936. Estimates for various years, and approximate trend. (*Derived from data on pp.* 60 *and* 66, *"Studies in Enterprise and Social Progress," National Industrial Conference Board, New York,* 1939.)

typical family has trouble making both ends meet without laying aside any considerable amount of savings. On the other hand, the few families toward the upper end of the income scale receive incomes so large that they would have serious difficulty in spending them all, even if they tried. They save whether they try to or not. A very large proportion of all saving takes place at the upper income levels; and, if none of these large incomes existed, it is probable that the amount of voluntary saving would be greatly reduced. This is made very clear by the figures of Table 33 • 1, which show the percentages of spending units (families and single persons) in various income classes in 1945, the percentage of the total liquid assets held by each group, and the percentage of the total saving done in 1945 by each group. We note that the

fifth of the families with incomes below $1000 held but 7 per cent of the liquid assets and actually had negative saving. This means that they not only

Table 33·1. Percentage Distribution of Spending Units, Liquid Asset Holdings, and Net Saving by Income Classes, United States, 1945 and 1946*

Income class	Spending units	Holdings of liquid assets in early 1946	Net saving during 1945
Under $1000....	20	7	−1
$1000–$1999....	27	14	11
2000– 2999....	23	17	14
3000– 3999....	15	16	24
4000– 4999....	7	10	12
5000– 7499....	5	13	16
7500 and over..	3	23	24
Totals.....	100	100	100

* *Federal Reserve Bulletin*, July, 1946, p. 717.

added nothing to their savings, but reduced them. At the other extreme, the 3 per cent of the spending units with incomes of $7500 and over held 23 per cent of the liquid assets and did 24 per cent of the year's saving.

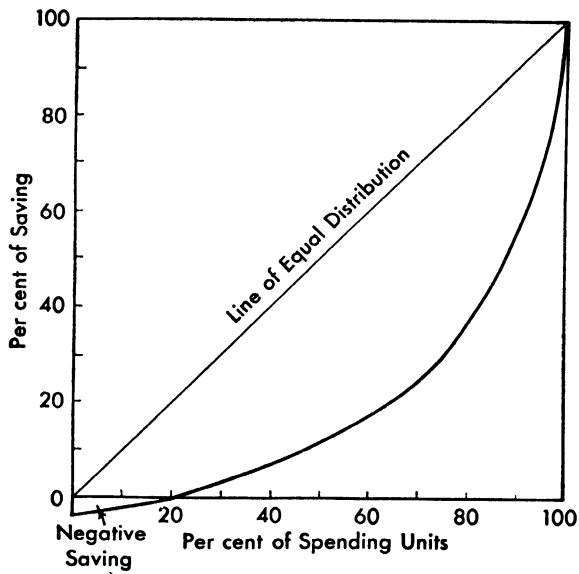


FIG. 33·3.—Lorenz chart of the distribution of saving in the United States in 1945. Compare with the distribution of incomes shown in Fig. 4·3 on page 45. (*Data of Table* 33·2.)

The situation is depicted graphically in Fig. 33·3, which is in the form of a Lorenz graph, explained in Sec. 4·3. The student will note by comparison

with Fig. 4 • 3 that saving is more unequal than the distribution of incomes.

Saving is advantageous to a society in that it gives rise to capital, and capital, being productive, increases the product available for consumption. As has been pointed out elsewhere, if the capital equipment of our society were to disappear, most of us would starve to death very soon. To be sure, saving is immediately productive to those who do the saving; but, as capital accumulates, the marginal productivity of capital declines and the productivity of labor increases. Ultimately and indirectly everyone gains by the saving. This desirable saving can be brought about in either of two ways: we can pay people interest to encourage them to save voluntarily, or we can force people to save. The latter policy of involuntary saving is carried out in

Table 33 • 2. Estimated Changes in Expenditures and Savings Which Would Have Resulted from Equalizing Incomes in the United States, 1935–1936*

	Disposal of consumer income, millions of dollars		Percentages	
	With existing unequal distribution	With equal distribution	With existing unequal distribution	With equal distribution
Expenditure on consumption...	$50,214	$55,756	84.7	94.1
Taxes and gifts.....	3,067	2,404	5.2	4.0
Net consumer savings.....	5,978	1,099	10.1	1.9
Total consumer income	$59,259	$59,259	100.0	100.0

* "The Structure of the American Economy," p 91, National Resources Committee, 1939.

part by taxation and in part by the setting aside of corporate surpluses. To the extent that we wish to rely on voluntary saving, we find that inequality in the distribution of incomes makes the saving more likely; and to the extent that we rely on involuntary saving through taxation, we find larger and larger proportions of the capital equipment of society falling into the hands of the government. Some people prefer public ownership of capital goods, while other people prefer private ownership. This issue we shall discuss later. But we are faced with the dilemma that private and voluntary accumulation of capital, which is desirable in a society based on private enterprise, is stimulated by inequality in the distribution of incomes.

An estimate of the actual influence of inequality of incomes on saving was made by the National Resources Committee for the year 1935-1936. The data are shown in Table 33 • 2. The average income in 1935-1936 was $1502 per

family, and the average family in the group with incomes from $1250 to $1500 saved 1.9 per cent of its income. Assuming that this rate of saving would have applied to all families if they received equal incomes of $1502, the total saving for the year would have been 1,099 million dollars instead of the actual 5,978 million which were saved with the existing inequalities of income. A similar very rough estimate based on the 1945 study from which the data of Table 33 • 1 were taken indicates that an equal distribution of income in 1945 might have reduced the total consumer saving to about 60 per cent of what actually occurred. This indicates a much smaller percentage decrease in saving than is suggested by Table 33 • 2, but at least a considerable part

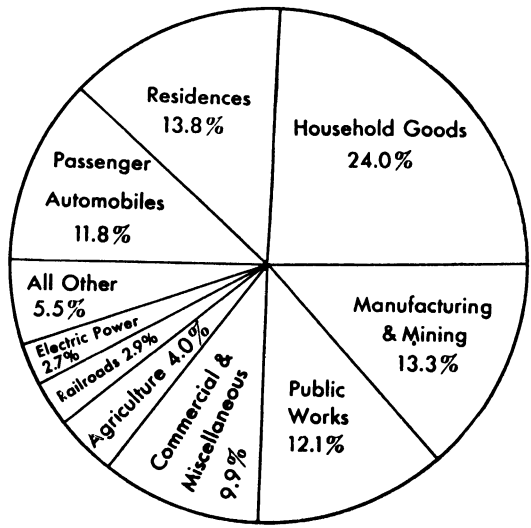


FIG. 33 • 4.—Uses made of national savings in the United States, 1919-1939. (*Federal Reserve Bulletin, September,* 1939.)

of that difference may be due to the fact that the 1945 average income was $2620 as compared with the average of $1502 in 1935-1936. The volume of saving is almost certainly more influenced by changes in the level of income than by changes in the interest rate.

33 • 11. Where Do Savings Go? The uses to which savings are put vary considerably from time to time. During the war, for example, durable consumer goods such as houses and automobiles were not to be had and investment in industrial war plants was greatly increased. But if we take the years 1919 to 1939 as being roughly typical of peacetime, including, as they do, periods both of prosperity and depression, the estimated expenditures for durable goods are shown in Fig. 33 • 4. Here we see that roughly a quarter of all savings go into durable household goods. If we add to this the expenditures

for residences and passenger automobiles, both of which are typically consumer rather than producer expenditures, we discover that almost half of the savings go into these three channels.

33·12. Determination of Interest Rate. As we have said before, the rate of interest is a price, and in the short run is determined, like other prices, by the forces of supply and demand. In the long run, the interest rate must be such as to make the capitalized value of durable goods equal the cost of producing them. When the interest rate is higher than this, it pays men to add to the stock of capital; when the interest rate is lower than this, it does not pay to replace capital as it wears out. But, in the short run, the rate of interest may be higher or lower than the cost.

In addition to comparing the actual rate of interest with the rate that would equate capitalized values and cost, it has been common in recent years for people to compare the actual interest rate with the marginal productivity of capital. If a businessman finds that the actual market rate of interest is so low that capital will produce more than it costs, he is tempted to borrow funds and expand his business. If, on the other hand, the market rate is higher than the productivity of the capital, it will not pay him to expand or even to replace capital as it wears out. The rate that would be determined by the actual productivity of the capital in use is called "natural rate of interest," as contrasted with the "market rate of interest" which is actually charged when one borrows capital. If the market rate exceeds the natural rate, business is static or declining; if the natural rate exceeds the market rate, business is expanding. Thus, or at least so the argument goes, the business cycle is determined by the relationship of these two rates. To be sure, it should be pointed out that exactly the same thing can be said of any other factor of production. Labor and land have natural rates dependent on productivity and, when the actual market rates exceed these natural rates, businessmen find it unprofitable to utilize the factors. When, on the other hand, the natural rate of any factor exceeds the market rate, businessmen tend to expand their use of the factor. The selection of any one factor, such as capital, as the one that determines the course of business is arbitrary and unrealistic.

The short-run determination of the interest rate is explained on the same basis as that of other short-run prices. For example, let us assume that Fig. 33·5 shows the supply of and the demand for capital at a given moment. The curve *DD* shows the demand, and we note that as the interest rate drops businessmen and consumers find it advantageous to borrow more capital. The curve *SS* shows the supply, and indicates that lenders are willing to offer more and more capital as the interest rate rises. The supply curve in the diagram is so drawn as to cross the base line. At its left-hand end it suggests that some capital would be saved even at negative rates of interest, even if the lender had to pay the borrower something to get him to take it. If we

remember that the pure rate of interest is a payment merely for the use of capital, without risk of loss,[1] we shall realize that, under some conditions, lenders would actually pay for the safekeeping of their capital. Such a situation would be unusual but not unheard of in practice. In times of great financial stress men not only get no interest on their capital but fear loss of the principal. If they can find an investment that appears to them safe, they are willing under the circumstances to pay for the safety as a man pays to put his valuables in a safety-deposit box. At one time during the depression of the thirties it actually occurred that investors lent short-term funds to the United States government at what amounted, after charges were paid, to a negative rate of interest. In our diagram we have assumed that men

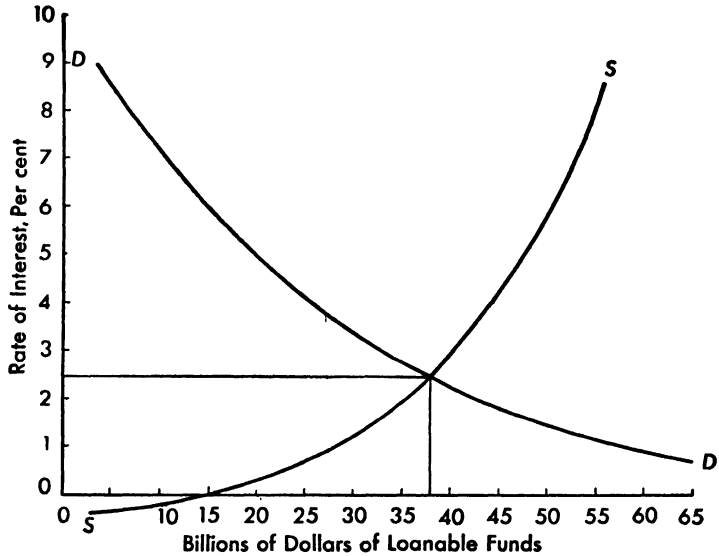


FIG. 33 • 5.—Determination of the short-run rate of interest.

would still be willing to invest 15 billion dollars or so even if the interest rate fell to zero if they could find a safe place for it; that they would save 5 billion dollars even if they had to make a small payment for the privilege, in order to get safety.

Of course, changes in the interest rate may change a man from the position of borrower to lender, or vice versa. A man who was anxious to borrow funds at a rate of 2 per cent might not borrow at all at 6 per cent; and, if the rate rose to 25 per cent, he might decide to lend what capital he did have in order to share in the high return. Interest does not differ from other prices in this regard, of course. The housewife who buys sugar at 8 cents a pound may stop buying at a price of 20 cents; if the price goes on up to $20 a pound,

[1] See Sec. 33 • 4.

she may become a seller and sell what stock she has in the pantry. Extreme price changes often change persons from buyers to sellers and back again.

Recalling our earlier analysis of prices, we realize that, under the conditions of Fig. 33 · 5, the interest rate would be set at about 2½ per cent, and that about 37.5 billion dollars of loanable funds would actually be borrowed (and lent). If the actual market rate of interest were higher, lenders would offer more than borrowers were willing to take at such a rate. If the rate of interest were lower, borrowers would be seeking more than lenders were willing to offer at the rate. Only at this rate of 2½ per cent will the market for loanable funds be "cleared."

Now let us ask what would be the effect of a change in time preference—in the attitude of men toward the future. In Fig. 33 · 6 let the two solid curves,

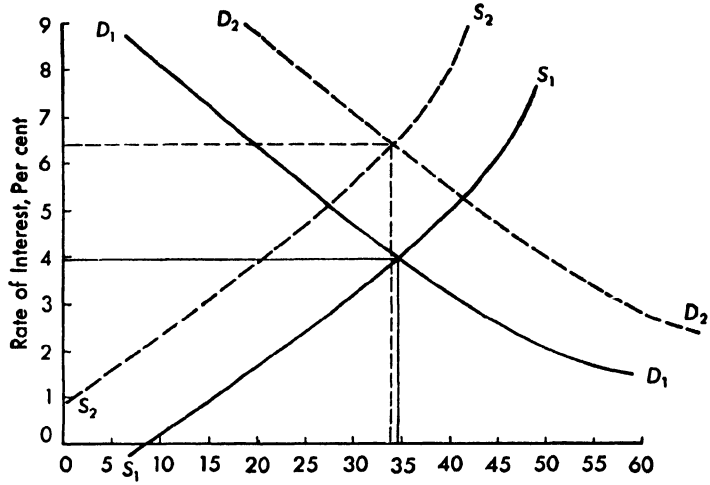


FIG. 33 · 6.—The effect of a change in time preference.

D_1 and S_1, represent the supply and demand situation on June 1. The June 1 interest rate is roughly 4 per cent and the volume of loanable funds borrowed is 35 billion dollars. By July 1, men have increased their time preference; in other words, they have come to set a higher value on present things as compared with future things—they have become more reckless of the future, more confident that the future will take care of itself, or more insistent on immediate gratification of wants. They discount the future at a higher rate. Under these circumstances, men will be more anxious to borrow than before, in order that they may have immediate satisfaction at the expense of future satisfaction. The demand for loans will increase, and we shall shift to the new broken curve D_2. If everyone is affected by this same change in attitude, lenders will be less willing to lend than before. They, too, will prefer to secure present enjoyment by using their savings immediately rather than by investing and

lending. Therefore the supply of loanable funds will be reduced, and we shall shift to the new broken supply curve S_2. On July 1, then, the two old solid curves have disappeared, and we have the two new broken curves. It will be noted that, under these conditions, the quantity of capital lent stays almost unchanged. In our particular diagram there is a very slight decrease but, had the slopes and the positions of the curves in the diagram been altered a little, this decrease might have disappeared or changed to a small increase. But the rate of interest, we note, has changed materially. The higher rate of time preference, or the higher rate of discount of the future, has brought about a higher interest rate.

The curves of Figs. 33 • 5 and 33 • 6 have been drawn more or less arbitrarily. Some students have tried to deduce on theoretical grounds the prob-

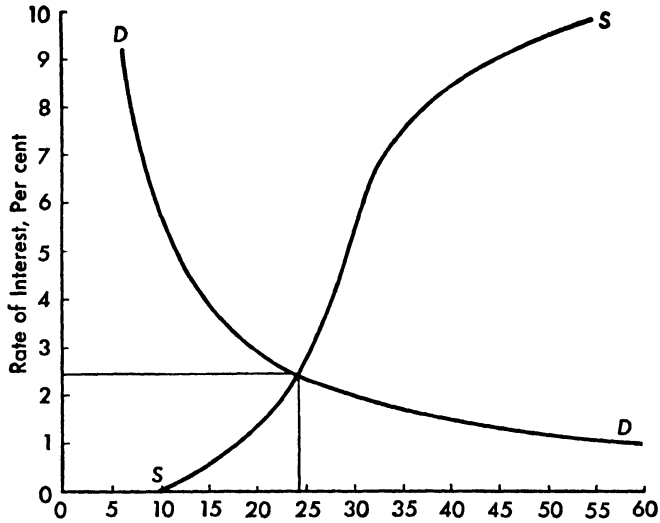


FIG. 33 • 7.—Suggested shapes of demand and supply curves for capital.

able shapes of these curves, and there is some dispute among them as to the facts. Frank H. Knight, one of the outstanding authorities in the field of the theory of capital, suggests, "It appears, for example, that in the United States the elasticity of demand for capital is around unity; that is, the total supply of capital, which includes the entire national wealth, would have to be doubled before the interest rate would be reduced to half its value . . ." [1] This would require that the demand curve take the particular shape shown in Fig. 33 • 7. Other authors have suggested that, at ordinary interest rates, the supply of capital is very inelastic, but that, when the interest rate is either unusually high or unusually low, the supply becomes elastic. If this is true, the actual

[1] KNIGHT, FRANK H., Interest, vol. 8, p. 136, "Encyclopedia of Social Sciences," The Macmillan Company, New York, 1932.

supply curve has the elongated S shape shown in Fig. 33 • 7. These conclusions are, however, highly speculative, and for purposes of exposition the conventionalized curves of Figs. 32 • 5 and 32 • 6 will serve.

33 • 13. Nature of Profits. There are more confusion and difference of usage in connection with the word "profits" than with any other word in the economist's vocabulary. Some economists think of it as a payment for the taking of risks, some think of it as a payment for the work of superintendence, some think of it as an unclassified residue which remains after all the other factors of production have been paid what is due them, some confine it to returns based on monopoly power, and some would argue that it either does not exist at all or at most is purely temporary in character. There is no way, of course, to ascertain which use is correct; but it is nevertheless important for us to understand enough of each other's meaning so that we can discuss usefully the problems involved. For that reason we shall define the word "profits" for our own use in understanding the problems of distribution.

Most people in our society are busy producing things. In this productive effort, use is made either of their personal efforts and exertions or of their property. In the course of any period of time, such as a year, great quantities of goods and services are produced—the distribuendum. These goods and services must all be allotted to someone and, as we noted in Chap. 30, the study of distribution is a study of the method of apportioning the distribuendum. Under ordinary circumstances one gets a claim to a share of goods by having participated himself in the productive process or by having furnished property that aided in production. As we have seen in this and the two preceding chapters, these agents of production share in the product, not at random, but in accordance with their marginal contribution. It is impossible to determine the total contribution of any factor and to say of the cotton crop, for example, "This is the part of the cotton crop which was produced by land, and this portion was produced by labor, etc." Without land, for example, there would be no product at all. But we do impute a product to each factor of production, and our imputation is on a marginal basis. We impute to each unit of any factor the amount of product that would be lost if that one unit (or, to be sure, any other one unit) of the factor were withdrawn from use. That is, we impute to each unit of the factor an amount of product equal to the marginal output; in actual practice, the payments made to the various factors are at least roughly equal to these marginal contributions.

Now we come to the point where we ask ourselves, "If we pay each unit of each productive factor just the amount of the marginal product, how much will be left over? Will such a system of payments distribute all the product, or will it call for even more product than is available—or will it just exactly come out even?" Any system of distribution must, in the long run, exactly distribute the product. We cannot distribute more than there is, nor in the

long run can we continue to produce quantities of goods that belong to no one at all. Any system of distribution must "come out even." We cannot divide a pie among six boys in such a way that each one gets a quarter; if we do divide it so that each one gets a tenth, there will be several pieces left which go to no one.

No one knows for certain whether our system would or would not exactly come out even in the long run, because life is dynamic and the situation is constantly changing. But the mathematicians tell us that a marginal system of distribution could work, and would work under certain specified conditions.[1] Be that as it may, in a world where things are constantly changing, the division is not always one that comes out exactly even. Sometimes there is a larger or smaller surplus left, and we call it a "profit." Sometimes there is not enough to go around, and we have a larger or smaller deficit which we call a "loss" or a "negative profit." Profits (or losses) arise under conditions when the distribution of the social dividend among the agents that produced it does not come out even, but leaves a surplus (or a deficit).

33 · 14. Profit Is Temporary under Competitive Conditions. We can illustrate the nature of profits most easily by referring to certain theorems discussed in Chaps. 14 and 17. Let us refer again to Fig. 14 · 3, discussed on page 257 and following. Here we see the familiar U-shaped average cost curve, *AC*, cut at its lowest point by the rising curve of marginal costs, *MC*. We noted in our previous discussion that if the market price were, for any reason, set at 80 cents, this firm would find it advantageous to produce 870 units of output. The average cost for such a volume of output is seen from the figure to be 72 cents per unit. Total income will be $696, total expense will be $626.40, and profit will be $69.60. When we pay the costs of $626.40, we have paid for the use of land, labor, and capital—for the use of all the factors of production. But we did not come out even in this case. We had a surplus of $69.60 which we call "profit." This, of course, the owner of the business is glad to appropriate even though it represents neither his own efforts nor the productive use of his property.

But, as noted in Sec. 17 · 4, such surpluses are temporary under competitive conditions. If other producers see that it is possible in this field to earn enough not only to pay off all the productive factors but also to leave a surplus, they will themselves enter the field in an effort to share the "spoils";

[1] The condition is evident in Euler's theorem, which tells us that if a homogeneous function in several variables of degree n is multiplied by n, the function will equal the sum of the products of each variable times the partial derivative of the function with respect to that variable. In economic nomenclature, these partial derivatives are the marginal productivities. The theorem tells us that a marginal system of distribution would work if the production functions were homogeneous; if, for example, doubling all the factors will double the product. This may seem at first contrary to the idea of diminishing productivity but, in the latter case, we do not double all the factors but hold at least one of them constant.

this will in turn increase the output and reduce the price below 80 cents. In fact, we discovered in Sec. 17 • 4 that, although the price may temporarily be above or below that point, in the long run under competition it must coincide with the very bottom point on the U-shaped average cost curve. In Fig. 14 • 3 this means that the long-run price must be approximately 68 cents. At this price the firm will produce approximately 775 units of output. The total income will be $527 and the total expenses will also be $527. There will be exactly enough income to pay off every unit of every productive factor—to recompense it for the work done—but there will be no surplus or deficit in the form of profit or loss.

33 • 15. Monopoly Profit May Be Permanent. The monopoly situation differs, however, in the fact that competitors cannot come in to share the profits and thus force the profits out of existence. There are no competitors to come in. This situation we also discussed in Chap. 17, and we repeat the bare outline of it here. Refer back to Fig. 17 • 7 on page 344. Here we have not only the two cost curves in their familiar relationship, but also the average revenue curve *AR* and the marginal revenue curve *MR*. The monopolist is free to choose whatever output he wishes, but he finds it most profitable to produce 575 units, since this is the point at which his marginal cost equals his marginal revenue. As was pointed out in Sec. 17 • 10, at this output he will make a profit of $143.75 after paying off all his costs. This time, since he is a monopolist, other producers cannot enter the field to take advantage of these profits. The monopolist can continue to enjoy them as long as he maintains his monopoly position. *Competitive profits are always unforeseeable, fortuitous, temporary phenomena, but monopoly profits may be permanent.* Losses are of course temporary in either case, since monopolists are no more immune to bankruptcy than anyone else.

33 • 16. Profit Is an Arbitrary Accounting Concept. This idea of profits is apt to strike the student as unrealistic. He knows businessmen who make profits every year without any element of monopoly advantage. How does the owner of the shoe store on Main Street, the insurance salesman on Pleasant Street, or the operator of the movie theater on Amity Street live if he does not make a profit? Must he work all year for nothing, getting barely enough to cover his costs, with nothing whatever left over for himself? These men would tell you that they make a profit every year. But they are using the term in another sense, and one that is admittedly just as "correct."

Of course these men get incomes, but the greater part of their incomes can be classified under other headings that we have used. The insurance salesman on Pleasant Street is spending most of his time and energy trying to sell policies and to assist with the adjustment of claims. Whatever pay he gets for these activities is not profit in the technical economic sense, but wages. The owner of the shoe store and the operator of the movie theater presumably

work in their own places of business and receive wages also. They may, of course, own their establishments outright instead of renting their quarters, and in this case part of their income is really interest on their invested capital. Some of it perhaps is rent on their building sites. The economist is not foolish enough to argue that these men get no incomes. He states merely that their incomes are, in the ordinary course of events, payments for their own effort (wages) or for the use of their property (rent and interest) rather than profits and that, if they are paid reasonable sums for their own effort and property, there will be nothing left to classify as profit in the economic sense.

The persistent student is still likely to ask, however, "But suppose there is actually something left over year after year. Aren't you going to classify it as profit then?" The economist's answer is that in such a case one thing or the other must be true. Either the man has monopoly advantages and his profit is a monopoly profit, or else his accountant has done a poor job. In the absence of monopoly, if a man makes enough to pay all the factors of production year after year and still has a surplus, it shows merely that he has not valued his productive factors properly. If he is so much more skillful a producer than his competitors that he can outdo them every year, he is worth a good deal himself, and perhaps he should be charging his own efforts at a higher rate. His "wages of management" have not been valued highly enough by the accountant. Possibly he has some other advantage which his competitors cannot get which enables him to do better than they do year after year—a patent, a unique source of materials, or something of that kind. If this is the cause of it, the accountant should evaluate the patent or the source of materials at its proper value. If he puts them in for what they are worth, his books will balance and there will be no "profit" left except the occasional one that arises from unforeseen circumstances. Even the monopoly profit in this sense is not a profit. If the monopolist sets up his accounts correctly, he will put such a value on his exclusive franchise, his patent, or his other source of monopoly power that his accounts show the payments assigned to that source rather than as an unsassigned and chance "profit." The problem turns out to be one of accounting, and completely arbitrary in its character. The economist insists merely that the payments be made to someone and for something—that they be not thought of as random payments without cause. In the long run, men do not in fact continue to pay for anything without cause. In the long run, payments are made in return for goods or services.

Questions

1. How is interest like, and how unlike, rent?

2. In the primitive society referred to in Sec. 33 • 1 a man borrows a bushel of wheat and a year later pays back a bushel of wheat and a bushel of barley. How much

interest was paid? What was the rate of interest? Comment on the difficulty involved in answering the latter question.

3. If, in question 2, wheat is worth $1 a bushel and barley is worth 65 cents a bushel, what was the interest rate? Why are interest transactions almost always computed in monetary form?

4. A man borrows $75 and a year later repays $85. What was the rate of interest?

5. Students who are held responsible for the matter of the Mathematical Appendix should answer the following:

   "In the South in the United States Negroes are said to borrow $5 at the beginning of the week and repay $7 at the end of it, an interest rate of 40 per cent per week."[1]

   What is the equivalent annual rate of interest, assuming that it is compounded weekly?

6. In Sec. 33·12, in connection with Fig. 33·5, the statement is made that some capital would be saved even at a negative rate of interest. Under such conditions, why do men actually pay interest for money?

7. "The fact that men save for the future even when not paid interest shows that they do not have time preference." Comment.

8. On a given transaction which would be larger: the contract interest or the pure interest? Why?

9. What is the relationship between changes in the general price level and the contract rate of interest?

10. A man borrows money for a year at 10 per cent contract interest. During the course of the year the price index number falls from 89 to 74. How does this affect the transaction? Give a numerical answer.

11. In time of war, prices usually rise sharply. Would the expectation of a price rise make businessmen more or less willing to borrow at any given interest rate? Why?

12. What is meant by the call rate of interest? Why do people prefer to borrow money on a call basis for certain kinds of transactions? For what kinds of transactions is call money actually used?

13. A businessman borrows from his bank on a 90-day note. The face value of the note is $4000, and it is discounted at 5 per cent bank discount. On this basis, the businessman receives an actual deposit of $3950, and 90 days later he repays $4000. What is the equivalent annual rate of interest if we assume that interest is compounded every 90 days? (To be answered by students who are responsible for the material of the Mathematical Appendix.)

14. A $1000 5 per cent 15-year bond is selling for $950. What is the effective yield to maturity? Give an approximate answer.

15. Explain the method of making your approximation in question 14.

[1] KNIGHT, *op. cit.*, p. 139.

16. Explain how changes in the rate of interest affect the value of durable property.

17. Suppose that the rate of interest rises from 3 to 4 per cent. What will be the effect on the value of durable property? Give a numerical answer.

18. At a time when inflation is threatened would you prefer to invest in stocks or bonds? Why?

19. Which has the greater time preference, the wastrel or the miser?

20. "Men borrow on the basis of their expectations." What does this statement mean?

21. Describe the fluctuations in volume of saving that accompany the swing of the business cycle.

22. Using Fig. 33 • 2 as a basis, describe in a paragraph the growth of real wealth per capita in the United States.

23. How is saving affected by inequality in the distribution of incomes?

24. In Table 33 • 2 an estimate is made of the saving that would have taken place in the United States in 1935-1936 if the actual total income had been distributed equally among the income earners instead of unequally. The estimate is based on the assumption that, if the income had been equally distributed, everyone would have saved as the average family did save which received an average income. Do you think that this assumption is reasonable? Why or why not?

25. How does the volume of consumers' capital compare with the volume of producers' capital?

26. What is the natural rate of interest? What is its relationship to the market rate?

27. Explain why it is that the rate of interest rises when men's time preference rises.

28. Suppose that, when the interest rate is 3.5 per cent, borrowers actually borrow 24 billion dollars. If there really is unitary elasticity of demand for loanable funds, how much will borrowers borrow if the interest rate rises to 4 per cent? To 5 per cent?

29. Why do we use a method of imputation to find the productivity of a factor instead of actually measuring what it produced? Of the total annual output of automobiles in the United States, how many would have been produced if we had used no land (natural resources)? How many, if we had used no labor (human effort)? How many, if we had used no capital? What is the difficulty in appraising the productivity of a factor as a whole instead of on a marginal basis?

30. "Profits and losses are not different phenomena, but are just as much the same as are temperatures above and below zero, or altitudes above and below sea level." Do you agree?

31. Suppose that a farmer tells you that he made a profit last year of $2500. You discover that he measures profit by subtracting from his total income the total amount of his payments to other people. He thinks of it as payments to himself as distinct from payments to others. Can you suggest the other distributive shares into which this $2500 might be divided by the economist if he refused to call them profits?

PART SEVEN. The Government and The Economic System

34. PUBLIC EXPENDITURES

34 · 1. Government as a Spending Agency. There has long been, and there continues to be, decided difference of opinion as to the proper functions of government. On the one hand, we have those who argue that every human activity should be carried on by those who are best qualified to do it; on the other, we have those who urge that men should be left free even if this means freedom to make their own mistakes. There is also difference of opinion on almost every issue as to who are the people best qualified to carry out the program involved. On the one hand, there are those who point to the admitted ignorance of the ordinary citizen on most technical matters of modern life; on the other, are those who assert that the bureaucracy of practical politics leads to control by those who are no better informed than the average citizen and who tend to make decisions for selfish and partisan benefit rather than for the benefit of the general public. The passage of time seems to bring us no nearer to unanimity on these matters.

Regardless of the opinion that may be held on any particular project, the fact remains that governments throughout the world have been taking over an increasing number of functions in recent years. This has obviously been true in the cases of the authoritarian governments of fascism and communism, but it also true in a somewhat lesser degree of the democratic countries. In the present chapter we shall confine our attention to the situation in the United States, which we find to be no exception to the general rule. Figure 34 · 1 shows the percentages of the total national expenditure spent in recent years by various government agencies, Federal, state, and local. Naturally, this percentage rose somewhat during the war, until in 1943 and 1944 half of the entire national expenditure was being made by government; but, even in the prewar years, the proportion stood at roughly one-fifth to one-sixth. To this extent, even in one of the leading "capitalist" countries, the economy was, so to speak, "socialized."[1] The government had become the most important

[1] We shall not stop to define here the words in quotation marks, but shall comment further on them in Chap. 36.

single "industry" of the country, far more important in volume of business transacted than agriculture, mining, retail or wholesale trade, and almost as important as all the manufacturing industries thrown together.

There is no way to tell how long the recent trend toward enlarged government functions will continue nor how far it will go. History seems to suggest, as a matter of fact, that the movement may be a cyclical one. Prior to the French Revolution, the governments of Europe exercised far greater powers, particularly in the economic sphere, than they have ever exercised since. Then came the period of individualism and *laissez faire,* during which the functions of government were reduced to the defense of the state from external attack, the preservation of internal order, and the adjudication of

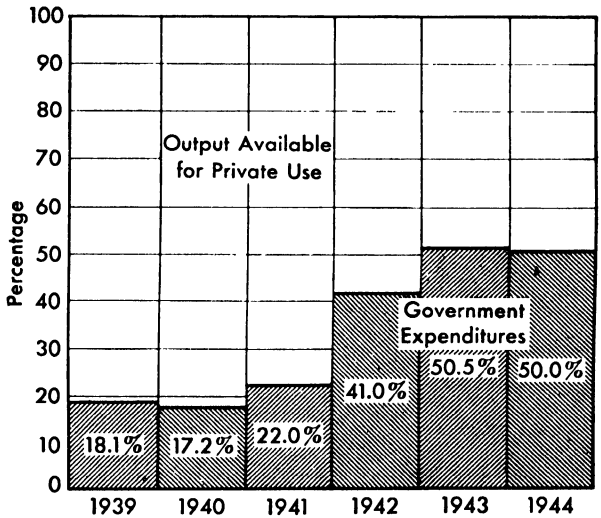


FIG. 34 • 1—Proportion of total national expenditures spent by Federal, state, and local governments, 1939-1944. (*Stat. Abs.* 1944-1945, *Table* 442, *p.* 406.)

individual rights. The doctrine that "that government is best which governs least" was widely preached. Since about 1870 or 1880, the present trend back toward increased state functions and responsibilities has been under way, and during the depression of the thirties and the New Deal in American political life the trend was particularly pronounced.

Government services, like other services, are costly. Governments may, to be sure, be costly without rendering a fair return in services, as a result of either inefficiency or graft. But, generally, the reason for such increased expense can be found in increased services by the government—by the assumption of new functions and the performance of new services.

34 • 2. Relationship of Government Income and Expenditure. Men commonly remark that the government is like an individual or a business—that its

economic problems are on a larger scale but are nevertheless fundamentally the same kind of problems. In some ways this is true, but the analogy is a poor one and must not be pressed too far. In the first place, governments are not operated for the purpose of making a profit. Their efficiency must be measured in other terms, and a government that succeeded in making a large profit might well be doing its job very poorly. In the second place, the basic nature of the economic problem of government is in some respects just the reverse of that faced by individuals and business firms.

A family, in meeting its economic problems, finds that its income is limited. Usually, as noted in Chap. 4, the income is altogether too low to enable the family to satisfy all its wants, and the necessity arises of fitting the expenditures to the income—of going without many things that the family would like in order to keep expenditures within the limits of the family's budget. But a government is in the envious position of being able to go at things the other way around. It decides first what it wants to spend and then adjusts its income to cover the desired expenditures. The individual adapts expenditures to a fixed income; the government adapts its income to cover desired expenditures. The typical governmental unit (local, state, or national) adopts a budget listing the expenses that it wishes to incur, and then alters the tax structure as may be necessary to provide the needed income.

For this reason, although we start to study consumer problems with a study of consumer incomes as the limiting factor, we begin the study of the economic problems of government with a study of government expenditures. We shall then take up in the following chapter the related and dependent problem of government income. Since, in the United States, we have several somewhat distinct levels of government and since they perform services that differ at least in emphasis if not in kind, it will be helpful if we consider separately the expenditures of the Federal, the state, and the local governments.

34·3. Purposes of Expenditure: Federal Government. For many years, especially since the outbreak of the first World War, the major expense of the Federal government has been the cost of past, present, and future wars. Past wars contribute a continuing and inescapable burden of expense in the form of care of the disabled, other types of veterans' benefits, interest on the public debt, etc. Expenditures for preparedness, battleships, military training, and weapons of all sorts, add to this burden as we look ahead to the possibility of future wars. During the actual prosecution of a war, the expenditures for its purposes dwarf all the rest of the budget. Armament has been a weary and an unremitting load, and for the past generation a rapidly increasing one.

Figure 34·2 shows the relative importance of Federal expenditures for various purposes in the years 1926, 1939, and 1944. These years are chosen as roughly typical of peacetime prosperity, peacetime depression, and wartime. It will be noted that the Federal budget is by no means fixed either in total

amount or in the direction of expenditure. In 1926, the major items were payments for past and future wars, with these items accounting for perhaps three-quarters of the total outlay. In 1939, the government was cast in a new role, that of dispenser of relief. Relief had theretofore been almost exclusively a state and local concern but, in 1939, the great depression had been under way

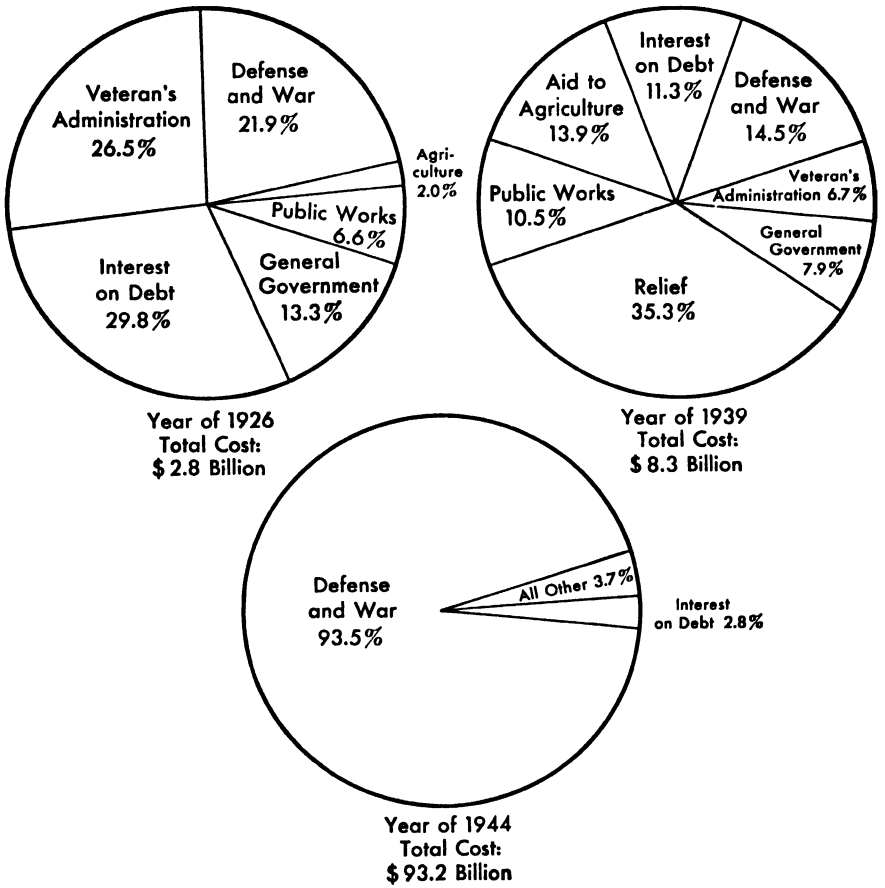


FIG. 34 • 2.—Relative importance of federal expenditures for various purposes in 1926, 1939, and 1944. Based on the cost of government exclusive of loans. (*Based on data in Paul W. Ellis, "The World's Biggest Business," National Industrial Conference Board, New York,* 1944, *p.* 25.)

for a decade, and the responsibility for poor relief had been taken over very largely by the national government. When we recall that a very large proportion of the expenditures for public works and for aid to agriculture were actually relief payments, we see that these items accounted for perhaps three-fifths of the 1939 outlays. In 1944, the war had supplanted everything else.

Over 95 per cent of the budget was directed toward either the current war or payments of interest on old war debts, and even the latter had dropped to a minor position. It must be remembered in interpreting these facts that, in

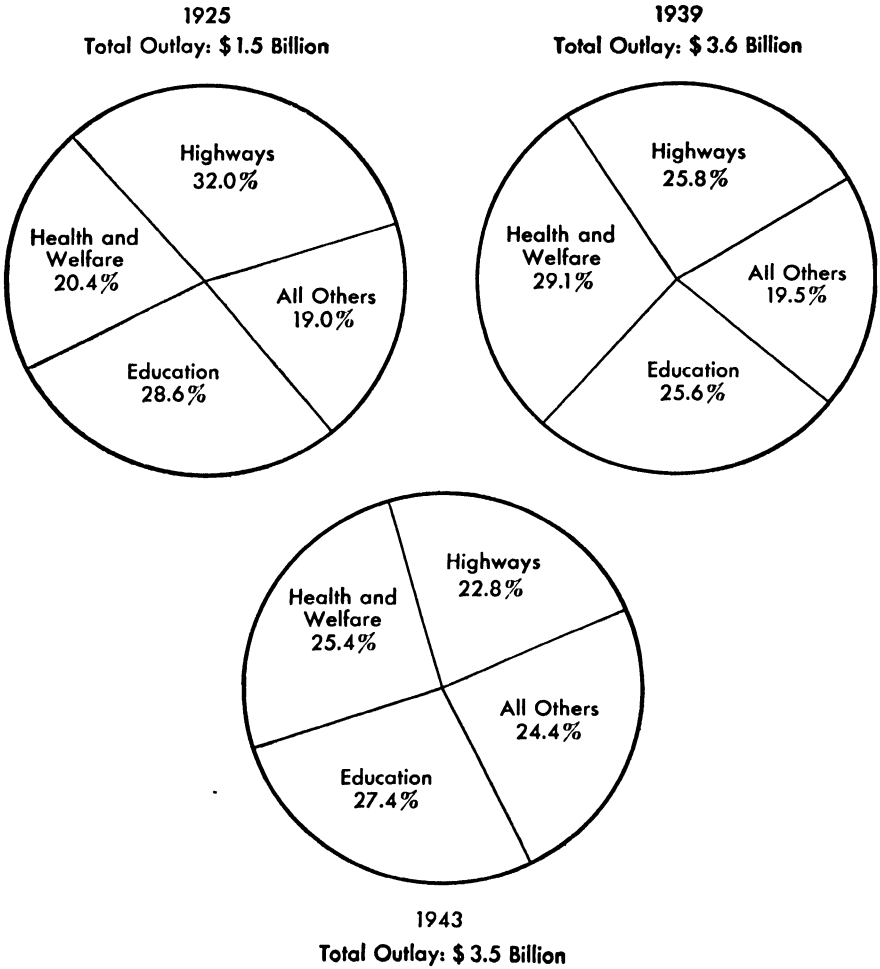


FIG. 34 • 3.—Relative importance of state expenditures for various purposes in 1925, 1939, and 1943. Based on expenditures from state funds only. Health and welfare includes health, hospitals, public welfare, and correction. This table does not include funds spent by states but furnished by the Federal government. It covers only expenditures from state funds. (*Paul W. Ellis, The World's Biggest Business, National Industrial Conference Board, New York,* 1944, *p.* 65.)

1944, not only had the relative importance of war expenditures increased, but also the total budget had swollen to roughly thirty times normal size. As one looks ahead, keeping in mind the staggering size of the present war debt which

we shall mention later, it seems certain that war expenditures will be the major item of Federal expenditure for a long time to come.

34·4. Purposes of Expenditure: The States. The purposes of state expenditures vary much less from time to time than do those of the Federal government. Year in and year out, in wartime and peacetime, in prosperity and depression, the three major purposes of state expenditure are education, highways, and welfare. Figure 34·3 shows the situation graphically for the years 1925, 1939, and 1943. Next most important among the expenditures, although not shown in the figure, were expenditures for "general control" including the usual legislative, executive, and judicial functions of government. This last item of expenditure ran in the neighborhood of 5 per cent of the total outlay.

The expenditures shown in Fig. 34·3 are not the total expenditures of the states. In addition to those listed, the states spent large amounts of funds which had been raised by the Federal government and which were disbursed by the states. These items were especially important in the fields of highways, education, and public welfare. Since these Federal expenditures have already been covered in Fig. 33·2 and do not represent an additional drain on the taxpayer, and since in addition they do not have to be covered by state tax systems, it has seemed wise to exclude them from the strictly state expenditures.

Comparison of the three parts of Fig. 34·3 will indicate that expenditures for welfare were, as is to be expected, proportionately more important during the depression than they had been before. The decrease in the relative importance of highway expenditures in 1943 was largely due to the scarcity of man power and materials during the war. Yet the striking thing about the three parts of Fig. 34·3 is not the differences, but the similarity—the stability of the distribution of major state expenditures.

34·5. Purposes of Expenditure: Local Government. When one has studied the expenditures of the Federal and the state governments, he has by no means completed his study of government expenditures. There must still be included the expenditures of over 155,000 such units of local government as counties, townships, fire districts, school districts, boroughs, cities, and metropolitan districts. Nor are these expenditures minor matters to be lightly dismissed. In ordinary times of peace, these local units have usually spent as much as the Federal and state governments combined; even in time of war, when the Federal budget rises almost beyond calculation, the local units always spend far more than the states.

Unfortunately, there are no complete or reliable figures covering these many small and overlapping units. At best we must speak of them in general terms. To a considerable extent, these local expenditures are for such things as schools, fire and police protection, highways, and public welfare. In cities of medium size, roughly one-fifth of the total expenditure goes for public

safety (police protection, etc.) and another fifth for education, with highway expenditures accounting for perhaps a tenth.[1] County governments, which take perhaps 15 to 20 per cent of the expenditures of local governments, spend primarily for public welfare, highways, and general control. The cities contribute perhaps 35 to 50 per cent of the local expenditures, and the many miscellaneous units make up the balance.

34 • 6. Volume of Government Expenditures. The amount of spending by state and local governments is fairly uniform from year to year, but the tremendous swings of Federal wartime spending have such an influence that

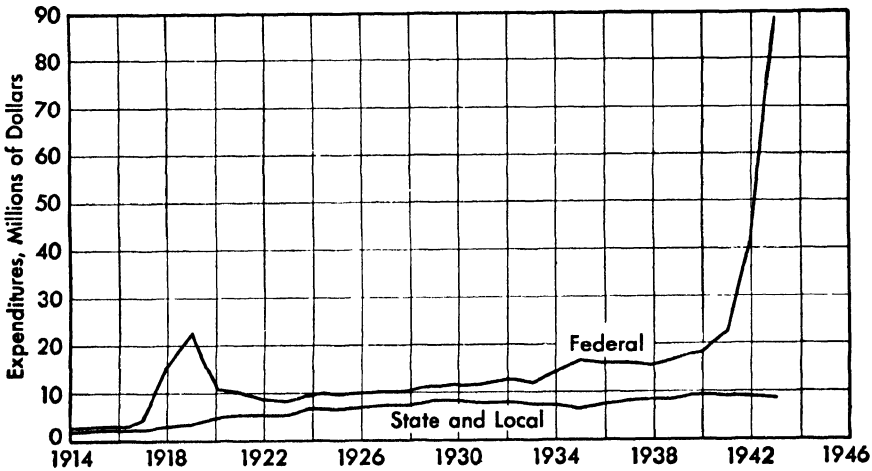


FIG. 34 • 4.—Expenditures of the Federal and the state and local governments, 1914-1943. This figure is based on the data of Table 34 • 1. The space between the lower line and the base line shows the amount of expenditures by state and local governments, while the expenditures of the Federal government are shown by the space between the two lines on the chart. The height of the upper line shows the total expenditures of all levels of government rather than the amount for the Federal government alone.

the total varies greatly. Table 34 • 1 shows the figures for total expenditures by Federal, state, and local governments since just before the first World War, and these data are shown graphically in Fig. 34 • 4. The student will note two things from these data: (1) In ordinary times, the units of local government, each one seemingly unimportant in itself, have had greater aggregate expenditures than either state or Federal governments, and much of the time greater than both of them together. Before the first World War the local governments accounted for about two-thirds of all government expenditures, after the war and before the depression for about half, during the depression and before

[1] These figures, and much of the other statistical material of this chapter, are based on Paul W. Ellis's very useful study, "The World's Biggest Business," *passim*, National Industrial Conference Board, New York, 1944.

the second war about a third, and during the war periods for a relatively small part because of the overwhelming influence of Federal war spending. (2) There

Table 34·1. Expenditures of Federal, State, and Local Governments in Millions of Dollars, 1914–1945*

Fiscal year	Federal	State	Local	Total
1914	$709	$421	$1,930	$3,060
1915	727	483	2,073	3,283
1916	697	497	2,087	3,281
1917	1,925	505	2,281	4,711
1918	12,648	553	2,460	15,661
1919	18,459	625	3,041	22,125
1920	6,308	789	4,035	11,132
1921	4,989	952	4,429	10,370
1922	3,213	1,179	4,464	8,856
1923	3,054	1,195	4,291	8,540
1924	2,810	1,388	5,485	9,683
1925	2,801	1,479	5,196	9,476
1926	2,779	1,487	5,666	9,932
1927	2,738	1,591	5,923	10,252
1928	2,799	1,753	6,067	10,619
1929	2,958	1,928	6,469	11,355
1930	3,152	2,152	6,071	11,375
1931	3,560	2,282	5,550	11,392
1932	4,640	2,245	5,824	12,709
1933	4,611	2,220	5,222	12,053
1934	7,362	2,195	5,012	14,569
1935	9,738	2,396	4,583	16,717
1936	9,098	2,614	4,651	16,363
1937	7,754	3,094	5,066	15,914
1938	6,977	3,391	5,233	15,601
1939	8,361	3,553	5,046	16,960
1940	8,824	3,612	5,628†	18,064
1941	13,862	3,585	5,553†	23,000†
1942	33,980	3,644	5,213†	42,837†
1943	79,604	3,503	4,828†	87,935‡
1944	95,129			
1945	97,309‡			

* ELLIS, PAUL W., "The World's Biggest Business," p. 3, National Industrial Conference Board, New York, 1944.
† Preliminary. ‡ Estimated.

has been a steady increase in the volume of government spending. Local governments reached a peak in 1929 and fell off during the ensuing depres-

sion; but state and especially Federal expenditures continued to rise, as did the grand total. The wartime peaks, as has been said before, broke the evenness of the trend but did not halt the general long-run tendency.

The fact that this tendency has run back much further than the period of the first World War is evident from an inspection of the data in Table 34 • 2, which show the Federal expenditures annually since 1792. Figures for state

Table 34 • 2. Expenditures of the Federal Government for the Fiscal Years 1792–1945*

Year	0	1	2	3	4	5	6	7	8	9
	Millions of dollars									
1790	. .	. .	$5 1	$4 5	$7 0	$7.5	$5.7	$6.1	$7.7	$9.7
1800	$10.8	$9 4	7.9	7.9	8.7	10.5	9.8	8.4	9.9	10 3
1810	8.2	8.1	20 3	31.7	34.7	32.7	30.6	21.8	19.8	21.5
1820	18.3	15.8	15.0	14.7	20.3	15.9	17.0	16.1	16.4	15 2
1830	15.1	15.2	17.3	23.0	18.6	17.6	30.9	37.2	33 9	26.9
1840	24.3	26.6	25.2	11.8	22.3	22.9	27.8	57.3	45.4	45 0
1850	39 5	47.7	44.2	48.2	58.0	59.7	69.6	67.8	74 2	69.1
1860	63.1	66 5	474 8	714.7	865.3	1297.6	520.8	357.5	377.3	322 9
1870	309 7	292 2	277.5	290.3	302.6	274.6	265 1	241.3	236.9	266 9
1880	267.6	260 7	258.0	265.4	244.1	260.2	242.5	267.9	267.9	299.3
1890	318.0	365.8	345.0	383.5	367.5	356.2	352.2	365.8	443.4	605 1
1900	520 9	524.6	485 2	517.0	583.7	567.3	570 2	579.1	659.2	693.7
1910	693.6	691.2	689 9	724.5	735.1	760.6	734.1			
	Billions of dollars									
1910	.....	.....	.....	. . .	.....	.....	.....	2.0	12.7	18.5
1920	6 4	5.1	3.4	3.3	3.0	3.1	3.1	3.0	3.1	3.3
1930	3.4	3.7	4.5	3.9	6.0	7.0	8.7	8.2	7.2	8 7
1940	9.0	12.7	32.4	78.2	93.7	100.4				

* "Annual Report of the Secretary of the Treasury on the State of the Finances for the Fiscal Year Ended June 30, 1945," pp. 445, 447, and 449, U S Government Printing Office, Washington, D. C., 1946. The figures are for expenditures, including debt retirement.

and local expenditures are not available at such early dates. The student will find it instructive to compare the date of Table 34 • 2 with those of Table 12 • 1 and to compute the per capita expenditures for those dates on which population data are available. Reference to Fig. 34 • 1 will indicate that these expenditures have now risen to the point where they take up in normal years roughly a fifth of the entire national income.[1]

[1] The figures of Tables 34 • 1 and 34 • 2 are not directly comparable, since they are computed on somewhat different bases. Yet it will be seen at once that the figures are of the same order of size and that they display the same general tendencies.

34 · 7. The Public Debt. Before we leave the question of government expenditures, it should be pointed out that an increasingly large item of expense has been the payment of interest on the public debt. This, of course, has been because of a tendency for public expenditure to exceed public revenues and for the debt itself to increase in size. The interest rates paid by public borrowers are, on the whole, rather low. In 1944, over half the outstanding debt of the United States government bore interest of from 2.0 to 2.99 per cent, and less than a quarter of it paid 3 per cent or over. The most common interest rate on state borrowing has been either 4 or 4¼ per cent, while cities have typically borrowed at about 4 to 4½ per cent.[1] Yet, despite these low rates, the volume of indebtedness has become so great that the interest burden is a major one. Reference to Fig. 34 · 2 will show that it is one of the principal items of Federal expenditure, but it looms large also with state and local governments.

The occurrence of two major wars in a generation and the unbelievable cost of fighting a modern war have raised the total public debt to unprecedented heights and have made it certain that the burden of interest and of debt repayment will continue to be a major expense, especially of the Federal government, for a generation or more. The general situation is shown in Table 34 · 3, which gives data on the size of the Federal, state, and local debts each year since just before the first World War. At that time, the debts of local governments were far greater than the debt of the national government, but the coming of the war in 1917 very quickly reversed the standing and, from that time on, the Federal government has been the largest debtor of the three. The combined state debts have never been large in comparison with either of the other two.

The burden of state and local debt has tended to increase slowly and without much interruption but, after the first World War, during the prosperity of the twenties, the Federal debt was cut by about a third. Then came the depression, with its load of Federal expenditure for relief and welfare, and the Federal debt started up again. The coming of the second World War sent it to levels which had not been dreamed of—not thought economically possible—before.

As with expenditures, the figures on Federal debt are more accurate and more complete than those for the other levels of government. We can get a reasonably clear picture of how the Federal debt was built from Table 34 · 4, which shows the amount of the Federal surplus or deficit in each fiscal year since 1792. A close study of this table will indicate that war has

[1] These figures, based on Ellis, *op. cit.*, are the rates quoted in the bonds or other evidences of indebtedness, and do not take into account any discount or premium at which the bonds were sold. They are not, therefore, entirely reliable indicators of the actual effective rate of interest paid by the governmental units.

always been the most spectacular contributor to Federal deficits, but that periods of depression have also ordinarily made their contribution. The

Table 34·3. Debt of Federal, State, and Local Governments in Millions of Dollars, 1914–1945*

Fiscal year	Federal	State	Local	Total
1914	$1,188	$477	$3,935	$5,600
1915	1,191	504	4,191	5,886
1916	1,225	559	4,456	6,240
1917	2,976	608	4,755	8,339
1918	12,244	625	5,066	17,935
1919	25,482	648	3,321	31,451
1920	24,299	774	6,126	31,199
1921	23,977	900	6,961	31,838
1922	22,963	1,026	7,832	31,821
1923	22,350	1,152	8,294	31,796
1924	21,251	1,408	9,660	32,319
1925	20,516	1,540	10,493	32,549
1926	19,643	1,645	11,495	32,783
1927	18,512	1,729	12,371	32,612
1928	17,604	1,851	13,236	32,691
1929	16,931	2,045	14,198	33,174
1930	16,185	2,182	14,718	33,085
1931	16,801	2,370	13,800	32,971
1932	19,487	2,611	13,982	36,080
1933	22,539	3,054	14,154	39,747
1934	27,734	3,238	14,173	45,145
1935	32,824	3,443	13,814	50,081
1936	38,497	3,363	13,701	55,561
1937	41,089	3,276	13,512	57,877
1938	42,018	3,301	13,354	58,673
1939	45,890	3,449	13,057	62,396
1940	48,496	3,514	13,146	65,156
1941	55,332	3,210	13,455	71,997
1942	76,991	3,021	13,114	93,126
1943	140,796	2,700	12,418	155,914
1944	202,626			
1945	251,564†			

* ELLIS, PAUL W., "The World's Biggest Business," p. 10, National Industrial Conference Board, New York, 1944.
† Estimated.

present Federal debt situation, however, can be accounted for almost completely by the long series of deficits beginning in 1931, rising to levels that

Table 34·4. Surplus or Deficit of the Federal Government for the Fiscal Years 1792–1945, Millions of Dollars*

Year	0	1	2	3	4	5	6	7	8	9
1790	....	....	$−1.4	$0.2	$−1.6	$−1.4	$2.7	$2.6	$0.2	$−2.1
1800	$0.1	$3.5	7.1	3.2	3.1	3.1	5.8	8	7.1	−2.5
1810	1.2	6.4	−10.5	−17.3	−23.5	−17	17.1	11.3	1.8	3.1
1820	−0.4	−1.2	5.2	5.8	−0.9	6	8.2	6.8	8.4	9.6
1830	9.7	13.3	14.6	10.9	3.2	17.9	20	−12.3	−7.6	4.6
1840	−4.8	−9.7	−5.2	−3.6	7	7	1.9	−30.8	−9.6	−13.8
1850	4.1	4.9	5.7	13.4	15.8	5.6	4.5	1.2	−27.5	−15.6
1860	−7.1	−25	−422.8	−602	−600.7	−963.8	37.2	133.1	28.3	48.1
1870	101.6	91.1	96.6	43.4	2.3	13.4	29	40.1	20.8	6.9
1880	65.9	100.1	145.5	132.9	104.4	63.5	94	103.5	111.3	87.8
1890	85	26.8	9.9	2.3	−61.2	−31.5	−14	−18.1	−38	−89.1
1900	46.4	63.1	77.2	44.9	−42.6	−23	24.8	86.7	−57.3	−89.4
1910	−18.1	10.6	2.7	−0.4	−0.4	−62.7	48.5	−853.4	−9,033.3	−13,370.6
1920	212.5	86.7	313.8	309.7	505.4	250.5	377.8	635.8	398.8	184.8
1930	183.8	−902	−2,942.1	−2,245.5	−3,255.4	−3,782.9	−4,952.9	−3,252.5	−1,449.6	−3,600.5
1940	−3,740.2	−5,167.7	−19,692.2	−55,900 7	−49,594.6	−53,948				

* "Annual Report of the Secretary of the Treasury on the State of the Finances for the Fiscal Year Ended June 30, 1945," pp. 445, 447, and 449, U.S. Government Printing Office, Washington, D. C., 1946. The figures are for gross surplus and deficit, including debt retirements.

had never been reached before save in wartime and culminating in the unprecedented deficits of the second World War.

Of course, we have learned that it is always difficult to compare monetary situations at different times and places, because of the changing value of the dollar itself. This is as true of government revenues, expenditures, and debt as of anything else. We also realize that a nation of 140 million people can carry much more debt than could a nation of 10 million. These tables which we have been studying need interpretation in the light both of changes in price level and of changes in population. Table 34 · 5 gives the relevant material for Federal debt for each year since 1853. First is shown the actual amount of Federal debt per capita for each year, and then the amount of debt which there would have been if the dollar had maintained a constant purchasing power equal to that which it had in the period 1910 to 1914. One or two noteworthy things about these figures merit mention: (1) There is again the general tendency for an increase. War periods swell the per capita debt and the per capita debt burden, and between wars there is usually some slight reduction; but the debt position is never restored after a war to the antebellum level. (2) We note that there is a tendency for per capita debt to rise in times of depression, and a comparison of the two columns of Table 34 · 5 shows, as we would expect, that the burden of the debt rises at such times even more than does the amount of the debt itself. This is because of the fact that, at these times, the general price level is falling, and not only is the amount of the debt larger, but it is harder to pay in terms of goods and services. By some strange perversity of nature, the times when men's debts are heaviest are also the times when it is hardest to pay them on other grounds.

34 · 8. Growing Cost of Government. Just as Table 34 · 5 indicates with regard to the burden of debt, so we find that other government expenditures have increased when considered on a per capita and "corrected dollar" basis. The real burden of government has become heavier and heavier with each passing generation. Yet, on second thought, we realize that growing cost is not in itself either a desirable or an undesirable thing. Perhaps it is unfortunate that desirable things can be had in our world only at a cost; be that as it may, we are all familiar with the fact that usually we can get the things that we want only by paying for them—by giving up our time, our effort, or our money. More important than, "How much did it cost?" is the question, "Is it worth what it cost?" This latter question is the one that should be applied to the cost of government. The former question can be answered at least roughly in quantitative terms, as we have been doing in the figures and tables of this chapter; but the latter question requires a sort of appraisal which is by no means so simple even if it is far more important.

Governments today perform many more services than did those of a generation ago. Free public education has become almost universal, and the

Table 34·5. Federal Government Debt per Capita, in Current Dollars and in the Equivalent of Dollars of the Years 1910–1914, for Fiscal Years 1853–1945*

Year	Debt per capita		Year	Debt per capita		Year	Debt per capita	
	Current	Equivalent		Current	Equivalent		Current	Equivalent
1853	$ 2.36	$ 2.43	1884	$29.60	$31.90	1915	$11.83	$11.70
1854	1.62	1.50	1885	28.11	33.10	1916	11.96	9.56
1855	1.32	1.20	1886	27.10	33.00	1917	28.57	16.60
1856	1.15	1.10	1887	24.97	29.40	1918	115.65	60.50
1857	1.01	0.91	1888	23.09	26.90	1919	240.09	119.00
1858	1.53	1.65	1889	20.39	25.20	1920	228.33	101.00
1859	1.93	2.03	1890	17.92	21.90	1921	221.10	154.00
1860	2.06	2.22	1891	15.75	19.20	1922	208.97	148.00
1861	2.83	3 18	1892	14.88	19.60	1923	200.10	136.00
1862	16.03	15.40	1893	14.49	18.60	1924	186.86	131.00
1863	33.56	25.20	1894	15.04	21.40	1925	177.82	118 00
1864	53 33	27.60	1895	15.91	22.40	1926	167.70	115.00
1865	77 07	41.60	1896	17.40	25.60	1927	156.05	112.00
1866	77 69	44.40	1897	17.14	25.20	1928	146.69	104.00
1867	73.19	45.10	1898	16.90	23.80	1929	139.40	98.50
1868	69.87	44.10	1899	19.33	25.10	1930	131.49	104.00
1869	67.41	44 60	1900	16.56	20 20	1931	135.37	126.00
1870	63.19	46.80	1901	15.71	19 40	1932	155.93	164.00
1871	58.70	45.10	1902	14.89	17.30	1933	179.21	187.00
1872	54 44	40.00	1903	14.40	16.60	1934	213.65	196.00
1873	51.72	38 90	1904	13.88	16.00	1935	225.07	193.00
1874	50.47	40 00	1905	13.60	15.50	1936	263.01	222.00
1875	49 06	41.60	1906	13.50	15.00	1937	281.80	224.00
1876	47.21	43 00	1907	13.33	14.10	1938	285.41	248.00
1877	45.47	43.00	1908	13.46	14.60	1939	308 29	273.00
1878	45.37	49.90	1909	12.91	13.10	1940	325.63	283.00
1879	47.05	52.40	1910	12.69	12.30	1941	367.54	289.00
1880	41.69	41.70	1911	12.28	12.90	1942	537 35	374.00
1881	39 55	38.10	1912	12.48	12.40	1943	1,001.55	667.00
1882	35.37	32.70	1913	12.26	12.00	1944	1,455.52	958.00
1883	32.07	31.80	1914	12.00	12.10	1945	1,853.01	1,200.00

* The figures for actual per capita debts in current dollars are from "Annual Report of the Secretary of the Treasury on the State of the Finances for the Fiscal Year Ended June 30, 1945, pp. 531–532, U. S. Government Printing Office, Washington, D. C., 1946 These figures have been "corrected" for changes in the general price level by using the index of the wholesale prices of all commodities, and the last column indicates the changes in the real burden of debt as it would have been had the dollar retained a constant purchasing power equal to that which it had on the average in the years 1910–1914.

quality of it has steadily improved along with the increase in the quantity. The highway system, especially since the coming of the automobile, has developed far beyond anything that was dreamed of even when great conquerors built their highways with slave labor. The government has taken over important activities in the field of scientific research. It has developed systems of insurance and social security. It generates electricity. It raises timber. Its supports hospitals and libraries. It employs policemen and firemen. It inspects food and drugs. In fact, its varied activities are far too numerous to catalogue here. And with each one of these activities the sensible question is not, "How much did it cost?" but "Was it worth what it cost? Could we have got the services better from some other source, or would we have preferred to have had other services in place of those which were furnished by the government?"

One difficulty that we face in answering such questions is that, even if we could tell exactly what it costs to perform a particular government service, we would find that the people who enjoy the benefits of government expenditure are not the same ones who pay for them, or at least the services are not distributed (nor is it intended that they should be distributed) in proportion to the payment. Thus, a poor family with six children may make more use of the public schools than a wealthy bachelor, yet the latter may pay far more toward the support of public education. If we ask the father of the poor family, "Was the public education worth what it cost?" he may well answer that he has no idea what it did cost and has not felt in any sense the pinch of it. If we ask the rich bachelor if it was worth what it cost, he may answer that he got nothing in return for his payment, or he may take a deeper, long-run view and say that such payments have enriched and improved the society in which he lives in such a way that his own life is better for it. Certainly, it should be clear that different people will appraise both the cost and the return in very different lights.

34·9. Government and Private Expenditures Contrasted. Private expenditures are made for the purpose of increasing the satisfaction of the spender—for getting those particular goods and services which the spender wants. Government expenditures, on the other hand, are made, at least ostensibly, for the general improvement of society. To be sure, there have been many cases in which they have been made for the selfish and partisan advantage of geographic sections or political parties, but these are perversions of the real purpose of government expenditure.

Private expenditures tend to encourage the production of those goods which individuals want produced, and to allocate the productive resources of society to those industries which yield the goods that are most desired. To be sure, in "voting" for the production of these goods, consumers are not on an equal footing. Those who have the most money carry the most in-

fluence. In the case of government expenditure, at least in democracies, men are on a more nearly equal basis. Insofar as they are eligible to vote, consumers cast one vote apiece in government matters. Yet this difference is perhaps more apparent than real for two reasons: (1) Anyone who has had any political experience knows that people differ in political influence almost as much as they differ in income. Government policies are determined, even in a democracy, by a relatively small group of people. (2) There is the difficulty in the case of government policies that the voter has no opportunity to vote on particular issues. He casts a vote for this or that candidate and, if his candidate wins, he has to go along with all the things for which that candidate stands. If the candidate loses, he finds that his vote carries, in most cases, no weight whatever. Thus the common statement that government sovereignty is diffused equally among the people, while economic sovereignty is unequally distributed, is apt to be misleading.

In the case of private expenditure the consumer has, within the limits of his budget, virtually complete freedom of choice. He selects the things that he wants most. In the case of government expenditures, this is by no means true. Government expenditure carries always the element of compulsion. The taxpayer spends his money for battleships, highways, and education, whether he wants them or not.

Some authors have suggested that government expenditure differs from private expenditure in another way; namely, that governments are much longer lived than individuals or business concerns and that, as a result, governments are able to undertake great systems of permanent improvements which are both beyond the means of individuals and too long-lasting to be worth the while of the individual. Although there is something to this suggestion, it should not be overemphasized. The student of history will realize that few governments are really very long-lived; especially is this true if we are to keep in mind that political overturns often alter government policies radically without a country's actually going out of existence. There are many, many business firms that are older than any existing government; and there are some business firms whose resources are far larger than those of most governments. The important thing, perhaps, is not that governments are long-lived, but that they often act as though they were going to be. In this way they do often make plans and initiate undertakings of a permanent or semipermanent nature.

34·10. The Budget. Governments, like individuals, may, if they wish, handle their financial affairs in a haphazard manner without plan. In practice, however, most governments find it advisable to plan their finances, as to both receipts and expenditures. The financial plan of the government, embodying advanced estimates of receipts and of expenditures, is called the "budget."

The budget of a government is usually a long and complicated document, consisting of several parts. First of all, there is a report of the existing state of the treasury—the funds on hand, the surplus or debt, and an account of the operations for the preceding year. Second, comes a list of expected expenditures for the coming year (or other fiscal period), showing the purposes for which funds are to be expended and the estimated amounts of funds necessary to accomplish these purposes. Third, comes a list of the sources from which these funds are expected to come: tax measures, borrowing, etc. Fourth, there is a series of revenue and appropriation acts, since in modern governments no funds can be expended and no taxes collected or monies borrowed save on the authorization of law. Finally, provision is made for audit of the expenditures by the various agencies of the government, and for the keeping of records of expense, so that the materials may be ready at the end of the period to start the series of activities anew.

The use of budgets by governments is by no means so old as one would imagine. The government of the United States first adopted a budget in 1921, but the movement in this country had started with municipal finance at the beginning of the present century and, turning out to be useful there, had been adopted in turn by many of the states. Wherever it has been tried, whether at the national, state, or local level, the adoption of budgetary procedures has had two major results: (1) It has brought some plan into public finance where there was no plan before, making it easier for governments to operate within their income. (2) It has strengthened the hand of the executive, be he mayor, governor, or President, concentrating in him more authority and power. Government in the United States has tended to become more centralized and authoritarian ever since the adoption of government budgets, and without question largely because of them.

Questions

1. In what sorts of ways is a government more expert in planning expenditures than are its individual citizens? In what ways less expert?

2. Describe in a short paragraph the present importance of government as a spender in our society.

3. List some functions which you think ought not to be performed by government. How would you decide what ought and what ought not to be assigned to government?

4. How are government economic problems like, and how unlike, the economic problems of a family?

5. What are the major things for which the Federal government spends money? The state governments? Local governments?

6. How do Federal expenditures vary from time to time? What are the major causes of change in the direction of Federal expenditures?

7. One would expect the total cost of government to grow as the population grew; but how do you explain the fact that the cost of government per capita has been increasing?

8. Pick out from Table 34 • 2 the major increases in Federal expenditures, and explain why each took place.

9. Using the population figures of Table 12 • 1 and the expenditure figures of Table 34 • 2, compute the per capita expenditures for each year.

10. Plot on graph paper the per capita expenditure figures computed in question 9. In a paragraph describe the significant characteristics of the graph.

11. "Taxation and government expenditure are, in effect, a method of changing the distribution of income in the country." Comment.

12. What are the major differences between government expenditures and private expenditures?

13. Using the "Statistical Abstract," the "World Almanac," the annual reports of the Secretary of the Treasury, or any other convenient source, bring up to date as many of the tables in this chapter as possible.

14. A budget is often thought of as a list of expected expenditures, agreed on in advance. A government budget, however, includes much more. What are the other things that are included?

35. PUBLIC REVENUES

35 · 1. Sources of Public Revenue. As noted in Chap. 34, the cost of government is very high, taking up approximately a fifth of the entire national income in normal years of peace (see Fig. 34 · 1). The funds to meet these heavy expenses may be raised in many different ways, but the four commonest sources are the following:

1. *Sale of Goods and Services.* This source of funds is rather common at the municipal level of government, where cities commonly sell water to their residents and sometimes other services. The Post Office is perhaps the best illustration at the Federal level, although the Federal government also sells grazing rights on national forests, sells electricity, and raises some other revenue from sales. When a government raises revenue from these sources, it sets a price on the good or service and the public is free to buy or not as it sees fit. There is no element of compulsion. Yet there is one important difference between sales of goods by the government and those by private business in that the government does not ordinarily try to make a profit on its sales. The aim is social benefit rather than profit, and the transactions are carried on either at cost or possibly even at a loss. In the latter case it should be clear that the government must subsidize the activity at least in part from other sources.

The sale of goods and services by the government has not been at all a popular method of raising government revenue in the United States. Our predilection has been definitely in favor of private enterprise, and the general public has usually objected strongly to government competition with private business. Many countries have gone further in the direction of government business activity than has the United States, especially in the direction of national ownership and operation of public-utility enterprises—railroads, telephone and telegraph lines, etc. In the United States the amount of government revenue from these sources is relatively unimportant.

2. *Borrowing of Funds.* Governments in need of funds often borrow at least part of them. This is obviously only a temporary expedient, however.

Ultimately, the government will be forced to raise funds by other means to pay back the debt and the interest on it. Yet governments find, as do private individuals, that the needs for funds vary considerably from time to time while the income is more or less uniform and steady. Borrowing makes it possible to "even out" the fluctuations, taking care of sudden needs for funds until such time as the revenue system can be adjusted to new needs. As noted in Chap. 34, most governments have borrowed enough so that the interest payments on their public debt are a real burden—one of the larger expenditures; and the borrowing of the Federal government has reached such proportions as a result of war and the government financial policy of the preceding depression as to make the problems of ultimate repayment exceedingly difficult.

3. *Taxes.* This is by far the commonest source of government revenue. It differs from the sale of goods in that the individual is no longer given freedom in deciding whether he wants to make the given expenditure or not. When a municipality sells water, each resident is permitted to decide whether he wants to use city water or not; if he does want to use it, he is allowed to use as much or as little as he wants. On the contrary, when the government decides to spend $50,000 for a new school building or a million dollars to pave a highway or a billion dollars to strengthen the navy, the individual citizen is not asked whether he wants to buy his share of the school building, the highway, or the navy. Taxes are always compulsory. This is one of their distinguishing characteristics. A permissive payment is never a tax, even though it may be called so colloquially. Taxes have the additional peculiarity that they are not apportioned according to the benefit that the taxpayer receives. The man who has several children in the public schools pays no more school tax than does the man whose children are sent to private schools; the man with a long police record pays no more for the support of the state penitentiary than does the man of unblemished character. This, of course, contrasts with the situation that exists when the state sells goods or services. When a city sells water, the large user pays more than the small user. The man who writes many letters pays more to the post office than does the man who writes few. If a payment is really a tax, its size is not connected with the amount of benefit that the taxpayer gets from it. We might state, by way of definition, that *a tax is a compulsory payment to a government to be used for paying the general expenses of the government and paid without reference to the amount of benefit received by the taxpayer.*

4. *Fees and Special Assessments.* In connection with its ordinary functions carried on for general benefit, the government not uncommonly performs services that are of especial benefit to particular persons. Thus, for example, the government regulates the practice of medicine, law, and dentistry because such regulation is believed to be of advantage to the public at

large. Its intent is not to regulate these occupations for the benefit of the persons within them, but for the public as a whole. Yet, in carrying out this regulation, the public incurs certain expenses, and persons wishing to enter or to continue in these occupations are a source of expense to the government. The government often in such cases charges a fee to cover, or partly cover, the cost incurred. License fees for owning dogs, hunting and fishing, getting married, driving automobiles, etc., are all illustrations of this kind. These payments are not taxes for two reasons: (1) They are not compulsory. One need not drive an automobile if he does not wish. (2) They are at least in considerable measure apportioned according to benefit. The man who registers three automobiles or who marries three wives pays more than the man who is content with one.

If a city decides, for the benefit of all its people, to pave a particular street, lay a public sewer along a certain street, or build a sidewalk in a particular section, some people will gain much more than others. Everyone in the city gains from good roads, but the people who own property abutting on the new street gain more than anyone else. In such cases, it is not uncommon to recover at least part of the cost of the improvement from a special charge levied against those who gain the most. In this case also, it is obvious that the charge is not really a tax, since it depends on benefit received. *A fee is a payment made to a government in return for either some special benefit received or some special cost incurred by the government to the advantage of the taxpayer.* In those particular cases where the benefit received consists of increasing the value of real estate, the fee is ordinarily called a *"special assessment."*

The student will notice at once that he cannot depend on everyday usage to determine whether a given payment is or is not a tax. For example, many states levy what is called a "gasoline tax." If the proceeds of this tax were used for general government expenditures, such as paying the costs of the courts, the schools, and the fire department, it would indeed be a tax. But, in practice, the proceeds of the gasoline tax are commonly set aside for use in construction and repair of the highways and, since expenditures for gasoline are in our society closely correlated with use of the highways, it is easy to see that this payment is rather closely apportioned according to the benefit received by the purchaser of the gasoline or according to the cost to the government of repairing and maintaining the roads. What we call a gasoline "tax" is actually a form of fee. On the other hand, any honest-to-goodness fee never exceeds the amount of the cost involved; if any so-called "fee" does exceed the actual cost, it is to that extent a tax no matter what it is called.

35·2. Objectives in a Tax System. To be sure, the immediate objective in any tax system is that of raising revenue. But different governments use

very different systems of taxation, and two systems that raise exactly the same amount of revenue may differ very much in other ways. The incidental effects of any system of taxation are far-reaching and important, and every government chooses among alternative possibilities largely on the basis of these incidental effects. In addition to the raising of revenue, then, what would we like to accomplish with a tax system? What other requirements would we impose beyond the requirement that the system be productive?

Different people give somewhat different answers to this question. For example, it is an old saying in tax circles that "in taxing, as in plucking geese, the idea is to get the greatest possible amount of feathers with the least squawking!" Those who accept this rule would try to select those taxes which will cause the least public complaint. They will prefer taxes that the ordinary citizen does not see and does not pay directly. For example, if we put a tax on the milling of wheat flour, the tax will be paid by the relatively small number of millers. They, in turn, will add the tax to the price of their flour, and ultimately the housewife will pay it in an increased cost of bread. But the thousands of housewives will not realize that they are paying a tax at all, and consequently they are not likely to "squawk" about it, as they would if the tax were a tax on the sale of bread at retail.

In opposition to this idea, we have many economists who would argue that every tax ought to be open and plain, where the taxpayer can tell just what he is paying and why. They would argue that it is hard to curb the rising cost of government if the people who bear the real burden of that cost do not realize that they bear it. If a man is forced to pay $500 cash in taxes, he is likely to complain but, if this same $500 is taken from him little by little without his knowing it, he is far less likely to object. As we said in Chap. 34, the important question in public finance, as in private economy, is not "How much did it cost?" but "Was it worth what it cost?" If, however, one cannot find out what it cost he can answer neither question.

Adam Smith, the famous English economist of the late eighteenth century, suggested four rules or Canons of Taxation as guides in the selection of taxes. His canons have received so much comment that we may well use them as a basis for our own discussion of the objectives of a tax system. They state

1. The subjects of every state ought to contribute toward the support of the government as nearly as possible in proportion to their respective abilities; that is, in proportion to the revenue that they, respectively, enjoy under the protection of the state.

2. The tax that each individual is bound to pay ought to be certain, and not arbitrary. The time of payment, the manner of payment, the quantity to be paid, ought all to be clear and plain to the contributor and to every other person.

3. Every tax ought to be levied at the time or in the manner in which it is most likely to be convenient for the contributor to pay it.

4. Every tax ought to be so contrived as both to take out and to keep out of the pockets of the people as little as possible over and above what it brings into the treasury of the state.

These maxims are perhaps to some extent conflicting. It is possible, for example, that in an effort to arrive at justice under the first maxim we might have to sacrifice certainty under the second. Moreover, the maxims are by no means complete. Modern authors would suggest certainly some others, of which we might list the following:

5. The tax system should be elastic; that is, it should be possible to increase or to decrease the amount of revenue raised by the tax easily and quickly in accordance with changing needs.

6. The tax system should be dependable and productive. Although it should be possible to alter the revenue at will, the revenue should not vary unexpectedly or haphazardly.

7. The tax system should not bring with it indirect and incidental reactions that are socially undesirable. Thus a tax system might raise large amounts of revenue but only by discouraging desirable enterprises or encouraging undesirable ones.

These seven maxims or canons can be classified as follows:

1. General social goals:
   - The canon of universality.
   - The canon of uniformity or equal sacrifice.
     (These two principles comprise Canon 1 above).
2. Fiscal canons:
   - The canon of adequacy (Canon 6 above).
   - The canon of elasticity (Canon 5 above).
3. Administrative canons:
   - The canon of economy (Canon 4 above).
   - The canon of certainty (Canon 2 above).
   - The canon of convenience (Canon 3 above).
   - The canon of harmlessness (Canon 7 above).

Although each of these canons is important, combining to make a good tax structure, perhaps the doctrines of universality and of equality of sacrifice have received the most comment—the doctrines that the tax system should bear on everyone, and on everyone with the same weight. We shall return for further comment on this matter after defining one or two terms that we shall need to use.

35·3. The Tax Base. Governments could dodge many problems by adopting a simple system of levying the same amount of tax against everyone. We noted in Chap. 34 that, even in times of peace, the total expenditures of all levels of government in the United States run to roughly 20 billion dol-

lars per year[1] and, if we think of the population as amounting to roughly 140 million persons we see that a system of actual equality would take about $140 or $145 from each man, woman, and child. A family of five would receive an annual tax bill of about $700, possibly even more. When we look back at Table 4 · 1, we see that such a tax is larger than the entire annual income of perhaps 10 per cent of the people of the country in average years. Even if a family had an income of $1000, the payment of a tax of $700 would be utterly impossible. In other words, actual mathematical equality of tax payments is impossible and the effects would be undesirable. It is necessary for us to find some other basis for determining how much tax each person or family will pay.

In any actual case of tax administration, the tax varies with something even if it does not vary with the number of persons in the family. It is based on some indicator of taxability: the amount of income, of property owned, of money spent, or something else. This criterion which determines the amount of tax liability is called the "tax base." Different taxes have different bases. In the case of the personal income tax, the base is the taxpayer's net income after the making of certain deductions. If we want to know how much income tax he owes, it does us no good to learn how many cows he has or how old he is. The base is the amount of his net income. On the other hand, in the case of the general property tax, it does us no good when we are trying to estimate a man's tax liability to know his income, but it does help us now to know how many cows he has and what their value is, how much his house is worth, how much his land is worth, and the value of his other property. The common tax bases are wealth owned, income received, or expenditures made (consumption).

35 · 4. Proportional, Progressive, and Regressive Taxes. After we have established a tax base, we may levy taxes in direct proportion to it. Thus, if the base consists of property, a man owning twice the amount of property of another will pay twice the tax. Such taxation, in which the amount of the tax is always the same proportion of the tax base, is called "proportional taxation." Property taxation is often, at least superficially, proportional. Thus, for example, a town may lay a tax of "25 mills on the dollar" (alternatively, called in some areas a tax of "25 dollars on the thousand"). This means that every property holder must pay a tax of 25 mills for every dollar's worth of property that he owns (or a tax of $25 for each $1000 worth of property that he owns). Since 10 mills make a cent, 25 mills amount to 2½ cents, and a tax of 25 mills on the dollar (like a tax of $25 on the thousand) amounts to a tax of 2.5 per cent of the value of the property. If *A* owns ten times as much property as *B*, he will pay ten times as much tax, since each pays 2½ per cent of the value of his property.

It seems to most people that equal taxation would be unfair—that

[1] See Table 34 • 1.

charging, as we have said, $140 to $150 taxes to each man, woman, and child would not really be an equal burden. It would entail much more real sacrifice for some families than for others. The proportional scheme sounds better. If each man has to give up 2 or 10 per cent of his property or wealth to the state, the burdens are probably more nearly commensurate and the canon of equality more nearly satisfied than would be true if each person paid the same absolute amount. *The principle of equality calls for equality of sacrifice or equality of burden rather than for arithmetical equality in the size of the tax.*

Many people would argue, however, that even proportional taxes do not make for equal burdens. They would illustrate by comparing two families with, say, incomes of $500 and $50,000 per year. Suppose each of these families is charged an income tax of 10 per cent. The rich family pays a tax of $5000 while the poor family pays a tax of $50. The rich family pays far more in absolute amount, but most people would say that the poor family still makes the greater sacrifice. The rich family, giving up $5000, gave up luxuries that they will not badly miss. The poor family, giving up $50, has been forced to give up medicine, clothing, or food that was desperately needed. Most people would agree that, even with proportional taxation, the actual tax burden will rest more heavily on the poor. They, therefore, advocate levying taxes on the rich at a larger percentage of the tax base than those levied on the poor. In such a case, where the tax becomes a larger and larger percentage of the tax base as the tax base increases in size, we say that the tax is "progressive."

The Federal income tax in the United States is a good example of a progressive tax. In computing taxes on personal incomes for the year 1945, for example, taxpayers paid 3 per cent of their net income (minus exemptions) as a normal tax. The allowance of exemptions in itself makes the tax progressive, as we can see by an example. Imagine several families with different net incomes, each with an exemption of $500. If they pay 3 per cent on their net incomes after deducting $500, their tax payments will be as shown in the accompanying table.

	Family A	Family B	Family C
Net income	$500	$3000	$10,000
Deduction	500	500	500
Taxable income	0	2500	9,500
3 per cent tax	0	75	285
Tax as per cent of net income	0%	2.5%	2.85%

As the income gets larger and larger, the rate of the tax also gets larger and larger, approaching a rate of 3 per cent. But, in addition to this

normal tax, the taxpayer was charged a surtax on his surtax net income, which was computed by taking even further deductions. This surtax amounted to 20 per cent on his first $2000 of surtax net income, 22 on the next $2000, 26 on the next $2000, 30 on the next $2000, and so on up until finally taxpayers with very large incomes paid surtaxes of 91 per cent of any surtax net income above $200,000.

A tax may, of course, be proportional on one base and not on another. Let us imagine, for example, that a state charges sales taxes of 2 per cent on every purchase of any kind. The tax base is the amount of expenditure, and the tax is obviously proportional—always 2 per cent. But such a tax is not apt to be proportional when compared with income. Poor families find it necessary to spend their entire income in order to keep going, so that their income and their expenditures are the same. Rich families save considerable parts of their incomes, so that their expenditures are smaller than their incomes. We noted in Table 33 • 1 that most of society's savings occurs in high-income families. In 1941, the average family with an income under $500 not only made no savings but had a deficit of $88; the average deficit for families in the income class from $500 to $1000 was $16; the average savings of families with incomes from $1000 to $1500 were $43; the average saving for families with incomes from $1500 to $2000 were $123; and so the average savings rose as the income rose until families with incomes between $3000 and $5000 saved an average of $483.

Let us compare two of these typical families, one with an income of $750 and the other with $4000. If these two families are representative, the first will spend $16 more than its income while the second spends $483 less than its income. If each pays a sales tax of 2 per cent on all purchases, their tax situations will be as shown in the accompanying table. The tax is pro-

	Family A	Family B
Total income	$750	$4000
Total expenditures	766	3517
Tax, 2 per cent of expenditures	15.32	70.34
Tax as per cent of income	2.04%	1 76%

portional when compared with expenditures but, if we compare it with incomes, we find that the higher the income the smaller the proportion taken in taxes. Such a tax is called a "regressive tax." If, as the tax base gets larger, the proportion taken by taxes remains constant, the tax is proportional. If, as the tax base gets larger, the proportion taken by taxes gets larger, the tax is progressive. If, as the tax base gets larger, the proportion taken by taxes gets smaller, the tax is regressive. The principal objection that most students

make to the general sales tax is that, despite its dependably high yield and the relative ease of collecting it, it tends to be regressive when compared with income. Often, when one speaks of a tax as being regressive or progressive, if nothing else is said, it is understood that the speaker means "regressive or progressive with respect to income," or "with respect to ability to pay" measured on any basis.

There is substantially unanimous agreement among students of the subject that a progressive tax system is necessary if there is to be even a close approach to equality of sacrifice on the part of taxpayers. While this must be true for the tax system as a whole, however, it need not be true of every single tax within the system. Most governments rely on a combination of taxes, hoping to get high yields with some, elasticity with others, and still other purposes with others. As we shall see later, for example, many of the states are making use of sales taxes to a considerable extent at present, despite their largely regressive nature, yet often these same states combine the sales tax with progressive income taxes in such a way that the entire tax system, taken as a whole, is progressive.

35 · 5. Measuring Ability to Pay. If we are to get substantial equality of sacrifice, we must adapt the size of each man's tax to his ability to pay. Yet we must not rely on any purely subjective appraisal of this ability to pay. The tax that each man owes must be definite and certain. It must be possible to establish its amount beyond dispute. Therefore, in practice, a tax base is always selected which reflects at least roughly the taxpayer's ability to pay, yet which can be measured with reasonable accuracy and objectivity.

During those early years of American public life in which our tax systems were being established, our economy was almost exclusively agricultural. At such a time, when 9 out of every 10 men made their living from the farm, the size of a man's holdings of land and other real estate was a fairly good indicator of his prosperity and his ability to pay. At that time, it was natural for governments to rely heavily on property taxes. Such taxes may be levied against particular sorts of property, or they may be levied against all property. In the latter case they are called "general" property taxes. We shall see shortly that these taxes still form the backbone of tax systems at the local level; but, as our economic life has changed, they have come less and less to represent ability to pay. A farmer with large holdings of land and buildings may today make an income of $2500 per year, while a merchant or trader who earns $25,000 per year may work in rented quarters, live in a rented house, and own little or no property. Under a system of property taxes, the farmer would pay more tax than the trader even though his ability to pay was less.

The difficulty with property taxes is magnified by the fact that many kinds of property escape taxation almost completely. The taxing authorities

can see a man's land and buildings. They are impossible to hide. If a man owns stocks and bonds, they can be hidden away in his safe-deposit box where no one can see them, and the tax gatherers are likely to overlook them entirely. In practice we find that "general" property taxes usually bear especially heavily on real estate, and that personal and intangible property usually escape in large measure.

As men's ownership of property has become less and less valid as a measure of ability to pay, more and more dependence has been given to incomes as a basis for taxation. There are, of course, many technical problems in defining exactly what is and what is not income for tax purposes, and naturally ability to pay is more closely associated with net than with gross income. The Federal government could not, under the restrictions of the Constitution, use income taxes until the passage of the Sixteenth Amendment in 1913. Since that time the income tax has become increasingly important in Federal finances, and many of the states have also relied on it heavily.

The income tax has one serious weakness from a fiscal standpoint, however, which became especially apparent early in the depression of the thirties. In such times, the need for public funds increases as programs of public works and relief are inaugurated; but men's incomes shrink, and the revenue from the income tax shrinks with it. The revenue from the income tax is not dependable enough at such times. Because of this difficulty, those public officials who have the responsibility for planning tax programs have hunted for some measure of ability to pay other than ownership of property or receipt of income. In many cases they have come to the conclusion that a man's expenditures—his actual purchases—can be used as a measure of ability to pay. They argue that if one man is able to spend ten times as much as another man for ordinary things, he can also spend ten times as much for the support of the government. For this reason, we have had a considerable increase in recent years in sales taxes.

Just as property taxes may be general or may apply only to specific kinds of property, so sales taxes may be general or may apply to specific kinds of purchases. Often they apply exclusively, or at unusually high rates, to luxury products on the assumption that these products will be purchased mainly by people who are comfortably well to do. In other words, in selecting the things to be taxed an effort is made to exclude the necessities of life like food and clothing, which are characterized by inelastic demand; and to include luxuries like cameras, sporting goods, liquors, or admission to the movies. These specific sales taxes may be so planned as to be progressive with respect to income and to fall in largest part on those people who have high incomes. If this is to be done, the taxes must lie mainly on those goods for which the demand is elastic, but in such cases there is the fiscal difficulty

that the tax will add to the price of the goods and thus reduce very decidedly their sales. A taxgatherer who wants a dependable source of revenue and who has no scruples about ability to pay will prefer to lay his taxes on goods for which the demand is inelastic, since people will continue to buy these goods despite the imposition of the tax and the raising of the price.

35·6. Special Problems of Property Taxation. We have now described in a very general way the nature of the basic kinds of taxes, and we shall turn our attention to them one by one in an effort to learn something of the problems connected with each.

The actual administration of a property tax involves several steps. First, it is necessary to decide how much money we wish to raise from the property tax. This may be decided by a legislature, board of aldermen, board of county commissioners, an old-fashioned New England town meeting, or by any other competent public body. Thus, for example, the Town of Mansfield, Conn., in making its plans for the year 1946, adopted a budget that included total expenditures for schools, roads, courts, and all other items of $126,487.50. It was estimated, however, that the town would receive income from other sources (grants from the state, fines in the town court, dog license fees, interest on invested funds, etc.) of $54,156.27. This left $72,331.23 to be raised by the property tax.

Having decided how much to raise, the problem still remains of discovering how much is to be contributed by each taxpayer—how the burden of $72,000 is to be distributed. In the case of a property tax, it is distributed, of course, in proportion to the value of property owned. This can be done only by compiling a list of the property in town, showing not only the total amount of property but the ownership of it. In every community there is some property that is exempt from taxation. Thus, it is usual for the law to exempt all government-owned property, such as the town hall, the county courthouse, the state park, and the schools. Also, it is customary to exempt from taxation the property of eleemosynary organizations such as private schools and colleges, churches, etc. Therefore, what is really needed is not a list of all the property in the taxing unit, but a list of all the taxable property. If we continue with our illustration from the Town of Mansfield, we can look in the official town report and find a list of the names of all property owners in the town, with a statement of the value of various kinds of property owned by each. A summary of this list shows that the total taxable property in the town has a value of $2,530,702. Certain additions are then made in the form of penalties imposed on persons who failed to report their property accurately, and certain deductions are made for personal exemptions to veterans and others who are allowed under the law a considerable exemption from taxation. Having made these additions and deductions, there remains an amount of $2,411,041, which is called the "grand list" of the town. The grand list of any taxing

unit is a list of property owners showing for each the value of their taxable property not exempt from taxation.

From here on, the matter is one of simple arithmetic. If the town must raise $72,331.23 to cover its budget and if it is to be raised by laying taxes against a grand list of $2,411,041, it will be necessary to raise 30 mills on every dollar of property on the list, or $30 on every $1000 worth of property on the list. We would say that the "tax rate" is 30 mills, or in some localities it would be said that it is $30. The tax rate is found by multiplying the budget by 1000 and dividing by the grand list. In algebraic terms we can write our formula as follows:

$$\text{Tax rate} = \frac{1000\ (\text{budget})}{\text{grand list}}$$

In our example, where the grand list was $2,411,041 and the budget to be raised by property taxation was $72,331.23, our formula gives us

$$\text{Tax rate} = \frac{1000\ (\$72{,}331.23)}{\$2{,}411{,}041} = 30$$

The answer arrived at by use of this formula is always in terms of mills per dollar or dollars per thousand; so in our present case we can see that the tax rate for the Town of Mansfield will be 30 mills on the dollar, or, as it is commonly stated, merely 30 mills.

The problem still remains, of course, of determining the amount of the tax bill for each individual taxpayer. Each such taxpayer will have to pay a tax of 30 mills (3 cents) for each dollar's worth of property that appears opposite his name on the grand list. If, for example, John Doe is recorded as owning $6908 worth of property, his tax bill will amount to 6908 times 3 cents, or $207.24.

35·7. Assessment and the Tax Rate. The student will note from the formula in the preceding section that the budget, the grand list, and the tax rate are so connected that, if any two of them are known, the third can be computed at once. Their interconnection is such that, if we change one of them, we must change at least one of the others. The budget, or list of expenditures, is usually determined within fairly narrow limits by the needs of the town or other taxing unit; but the grand list and the tax rate can be, and often are manipulated more or less arbitrarily to the confusion of the uninitiated.

The grand list is a list of the taxable property of the taxing unit, showing for each piece of property its value for tax purposes. Values, as we know, are often hard to determine accurately and, if two competent men are each asked to appraise the value of a given farm, they may easily differ by 10 per cent or more in their estimates. Values of property for tax purposes are determined by a public board of appraisers, called in most jurisdictions a

"board of assessors." This board receives lists of property from taxpayers and others, and assesses or appraises the property for tax purposes. Often the owner of the property is allowed to suggest what seems to him a reasonable value, but the board is in no way bound to agree with his evaluation and the real responsibility is theirs. The law often stipulates that assessors must enter property at its full and fair market value; but this requirement is unfortunately honored more in the breach than in the observance. In practice, boards of assessors very commonly appraise property at half its real value, or even less.

The effect of such underevaluation is obvious. Suppose, for example, that the assessors of the Town of Mansfield follow the practice of appraising property at half its actual value. Their grand list records $2,411,041 worth of property, so there must really be twice that amount or $4,822,082 worth of property in the town. If the grand list has been made up correctly, we would have used this latter and larger figure in computing the tax rate and, since the grand list would have been twice as large, the tax rate would have been but half as large as before, or 15 mills instead of 30.

Now this new tax rate of 15 mills would not represent any actual change in the value of the property in the town, but merely a difference in the method of assessment. If we go back to the case of John Doe whose property was formerly listed as worth $6908, under the new appraisal the grand list will show that he has $13,816 worth of property. At first, it appears as though Mr. Doe would suffer from the change, since the appraised value of his property has doubled; actually we now apply a tax rate of 15 mills against his property instead of the former tax rate of 30 mills, and his tax will be 13,816 times 1.5 cents, which makes his tax $207.24 exactly as before. Doubling the assessment rate has made it possible to cut the tax rate in half without charging anyone any more taxes.

This is often overlooked by taxpayers who worry about high tax rates. A tax rate alone tells us nothing about the tax burden. We must know whether property is assessed at 100 per cent of its value or at 50 or 30 per cent. The assessment rate is just as important as the tax rate in determining the actual tax burden. The two fundamental factors that determine whether John Doe's tax burden will be heavy or light—whether he will have to pay $200 in taxes or $500—are the budget of the taxing unit, and the proportion of the total property in the grand list that belong to him.

There are many reasons for preferring that all property be actually assessed, as the law ordinarily requires, at its full sale value. In practice, however, assessment is usually in the hands of untrained and inexpert part-time workers—farmers, merchants, laborers, small businessmen. In most cases, they do the best they can, relying in part on the estimates of value given them by the owners (such estimates being naturally biased downward when

given for this purpose) and in part on the records of their predecessors in office. It has often been pointed out that, when an incorrect assessment is once entered on the books, it tends to be copied year after year by new assessors who have never seen the property and are not familiar with it, and who in many cases would be at a loss to fix a better value even if they did try.

35·8. Capitalization of Taxes. Let us suppose that Mr. Carter owns a farm which appears on the grand list at a value of $8000. The tax rate in Carter's district has for a long time been in the neighborhood of 20 mills, which means that Carter has paid an annual tax of about $160. Having paid his tax bill and all his other expenses, Carter has been making a net income of about $1500 a year from his farm. As we learned in Sec. 26·9, it is possible to capitalize this net income to learn the approximate sale value of the farm as distinguished from its assessed value. If the going rate of interest in the community is, say, 5 per cent, then the capitalization formula tells us that Carter's farm is worth roughly $30,000; in fact, the farm probably would sell for something like that amount.

Now let us suppose that a new tax is laid, and it is expected that the new tax will be permanent. Suppose that, under the new system, the tax rate rises to 30 mills, so that Carter's tax bill jumps to $240 per year. This is an increase of $80 and, if other things remain as before, Carter's net income will drop by $80 to $1420. If it is expected now that the farm will yield a net income of only $1420 per year, the capitalized value will be only $28,400. The value of the farm has fallen, as a result of the new tax, from $30,000 to $28,400, or a reduction of $1,600. The tax was increased from $160 to $240, or an increase of $80 per year. If we capitalize the amount of the tax, $80 a year, at the same rate of 5 per cent, we find that the capitalized tax amounts to $1,600 or that it is exactly equal to the reduction in the value of the farm.

This informal numerical example probably seems to the student very theoretical, hypothetical, and unreal—very textbookish. Yet every businessman is keenly aware that exactly this sort of thing always does happen in practice. Every farmer knows that, when we raise the taxes on his farm, we reduce the value of his farm and that, if we raised the taxes far enough, we could make his farm absolutely worthless. The same thing is true of factories, stores, dwelling houses, or any other durable property. As we raise the taxes, we reduce the net income and, as we reduce the net income, we reduce the value of the property. This is what men have in mind when they make the statement that taxation is to some extent a confiscation of property by the state. What perhaps they have not realized is the simple point that is evident from our numerical example; namely, *if taxes on durable property are increased and if it is anticipated that the increase will be permanent, the value of the property is reduced by an amount exactly equal to the capitalized amount of the tax.*

35·9. Who Bears the Burden of Taxation? If we state our last conclusion in another way, it may be illuminating. Mr. Carter had a farm worth $30,000. Now that the tax has been increased, its value drops to $28,400 and, if Mr. Taylor buys it today he will buy it at the new and lower figure. He has saved $1,600 on the purchase price. If, as we have assumed, the going rate of interest in the community is 5 per cent, Taylor can put the $1,600 which he has saved into other investments and earn 5 per cent on it. This will bring him an annual return of 5 per cent of $1,600, or $80. But $80 is exactly the amount by which the tax was increased. Taylor, who has now bought the farm, will never in any real sense have to pay this new tax of $80. The saving that he made in buying the property was sufficient to pay all the taxes from now until doomsday. And Carter, who sold the property, has received $1,600 less than would otherwise have been true. Carter has sacrificed enough to pay all the taxes forever. The important point to understand in this connection is that *old and established and permanent taxes (or those which are expected to be permanent) are not a burden to the people who pay them year after year in any real sense.* If they were wise, they considered the earning power of the property when they bought it, and paid for it only what it was worth after the deduction of taxes. The entire burden of a new permanent tax is borne immediately when the tax is first passed by the owner of the property at the time, and takes the form of a reduction in the value of the property. This reduction is sufficient so that, at the current rate of interest, it will pay all the taxes for the future.

35·10. Is Real Estate Overtaxed? It is common for students to point out, when appraising the property tax, that it falls almost completely on real estate. For example, the town report of the Town of Mansfield, Conn., for 1945, shows that over 78 per cent of the property carried on the tax books was real estate, and another 13 per cent consisted of motor vehicles which could not well be hidden from the tax collector because they had to be registered. This means that all the farm machinery, livestock, furniture, jewelry, etc., made up less than 10 per cent of the assessed value of the property in the town. In this university town but one taxpayer was listed as owning a library, and the total value of the watches, diamonds, and jewelry owned by the people of the town was $150 according to the assessor's records. In fact, the assessor finds it almost impossible to discover these items of personal property, and to a considerable extent they escape taxation.[1] Thus, in practice, almost all the property listed is real estate, which lies there for all to see, and it is commonly argued that real estate bears as a consequence an unduly heavy load of taxes, that land and buildings are overtaxed, and that their owners deserve some sort of tax relief.

[1] The Town of Mansfield is chosen for this illustration not because it is unusual, but because it is typical and because its town report happened to be handy. The same sort of situation can be repeated almost anywhere; in many cases, the situation is even worse.

In part, this argument is based on unassailable figures. Taxes do lie very heavily on real estate. But the conclusion does not follow from the premises. The heavy taxation of real estate is not a burden on its present owners except to the extent that the taxes are new taxes which have been imposed since the present owners took possession. If we go back to our numerical illustration, we can perhaps make this clear. We assumed that Taylor bought a farm from Carter after a tax had been imposed. If Taylor now cries about the tax burden, it is fair to point out to him that he bought the farm at a price that included his appraisal of the tax burden—that it was borne by the former owner, Carter. Suppose, not understanding this, we were to answer Taylor's plea by cutting off the new $80 tax and going back to the original $160. Now the value of the property will revert at once to its original $30,000, and Taylor will have received an unearned gift from the government in the amount of $1,600.

It is an old maxim among tax men that "an old tax is a good tax." This is commonly taken to mean merely that people always complain about new taxes to which they are not accustomed but that, after a tax has been enforced long enough so that people are used to it, they do not complain so vociferously about it. The saying has much more economic justification than that. After a tax has been in operation for some time, values of property become adjusted to it, and the old tax is a good tax in the sense that it is no longer a real burden on anyone. Those who bore the burden are dead and gone. But a new tax always confiscates the property of someone and in that sense is a "bad" tax.

In the consideration of any reform of the general property tax, this problem may well be faced squarely. Intangible and personal property pretty generally escapes taxation, and it seems that it is certain to continue so to do. For that reason, it would seem wise to abandon the idea of a general property tax and to concentrate property taxation on those kinds of property which can be easily and quite accurately discovered and evaluated, such as real estate and motor vehicles. It would also seem wise, where at all possible, to put assessment in the hands of specialists who have been trained for the purpose, getting away from elected or politically appointed boards of part-time workers and entrusting the problem to those who have some possibility of doing it well.

35·11. Special Problems of Income Taxation. In the preceding five sections we have given some attention to problems that arise especially in connection with the taxation of property. We now turn to special problems of income taxation.

As has already been pointed out, income taxation is much more recent than property taxation; yet in the few years since it has come into common use it has become one of the major sources of government revenue. Some states began to adopt this tax base as early as a century ago, but the early

acts were in general unsuccessful. The Federal government first tried the income tax in connection with the Civil War, but the measure was soon repealed. Another such measure was passed in 1894 but was declared unconsitutional without ever having actually been used. The Constitution declares[1] that "no capitation, or other direct, tax shall be laid, unless in proportion to the census or enumeration hereinbefore directed to be taken." Thus, under the Constitution, since the population of Mississippi in 1943 was 25 per cent greater than that of Connecticut, no direct tax could be levied by Congress unless the people of Mississippi paid 25 per cent more than did the people of Connecticut. Yet, in 1943, the total income of the people of Mississippi was only 42 per cent as great as that of the people of Connecticut. With income less than half as great they would be required, under any direct tax as described in the original Constitution, to have paid a quarter again as much. Such a restriction seemed so unfair that Congress was driven to abandon direct taxation until, in 1913, the Sixteenth Amendment to the Constitution made income taxation possible. Since that time, the income tax has not only come to be a major source of revenue for the Federal government, but has also been adopted in some degree by all but 15 of the states.[2]

35 · 12. Present Income Taxation in the United States. The outbreak of cartoons every March depicting the poor taxpayer mystified over the technicalities and complications of his income tax return reflects the general impression that income taxation is an esoteric art beyond ordinary comprehension. Although this is by no means the case and although income tax forms are really worked out carefully to make computation easy and avoid ambiguity, nevertheless it is true that the subject is too large a one for us to treat here in any detail. The most that we can do is to mention some of the more important characteristics of it, remembering that the details vary from year to year anyway.

In the first place, income taxes are levied on net rather than on gross incomes. The taxpayer is permitted to deduct from his gross income any business expenses incurred. In addition, some income is exempt from taxation. Most important of these exempt incomes are those received as social security benefits or as sickness or injury benefits, interest on certain tax-exempt government bonds, gifts, bequests, or inheritances, and proceeds of life insurance.

Having found the net taxable income, the taxpayer is allowed to make

[1] Art. I, Sec. 9.

[2] In 1943, all states taxed individual incomes with the exception of Connecticut, Florida, Illinois, Indiana, Maine, Michigan, Nebraska, New Hampshire, Nevada, New Jersey, Ohio, Pennsylvania, Rhode Island, Texas, Washington, and Wyoming. Of these, Connecticut, Pennsylvania, and Rhode Island levied corporation income taxes and are therefore sometimes included among the states using income taxation.

certain deductions. Chief among them are deductions (not exceeding 15 per cent of the taxable net income) for contributions made to religious or charitable organizations, for interest paid on personal debts, for certain sorts of taxes, for medical and dental expenses under certain conditions, and for losses of property from fire, storm, shipwreck, etc. Finally, he makes a deduction for his own normal-tax exemption, and his normal tax is computed on the remainder, being at present (1946) a flat 3 per cent for everyone after exemptions. Further deductions are then made for other dependents, and the surtax is computed at sharply progressive rates on the new remainder. The income tax that one pays is the sum of the normal tax and the surtax.

This all sounds much more complicated than it usually is in practice. Especially has it been true in recent years that men with low or medium-sized incomes can escape most of the computation by looking up their tax in a table, which is part of the income tax return. This table gives them their personal and family exemptions, and allows them a rough average amount for other deductions. Knowing their annual income, they are able to find the amount of their tax immediately from the table.

35 · 13. Withholding Taxes. During the war years, personal income taxes ran to such large amounts that it was well-nigh impossible for most people to save enough to pay them. Taxpayers failed to realize how large their tax was to be, saved too little to cover it, and discovered when it came time to file their returns that they did not have enough left to pay it. For this reason, and also to reduce somewhat the cost of collection, the policy has been followed in recent years of collecting a large part of income taxes at the source, before the income earner ever gets the income at all. The income is earned but never received. Employers are required to withhold specified proportions of income and to turn them directly over to the government. These amounts, withheld from the income earner, are credited to his account, and at the end of the year he finds that all or most of his tax has already been paid. If it happens that a larger amount has been withheld than he actually owes, the government refunds the difference to him. Withholding has become such an important part of tax collection that over 55 per cent of all individual income tax collections in 1945 were collected by this means.

35 · 14. Purpose of Income Taxation. The purpose of taxation is ordinarily the raising of revenue for the government. When, however, a government uses a progressive income tax which collects funds largely from the rich and uses it to furnish services which are to a considerable extent used by the poor, the effect of the tax is to bring about a redistribution of wealth. Many legislators openly advocate the use of income tax for this purpose, and many others at least unconsciously favor the result. To be sure, it is likely that the end itself would be unconstitutional if only openly avowed, but many things can be done by indirection which are forbidden as direct measures.

There has long been difference of opinion as to the desirability of government attempts to alter the distribution of wealth and income. It is interesting to note in retrospect that, when the income tax was first inaugurated, it was definitely stated that this was not its goal, and one of the chief economic advisers in the President's cabinet at the time when the constitutional amendment became effective said of the income tax, "There will, on the other hand, be those who will desire and try to use the income tax as a leveling measure rather than as a revenue measure without regard to equity. They will have to be controlled by those who believe in equal justice."[1] The matter is one on which there continues to be difference of opinion, but at least one must understand that the Federal government has no constitutional authority to alter the distribution of wealth or income, and whatever it does in this direction must be done unintentionally, incidentally, surreptitiously, under pretense of other goals. Legally, even if not always actually, income measures, like other tax measures, are for raising revenue.

35 · 15. Elasticity of the Income Tax. Throughout the early years of experience with the income tax, it was thought that one of its great advantages was its elasticity. The government could easily and quickly alter the amount of revenue coming in by changing the rates of taxation. But the revenue raised by an income tax depends not only on the tax rate, but also on the amount of income earned by the citizens of the country. This latter factor is by no means so easily controlled and, when the depression of the late twenties and early thirties brought sharp and sudden reductions in personal incomes, the receipts from the income tax fell off drastically despite increases in the rates. The government discovered that in times of depression, when public expenditures are usually rising and the need for revenue is increasing, the income tax is not a dependable source of income. In the emergency of the depression, when the revenue was badly needed, the income tax receipts of the Federal government fell off by over two-thirds and put the government in a serious financial position.

35 · 16. Direct and Indirect Taxes. In Sec. 35 · 2 it was pointed out that some taxes, such as a tax on the milling of wheat flour, will be paid in the first instance by the miller, but will be added by him to the price of the flour and thus passed along to the ultimate consumer. Such a tax, the burden of which is passed along from the original taxpayer to other people, is called an "indirect tax," while, if the original taxpayer has no way to shift the tax to others and has to carry the burden himself, the tax is called a "direct tax." When a tax is passed along from one person to another, we say that the tax is "shifted."

The tax collector, of course, is interested to know who is originally

[1] HOUSTON, DAVID F., "Eight Years with Wilson's Cabinet," vol. i, p. 76, Doubleday, Doran & Company, Inc. New York, 1926.

responsible to pay any tax and how much he is obligated to pay. But the statesman is also interested in the indirect effects. He wants to know who bears the ultimate burden. The pattern of original payments may be very different from the pattern of ultimate burden, or what the economist calls the "incidence" of the tax. He would say, for example, that a tax on the milling of flour is paid in the first instance by the miller, but "the incidence of the tax is on the ultimate consumer of the bread." The part of the preceding sentence in quotation marks means that the ultimate consumer of the bread is the person on whom the burden of the tax finally falls—the one who is unable to shift the burden and is forced to bear it.

The shifting of taxes is, unfortunately, often misunderstood. It is assumed that the producer of a product, like the miller of wheat flour in our example, derives some new power to raise prices as a result of the tax. After our earlier study of the pricing process, we realize that a manufacturer or merchant or other producer is usually at liberty to raise his price if he wishes, and presumably before the tax is laid he is already charging that price which yields the greatest possible net income. The mere fact that a tax has been laid does not make it desirable for him to raise his selling price and thus shift the tax to his customers. We have discovered that it will pay him to raise his price, and thus to pass the tax along, at least in part, if the tax is part of his variable expense, but that it will not pay him to raise his price and shift the tax in any degree whatever if the tax is part of his fixed expense.[1] We have also learned that a tax of some given percentage of the net income will not be passed along. Thus, our earlier study prepares us at least in some measure for an understanding of the problem of tax shifting.

A tax on the raw materials used by a producer, such as the tax on the milling of flour in our earlier illustration, is part of the variable expense. The greater the volume of output, the higher the tax to be paid. Thus, taxes on the manufacture or sale of commodities are usually shifted in large part to the consumers of those commodities. On the other hand, if the demand for the commodities in question is elastic, the increased price as a result of the tax will discourage consumption in a very considerable degree. Those consumers who continue to use the commodity will bear their share of the tax, but those who reduce their purchases will bear the burden on only the reduced amount that they continue to use, and those who give up the use of the commodity entirely will bear no burden whatever. Thus, in fact, only a part of the tax can be shifted to consumers and *the more elastic their demand for the goods taxed, the smaller the degree to which the tax can be shifted.*

The situation can be easily visualized by two diagrams, as in Fig. 35·1. In each diagram the solid curves represent the supply and demand curves for a commodity before a tax is laid on it. In the upper diagram the demand

[1] See Chap. 17.

is elastic; in the lower, it is inelastic. The supply curves are identical in the two diagrams. In each case it is assumed that a tax is now levied on the production of the commodity in question, and that the tax amounts to $1 per unit of output. This will change the supply curve. Consumers will still be willing to buy as much as before at any given price; their tastes and incomes have not changed. But, at any given price, the producer will now be willing to sell only the quantity that he formerly offered at $1 less. For example, he will now be willing to sell at $5 only as much as he formerly sold at $4, since

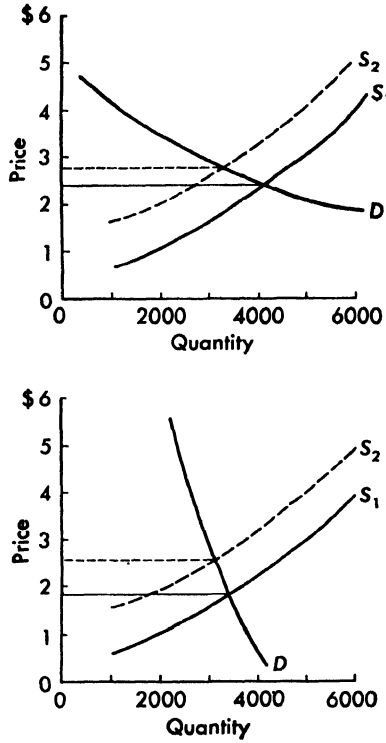


FIG. 35 • 1.—Effect of elasticity of demand on tax shifting.

after he pays the tax he will have but $4 left. The supply curve is therefore shifted vertically by an amount equal to the tax ($1) on the vertical scale. Thus, in each diagram, we get a new supply curve, shown by the broken line, parallel to the old curve and higher by the amount of the tax. Since in both cases we have assumed a tax of $1, the two supply curves are identical in both diagrams and the only difference is in the elasticity of demand.

If we notice what happens in the two cases, we see that where the

demand is very elastic, as in the upper diagram, the price of the commodity rises from about $2.40 per unit to about $2.70 per unit when the tax of $1 is raised. The tax is $1, but the price rises only about 30 cents. There is, however, a considerable reduction in the quantity sold. In the lower diagram, which represents an inelastic demand, the price rises from about $1.90 before the tax, to about $2.60 after the tax. A tax of $1 raised the price to consumers by about 70 cents and brought about only a slight reduction in the quantity consumed. Inspection of these diagrams will make it clear why *the producer of a commodity finds it easier to shift a tax by raising prices if the demand for his product is inelastic than if it is elastic.* This is perhaps another reason why selective sales taxes are often confined to luxury products, where the tax is naturally harder to shift.

Another easy rule to understand about the shifting of taxes is that *the more general the applicability of the tax, the less likely it is to be shifted.* For example, if a tax is laid on the use of land for growing melons, the producer can with relative ease shift his land to other products if the consumer does not assume the burden. If, on the other hand, the tax is somewhat more general and covers all land used for agricultural purposes, it will be harder for the owner to find other profitable uses in nontaxed fields, and he will be more likely to continue in production and to bear part of the tax himself. Finally, if the tax covers all land regardless of use, the owner of the land has no possibility of escaping the tax, since there are now no tax-free uses to which he can turn. He, therefore, tends under such circumstances to bear the whole burden of the tax himself. Similarly, a tax on the incomes of physicians that did not apply to incomes of other people would tend to be shifted to the patients, for otherwise the tendency would be, for some of the physicians at least, to turn to other occupations or for people who would have studied medicine to take up aviation or zoology instead. The number of physicians would fall off, and their services would become scarce and dear, thus passing the burden along to the patients. A general tax on all incomes cannot be passed along, since there is no way for the taxpayer to escape it save by getting along without any income whatever.

Keeping in mind these principles, we might summarize the situation with regard to actual taxes by saying, first, that probably every tax is borne in some part by the initial taxpayer and that most taxes are in some measure passed along. Instead of having but the two categories of direct and indirect taxes, we find all degrees of shifting from none at all to almost complete shifting. Yet there are great differences in degree, and we might well classify the commoner taxes by saying, second, that taxes that are shifted only to a slight extent are the general property tax, the net income tax, taxes on monopoly profits, inheritance or estate taxes, poll taxes, and taxes on economic rent. On the other hand, taxes that are largely shifted are sales taxes on

particular commodities, property taxes that apply to only specific kinds of property, import duties, etc.

35 • 17. Use of Various Taxes in the United States. Most taxing authorities in this country use not one tax, but a combination of several. They also ordinarily raise some revenue from nontax sources. But the types of taxes

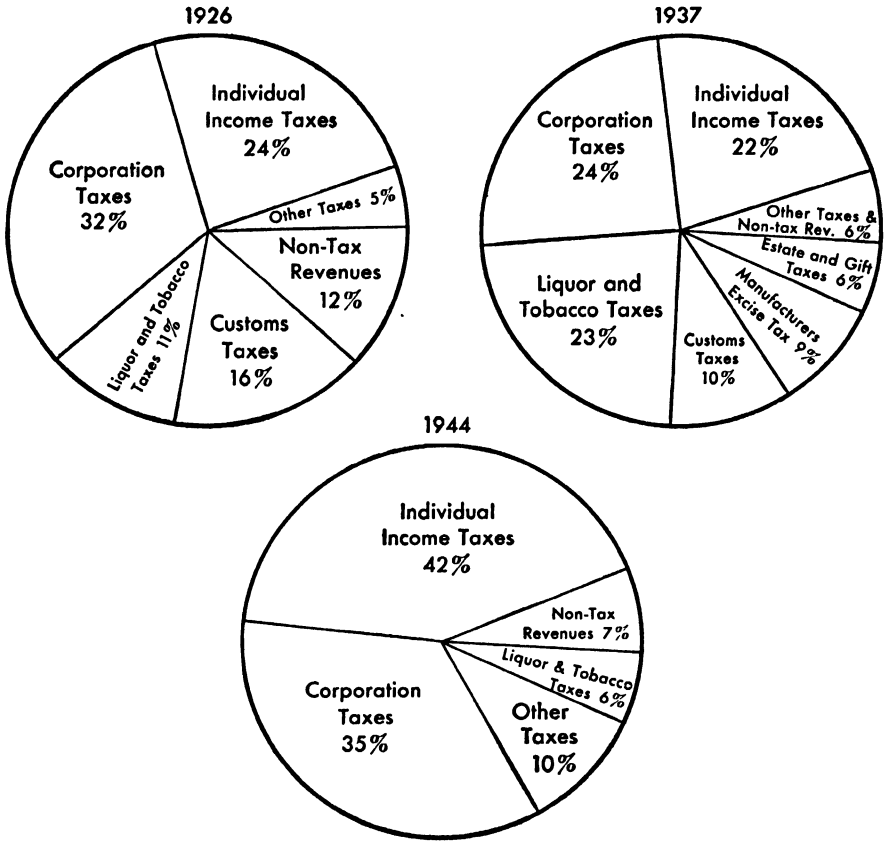


FIG. 35 • 2.—Sources of revenue of the Federal government in 1926, 1937, and 1944. (*Based on data in Paul W. Ellis, "The World's Biggest Business," National Industrial Conference Board, New York,* 1944, *p.* 41.)

used and their relative importance vary greatly from government to government.

The situation, as far as the Federal government is concerned, is depicted in Fig. 35 • 2, which shows the relative importance of different taxes and of nontax revenue in the years 1926, 1937, and 1944. Comparison of these diagrams will show that the individual income tax is a very important element in the system and is growing more important lately. Corporation taxes

also play a very important role. Customs taxes have been growing less and less important. During the depression some added emphasis was given to consumption taxes on luxury items, notably tobacco and liquor, but during the war the tremendous increase in revenues was handled very largely by additions to personal income and corporation taxes, until they came to com-

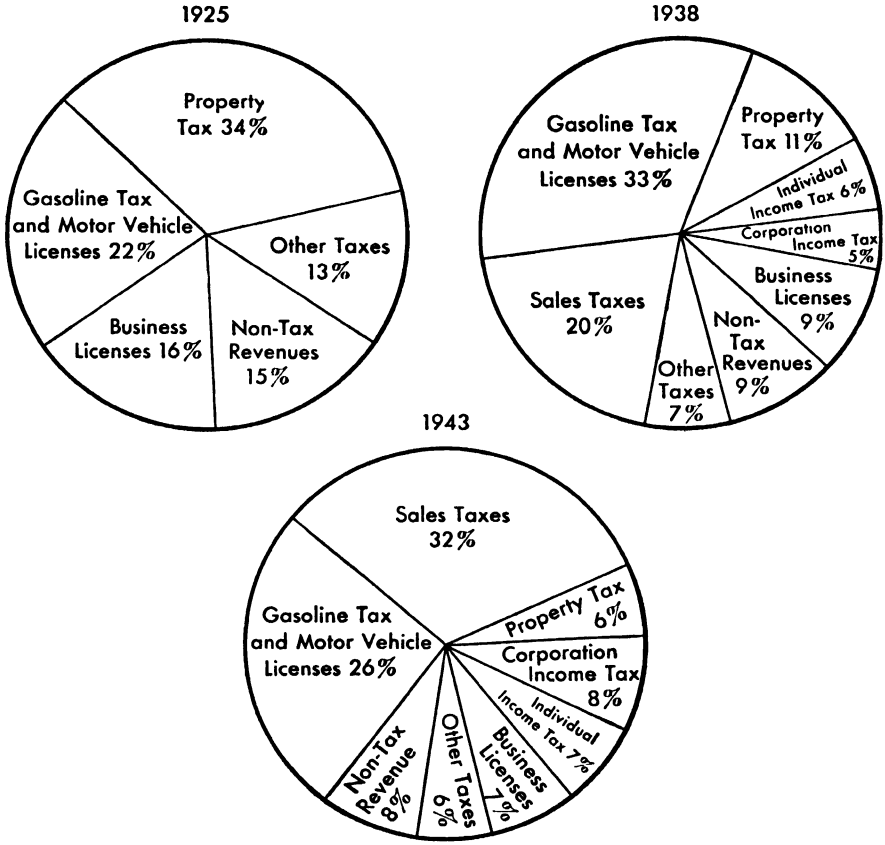


FIG. 35 • 3.—Sources of revenue of the state governments in 1925, 1938, and 1943. (*Based on data in Paul W. Ellis, "The World's Biggest Business," National Industrial Conference Board, New York,* 1944, *p.* 87.)

prise more than three-quarters of the entire Federal revenue. Nontax revenues have been consistently small.

The individual states differ greatly in their tax systems, but Fig. 35 • 3 shows the situation for the states as a whole (or what amounts to the same thing, the average situation among them). The rapid wane of the importance of the property tax is perhaps the most obvious feature on these diagrams, and the concomitant growth in the reliance on sales taxes. The states also rely

heavily on taxes on gasoline (one type of sales tax, of course) and on motor vehicle licenses.

Information on units of local government is very difficult to find, anc probably not particularly reliable. Some indication of the situation can be obtained, however, from Fig. 35·4, which shows the sources of county revenues and of the revenues of large cities in 1942. One glance at these charts is sufficient to demonstrate the continued reliance of local governments on the property tax almost to the exclusion of everything else. Cities and towns usually receive a considerable amount of income from the state as grants for the aid of education, for the building of highways, etc. In 1942, almost a fifth of the income of the large cities came from such sources, and it is likely that

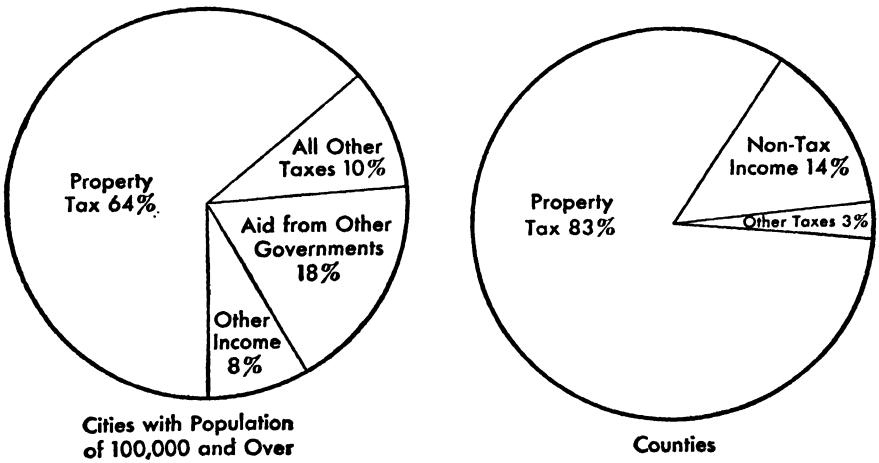


FIG. 35·4.—Sources of revenue of large cities and of counties in 1942. (*Based on data in Paul W. Ellis, "The World's Biggest Business," National Industrial Conference Board, New York,* 1944, *pp.* 129 *and* 134.)

smaller units received even larger proportions in the form of grants-in-aid. The three sets of diagrams in Figs. 35·2 to 35·4 will repay careful study on the part of the student, since they summarize a very large amount of information on the relative importance of various taxes at different levels of government, and on the trends of change therein.

35·18. Volume of Government Revenues. As we have repeated several times before, governments all over the world are tending to become more and more expensive as they tend to perform a larger number of functions. In part, the added expense has been reflected, as noted in Sec. 34·7, by a rising public debt. On the other hand, governments have been under the necessity of finding more and more lucrative taxes and other more productive sources of revenue to meet the increased demand for funds.

Table 35·1 shows the revenues of the Federal, state, and local govern-

ments in the United States each year since 1914. Comparison with the corresponding table on expenditures in Chap. 34 (see page 817) will show that,

Table 35·1. Revenues of Federal, State, and Local Governments in Millions of Dollars, 1914–1945*

Fiscal year	Federal	State	Local	Total
1914	$709	$412	$1,673	$2,794
1915	669	450	1,817	2,936
1916	743	454	1,904	3,101
1917	1,077	512	2,071	3,660
1918	3,618	576	2,230	6,424
1919	5,102	633	2,883	8,618
1920	6,609	791	3,410	10,810
1921	5,533	924	3,774	10,231
1922	3,958	1,061	3,773	8,792
1923	3,764	1,135	3,917	8,816
1924	3,777	1,246	4,285	9,308
1925	3,507	1,302	4,523	9,332
1926	3,647	1,529	4,895	10,071
1927	3,873	1,633	5,159	10,665
1928	3,741	1,805	5,470	11,016
1929	3,693	1,924	5,666	11,283
1930	3,890	2,030	5,886	11,806
1931	3,008	2,096	5,623	10,727
1932	1,905	1,949	5,506	9,360
1933	2,009	1,800	4,915	8,724
1934	3,052	2,035	4,857	9,944
1935	3,724	2,214	5,019	10,957
1936	4,062	2,718	5,030	11,810
1937	4,973	3,310	5,145	13,428
1938	5,609	3,456	5,309	14,374
1939	4,990	3,372	5,309	13,671
1940	5,175	3,613	5,450	14,238
1941	7,700	3,932	5,493§	17,125§
1942	12,560	4,240	5,527§	22,327§
1943	21,982†	4,286	5,552‡	31,820‡
1944	43,608†	4,430‡		
1945	43,838†‡			

* ELLIS, PAUL W., "The World's Biggest Business," p. 7, National Industrial Conference Board, New York, 1944.
† Includes taxes refundable after the war.
‡ Estimated.
§ Preliminary.

when expenditures undergo a sudden increase, as in times of war, the revenues increase also but more slowly. At least temporarily, the disparity is made up by an increase in the public debt. When, however, the expenditures again drop back toward their normal level, the high tax rates continue for a time, and there is a tendency for revenues to exceed expenditures and for the public debt to decrease. The most noticeable fact in Table 35 • 1, however, is the gradual but continued increase in the volume of revenues.

Table 35 • 2. Receipts of the Federal Government for the Fiscal Years 1792–1945*

Year	0	1	2	3	4	5	6	7	8	9
	Millions of dollars									
1790	....	....	$ 3.7	$ 4.7	$ 5.4	$ 6.1	$ 8.4	$ 8.7	$7.9	$ 7.5
1800	$10.8	$12.9	15	11.1	11.8	13.6	15.6	16.4	17.1	7.8
1810	9.4	14.4	9.8	14.3	11.2	15.7	47.7	33.1	21.6	24.6
1820	17.9	14.6	20.2	20.5	19.4	21.8	25.3	23	24.8	24.8
1830	24.8	28.5	31.9	33.9	21.8	35.4	50.8	25	26.3	31.5
1840	19.5	16.9	20	8.3	29.3	30	29.7	26.5	35.7	31.2
1850	43.6	52.6	49.8	61.6	73.8	65.4	74.1	69	46.7	53.5
1860	56.1	41.5	52	112.7	264.6	333.7	558	490.6	405.6	370.9
1870	411.2	383.3	374.1	333.7	305	288	294.1	281.4	257.8	273.8
1880	333.5	360.8	403.5	398.3	348.5	323.7	336.4	371.4	379.3	387.1
1890	403.1	392.6	354.9	385.8	306.4	324.7	338.1	347.7	405.3	516
1900	567.2	587.7	562.5	561.9	541.1	544 3	595	665.9	601.9	604.3
1910	675.5	701.8	692.6	724.1	734.7	697.9	782.5			
	Billions of dollars									
1910	....	....	....	....	....	....	....	1.1	3.7	5.2
1920	6.7	5.6	4.1	4	4	3.8	4	4.1	4	4
1930	4.2	3.2	2	2.1	3.1	3.8	4.1	5	5.9	5.2
1940	5.4	7.6	12.8	22.3	44.1	46.5				

* "Annual Report of the Secretary of the Treasury on the State of the Finances for the Fiscal Year Ended June 30, 1945," pp. 445, 447, and 449, U. S. Government Printing Office, Washington, D. C., 1946. The figures do not include receipts under Title VIII of the Social Security Act.

That this increase is a matter of long standing, at least on the Federal level, is indicated by the figures of Table 35 • 2, which lists the Federal revenues for each year since 1792. If we neglect the temporary aberrations of war years, we can say that the revenues of the Federal government have risen during this period from about 4 million to about 5 billion dollars a year. In other words, revenues have increased well over a 1000 fold. During this same period the population has increased from about 4 million to 140

millon persons,[1] which is an increase of about 35 fold. The figures seem to indicate, as a matter of fact, that the per capita revenues of the Federal government have increased roughly 35 fold during its history.

Our figures on the general price level in Chap. 21 do not go back quite so far as 1792, but an inspection of Fig 21 • 1 and of Table 21 • 3 will indicate that the present price level is at least roughly comparable with that of the late 1790's. Information from other sources leads us to believe that the price level of the late 1930's and early 1940's was, if anything about 10 per cent above that of 1792. Realizing that the available data are crude and approximate, we can conclude that the actual per capita burden of Federal government revenue has increased since 1792 perhaps 30 or 35 fold.

35 • 19. Nonrevenue Aspects of Tax Policy. Throughout this chapter we have considered taxes primarily as sources of government funds. This is their original and most important function. Yet it must be acknowledged that indirect effects of taxes, entirely outside the field of government revenue, are of tremendous importance.

One should note, in the first place, that there are limits to taxation. A system of taxes, so high or so badly arranged as to discourage production, or the incentive to work, or the willingness to save, may bring immediate advantages of large revenue at the expense of long-run social welfare. One must always compare the services performed by the government with the other things that taxpayers could have had for their money. If the government takes, in the form of taxes, funds that taxpayers would otherwise have spent for things more valuable to them than are the services of the government, then the taxes are too high. If, on the other hand, taxes are reduced and taxpayers thereby have more to spend, but if there is nothing for which they can spend it that they want so much as they want the services of government, then the taxes are too low. The problem is again one of alternative opportunities—of opportunity costs such as those treated in Sec. 30 • 9.

We should also emphasize the importance of the relation of taxation to distribution. It is a common statement that the government takes from its citizens in the form of taxes and returns to them in the form of services, that taxation involves no net loss, since the government gives back schools, roads, and navies in return. From the standpoint of the individual, however, the problem is by no means so simple. Taxation almost always involves a redistribution of wealth and income. The government takes more from some men than from others. Similarly, some people make far more use of government services than others. But there is little relationship between the amount that any individual pays in taxes and the amount of government services that he receives. The heaviest taxpayers are sometimes the ones

[1] See Table 12 • 1.

who make the least use of the public schools, libraries, and playgrounds. There is serious question how far the government can go in taking from the well to do and giving to the poor without demoralizing both. There is even some question as to the legitimacy of the use of the tax system for the indirect accomplishment of ends other than the raising of revenue.

Questions

1. When a government sells goods and services, it often sells them at or below cost because it is not trying to operate at a profit. Under such circumstances, however, it is clear that the government raises no revenue from the transactions. If we imagine that a city government were to manufacture illuminating gas and sell it to its citizens with the idea of raising as much revenue as possible, how, if at all, would its economic decisions (selling price of the gas, volume of gas sold, number of men employed, etc.) differ from those of a private concern in the same field?

2. Taxes are, of course, a burden on the people. What would prevent a government from eliminating this burden on its people entirely by borrowing all its funds instead of raising them by taxation, and then refusing to pay its public debt when it came due?

3. "If schools were supported by a tax on each school child, paid by his parents, the payment could not properly be called a tax at all." Do you agree?

4. What is the main difference between a fee and a tax? Between a special assessment and a fee?

5. The canon of universality states that every citizen should bear in some measure the burden of government—that none should escape entirely. Do you think that anyone does escape entirely in the United States? What bearing do indirect taxes have on your answer?

6. Would you expect a sales tax on overalls to be regressive, progressive, or proportional when compared with income? Why? How about a sales tax on steam yachts? On telegrams? On admissions to moving-picture theaters?

7. In a given income-tax system a man with an income of $1200 pays a tax of $30 and a man with an income of $6000 pays $100. Is the tax progressive, regressive, or proportional?

8. Explain why a general sales tax tends to be regressive when compared with income.

9. Someone has suggested that the imposition of a tax should leave the people in exactly the same relative positions that they occupied before it was imposed. Would this suggestion require, as you see it, proportional, progressive, or regressive taxation?

10. The property tax has in practice turned out to be primarily a tax on real property, and in our modern society the ownership of real property is no longer a good index of one's ability to pay taxes. Suppose, however, that it were possible to tax all property, including not only land and buildings but also personal property, cor-

porate securities, savings, and everything else. Would such a general property tax in your opinion be roughly in proportion to the taxpayer's ability to pay?

11. A government raised $150,000 with a tax rate of 25 mills. What was the size of its grand list?

12. A government wishes to raise $90,000 from property taxation. The grand list is $3,600,000. What will the tax rate be?

13. Explain the relationship between the assessment rate and the tax rate.

14. County A has a tax rate of 20 mills while County B has a tax rate of 28 mills. County A appraises property at 90 per cent of its true value while County B appraises at 75 per cent of its true value. What is the relative tax burden in the two counties? In other words, would a piece of property in County A be taxed more or less than an indentical piece in County B, and how much more or less?

15. A farm is yielding a net income of $3000 a year in a community where the rate of interest is 6 per cent. A new and permanent tax takes $100 per year from this farm. How much does it reduce the farm's value?

16. Explain why it is that a permanent tax imposed before a given piece of property was bought by its present owner is not in any real sense a burden on the present owner even though he has to pay it year after year.

17. In what sense is it true that an old tax is a good tax?

18. What is meant by the elasticity of a tax?

19. What is the difference between a direct and an indirect tax?

20. Suppose a new tax is laid on a commodity, and the producer tries to shift it by adding it to the selling price. What happens if the demand for the product is elastic? If it is inelastic?

21. Under competitive conditions, the demand for the products of any single producer is infinitely elastic. Suppose we were to lay a tax on just this single producer, taxing him $1 per unit of output but leaving his competitors untaxed. Could he or could he not shift the tax? Why or why not? How does your answer gibe with that to question 20?

22. Why do you suppose that the income from customs duties, as shown in Fig. 35 • 2, fell off in relative importance between 1926 and 1937? What happened to the volume of foreign trade in this period?

23. Using the data of Tables 35 • 2, 12 • 1, and 21 • 3, compute the real burden of per capita government revenues for each year listed in all three tables.

24. Plot on graph paper the data computed in question 23.

25. Try as far as possible to explain the major changes depicted on the graph made in question 24.

36. TYPES OF ECONOMIC ORGANIZATION

36·1. Economic Goal. Men are actuated by a mixture of motives, sometimes conflicting. In some measure they attempt to adopt those modes of conduct which will, as they see it, add to their wealth; at the same time, they are trying to reach other objectives based on love, religion, malice, ambition for social position or political prestige, etc. Under such conditions, it is by no means unusual for economic motives to be subordinated for the time being and for men to take action that runs counter to their own economic interest.

It is by no means easy to separate economic interests from other interests. As we noted in Chap. 2, economics has to do with the satisfaction of wants. These may be, and often are, wants for material goods but, on the contrary, they may center around such abstractions as desire for social position, desire for the approbation of one's fellows, or the desire to win the hand of some fair maiden. The fact that these latter wants do not involve material goods does not remove them from the field of economics. Economics is not necessarily materialistic. Yet, economics does not include all problems of want satisfaction, but only those which involve the use of scarce factors of production. The use of honeyed words to win one's love is outside the realm of economics but, as soon as reliance is placed on boxes of chocolates, diamond rings, or a flashy motor car, the problem has become at least in part an economic one. It is not the kind of want that determines whether or not a situation is economic, it is the method of satisfying the want.

When we dress the definition up in technical jargon, we say that *economics involves the allocation of scarce means between alternative desirable ends.* When we whisper sweet nothings in Mable's ear, there are just as many sweet nothings left to whisper to Kate. No scarce resources have been utilized. When we decide to invest a month's pay in a diamond, we are forced, whether we like it or not, to give up other satisfactions that we could have had with the expenditure of the money. Even if our primary motive is one of religion, malice, philanthropy, greed, love, or hate, if we seek to

reach our goal by the use of scarce resources, our problem becomes an economic one. Thus the problems of life are often mixed, and the field of economics is by no means confined, as the superficial thinker often supposes, to materialistic motives.

36 · 2. Social Organization and Economic Goals. Psychologists and sociologists tell us that men are gregarious—that they prefer to live in groups rather than in isolation. Yet the fact that men do usually live in groups could easily be explained on other and primarily economic grounds. Many things that men cannot do at all or which they can do only with great difficulty alone can be done and done easily by cooperation. Joint action of men makes it possible to satisfy wants more fully—to satisfy wants that could never have been satisfied at all by men living as hermits. In Chap. 10 we discussed some of the advantages of the division of labor and saw how cooperative effort may, and usually does, increase the amount of the product. Although men doubtless have other motives for coming together in societies, surely one of their strongest is the economic one.

It is perhaps reasonable, then, that men should judge the society in which they live on the basis of its success along economic lines. Different people have organized their societies in different ways, and what is more natural than that we should now ask ourselves whose organization works best? Is the Russian plan or the fascist plan better than ours? Would it be desirable to alter the structure of our own society in some ways? If so, how? How does our society work well, and how does it work poorly?

As we start out on a very brief treatment of the problems involved in answering these questions, it is important to emphasize again that men's motives are mixed. It is possible that men might prefer a society that was not efficient along economic lines because it had such great advantages in other directions. Men may prefer to live in poverty rather than to give up what seem to them certain spiritual advantages of a system to which they are accustomed. In this chapter we shall confine our attention almost exclusively to economic aspects of social organization, but wise is the student who will realize that he and his fellow citizens must give their attention to other aspects as well.

36 · 3. A Comment on Teleology. Before we go too far in any discussion of this matter, however, it should be made clear that any society grows to a considerable extent in an aimless way—that its organization is by no means entirely the result of careful forethought, planning, and purpose. If we take as an example the society in which we live, which some men call "capitalism," we ought to realize that to a considerable extent, like Topsy, it "just grew." No one "planned it that way." Men did not know that they were building a capitalistic system. In fact, this name was given to it only recently, and by people who objected to some of its characteristics.

The child who is told that "it rains in order that the grass may grow" may accept this purposive explanation if he likes, or he may reject it. The assignment of purpose to everything is a common human characteristic, especially with the childish. Similarly, it should be pointed out that quite possibly our social organization has no real purpose, no real aim or goal. Without doubt, individual men within the society have aims and motives,[1] but to ascribe goals to an entire society and to attempt to draw distinctions between the goals of a capitalist, a communistic, and a socialistic society, is to run the risk of overstating the role of purpose in men's lives. In what follows we are apt to state the case, for ease in understanding, largely in terms of purposes, but the student is warned in advance that many of the characteristics of any society took form without any conscious guidance or intent.

36·4. Two Basic Societal Characteristics. The social life of Iowa differs from that of Georgia. The social life of any large city differs from that of neighboring small towns. The social life of an agricultural region differs from that of a manufacturing center. There are all sorts and degrees of difference in social structure. When we consider the economic aspects of society and when we look at the problem in the large (as we are forced to do if we are to compare the three or four main types of economic organization), we find that there are two major kinds of decisions that must be made which will in turn color our decisions on many minor decisions.

In the first place, if our society is to be an efficient one, the people within it must cooperate in production. How can this cooperation be brought about? How can we get men to perform those tasks which other men need to have done? There are in general two ways in which we can induce desired action in other people: by reward or by punishment—by encouragement or by coercion. Some societies give men freedom to do what they want, but at the same time set up systems of incentives to encourage them to do things that will be for the social benefit. Other societies use force to get the desired ends. One has but to contrast a slave society with our own to see the fundamental nature of the difference.

In the second place, every society requires some planning in connection with its consumption problems. Somehow there must be decisions as to what goods will be produced and how the scarce resources of society will be used. In a society in which nothing was scarce it would be unnecessary to make such decisions. We would satisfy all the wants. But, in any earthly society, it is impossible to satisfy all the wants, and it is necessary to establish some order of priority—to decide which wants will be satisfied and which will not. The decisions with regard to such plans may be left in the hands of in-

[1] Even this statement would be too teleological in the eyes of a behaviorist psychology which denies that men have purpose any more than, say, sodium hydroxid and hydrochloric acid have a "purpose" when they combine to make salt and water.

dividual consumers or they may be made by governmental bodies of some sort. In a democracy, we are accustomed to thinking that "majority rule" is wise, fair, and desirable. Yet many of our decisions are not made on such a basis. For example, even if a majority of the people of the United States decide that they want to go to the movies once a week, the person who dislikes the movies need not go nor need he support the movies in any way. The government could, of course, decide how many movie theaters were needed and then by means of taxation force everyone to contribute toward their support. In some societies economic planning takes this general form. Economic decisions are made on a social rather than on an individual basis. In other societies, on the contrary, individual consumers are left free in most cases to make their own decisions. If they wish to run counter to the majority, or even to everyone else, in their consumption decisions, they are at liberty to do so. The difference is one of individual freedom of choice.

As we describe various proposals for social organization, the student will see these two distinctions arising over and over again. The way in which a society gets its members to cooperate in production decisions, and the extent to which it allows them freedom in the making of consumption decisions, are perhaps the two most important and fundamental characteristics of economic organization. No society ever adopts a completely extreme position on either characteristic. Every society depends to some extent on coercion in production and to some extent on reward. Every society leaves some consumption decisions to individuals and makes others on a community basis. The differences between various societies, although differences in degree, are nevertheless important enough to merit our careful attention.

36·5. Capitalism and Private Property. The economic system under which we live in the United States is, as we have suggested before, a nondescript system which developed slowly and piecemeal. Yet it has certain characteristics to which its critics gave the name "capitalism." As we shall see, no society has ever tried a purely capitalistic system, yet many of its attributes are to be found in almost any modern American community.

Foremost among these characteristics is that of private property, which we discussed in Sec. 3·4. As we saw there, property is a social institution which consists of a relationship between men and goods. It would be quite possible to operate a society without any property and, even if we had property, it could be public rather than private. Most systems that are suggested as alternatives to capitalism do not propose to abolish property, but rather to shift in whole or in part from a system of private to a system of public property. Some would turn over all property to the public, while others would turn over only production goods (capital) and leave consumption goods in private ownership. Especially, as we have seen,[1] do some people

[1] See Chaps. 11 and 31.

object to private property in land. Be that as it may, in our capitalist system all these kinds of goods are ordinarily privately owned, and advocates of the system argue that the right of private ownership gives a stimulus to production that makes such private ownership worth while.

To be sure there is a mixture of private and public property in our system. Few people would argue that public property ought to be completely abolished, and few people would argue that our system of private property is perfect as at present established. We must realize that private property is not a "natural" condition nor an inherent right but an artificial man-made institution which was established by society and which is guaranteed by it; that under such circumstances the property owner is under obligation to society not to turn his property to antisocial ends. Private property must be administered as a social trust and, where it is turned to social disadvantage, society can and should protect itself. The defender of the capitalist system bases his defense, therefore, not on any supposed inherent right nor on the basis of defending the *status quo* and the entrenched interests, but on the argument that private property works out to social advantage and that it is from the social standpoint the most satisfactory way of disposing of and utilizing wealth.

36·6. Inheritance. The right to control wealth during one's lifetime does not necessarily imply the right to dispose of it after one's death. It is quite possible to imagine a society in which private property was recognized but which did not recognize the right to bequeath property on death or the correlative right to inherit property left by others. Under such circumstances, it would be necessary, of course, to make some arrangement for the redistribution of a man's property after his death. If it seemed desirable, a society could take over the property of deceased persons, either directly or by means of heavy taxation of estates. Such a program would not be compatible with the continuation of a system of private property, however, since within a relatively few years under such a scheme all the property would have come into the hands of the state and private property would have disappeared. The maintenance of a capitalist system depends on the continuation of private property, and the continuation of private property, in turn, would be impossible if the state were to take over the property of deceased persons. An alternative, of course, would be for the state to distribute the property of deceased persons to private individuals on some basis. In practice, the state either allows the deceased to distribute the property himself in advance of death by leaving a will or, in case he dies intestate (without a will), the state determines how the property is to be distributed again to private parties. Again this is done because it is thought to be socially advantageous. It is supposed that the right to leave property by will stimulates productivity, that

men would not be willing to work so hard or so long or to be so frugal if they could not be assured that their property would be used in accordance with their desires.

The evils of inheritance are evident. In fact they are connected with one of the worst evils of the capitalist system; namely, the inequality in the distribution of wealth. These inequalities are established under a system of private property, but they are then perpetuated and passed along from one generation to another under the system of inheritance. Yet, despite this obvious and admitted disadvantage, those who favor the capitalistic system feel that the advantages outweigh the disadvantages, that the eradication of inheritance would bring more new evils than it would displace.[1] It is often true in practical affairs, as it is here, that it is not enough to find fault with an existing system. One must suggest an alternative which would work better.

36·7. Profit Motive. The capitalist system attempts to persuade people to do those things which other people want done by rewarding them rather than by coercing them. It tries to give men as great a measure of personal freedom as possible, but to make it worth their while to produce those goods which their fellows desire. In conformity with this general philosophy of allowing the maximum of personal freedom, the capitalist system includes the principles of *freedom of enterprise* and *freedom of contract.*

A system of free enterprise is one in which men are free to enter those occupations which they choose. To be sure, such freedom is not complete in our society. We impose limitations on the practice of medicine, law, and many other occupations. Even in these cases, anyone who can meet the qualifications is allowed to enter the occupation. We have no such caste system as exists in some parts of the world, under which the right to enter a business depends on the accident of birth. Anyone in our society who wishes to practice law or to become a butcher or an electrician can set out to acquire the necessary training and experience and, if he has the ability and the resources to carry him through, he is then allowed to follow the vocation of his choice. A society might be organized on the basis of making these decisions for men, deciding how many people should be trained in each line and which individuals should get which type of training. Under a system of free enterprise, the individual is allowed to make his own choice, but his choice is influenced and directed by the opportunities for making profit in various fields. If consumers want more of a given commodity than is being

[1] Perhaps it should be said again, as was stated in Chap. 4, that there is no reason to suppose that inequality is itself undesirable, but merely that we have too much of it. It would be difficult to justify complete equality on any ethical grounds, yet it is certainly impossible to justify the degree of inequality that exists in most nations today.

produced, their demand for the scarce commodity raises its price and makes the business more profitable, thus making the business more attractive to prospective entrepreneurs.

A system of free enterprise is usually combined with a system of free contract. *A contract is an agreement between two or more competent parties to do or not to do some specific lawful thing for a lawful consideration.* If Turner agrees to do something for Hoffman, but Hoffman does not promise to do something for Turner in return, there is no contract because there is no consideration. A contract must always be two-sided. Each party must promise to contribute something of value. The parties to the contract must also be competent, of proper age, sound of mind, etc. Subject to such restrictions, however, individuals in a capitalist society are left free to enter into contracts on such lawful matters as they wish and (this is the important part) having entered into such an agreement voluntarily, they will be forced by society to carry out the terms of the agreement.

Just as the concept of property implies much more than mere possession of goods, including social protection of the right to enjoy and use the goods, so also the concept of contract implies much more than an agreement between individuals. An agreement can be entered into voluntarily, but in the absence of social restriction any party to the agreement could also change his mind and withdraw at will. This is not the case with a contract, since, just as is true in the case of property, society is a party to every contract. When men have voluntarily entered into a contractual relationship, they can withdraw only by common consent and, if one tries to withdraw without carrying out his part of the agreement, society will force him to do what he promised.

Modern society depends to a very great extent on contracts. Purchases and sales of goods usually involve contractual relationships, as do the borrowing of money, the shipping of goods, the hiring of labor, the renting of durable property, etc. Division of labor could not be effected without either the use of enforceable contracts or the loss of individual freedom coupled with the imposition of duties under the force of the state. In a capitalist society, men are for the most part left free to enter into such contracts as they wish but, having entered into them, they lose their freedom to the extent of being forced to carry out the agreement that they voluntarily made.

A society that used coercion instead of the profit motive would not need to establish such institutions as those of free enterprise and free contract. If men's economic lives are to be directed by higher authority, they cannot of course choose their vocations and neither can they enter into voluntary contracts. The profit system and the institutions of free enterprise and contract that accompany it are systems devised to allow men as much individual liberty as possible.

36·8. Capitalism and Competition. In every society, human wants are many and productive factors are scarce. It is impossible to satisfy all the

wants of all the people, and decisions must be made as to which wants will take precedence, which will be satisfied and which will not. This involves decisions as to the allocation of the scarce factors of production. How much of the limited man power and land and equipment will be used to produce automobiles, how much to raise tobacco, how much to build highways, and how much to manufacture poker chips? If all wants could be satisfied, such problems would not arise, but in a world of scarcity every decision to produce one commodity is a decision not to produce others. Every time we put man power and equipment into the manufacture of poker chips we have that much less man power and equipment for use in the production of highways, tobacco, and automobiles. As we noted long since, *economic problems always involve the making of choices between desired ends.*

Decisions as to the allotment of productive factors to various industries can be made, and in some societies are made, by government authority. In a capitalist society, on the other hand, they are again left to free individual decision. Under a price and profit system[1] the productive factors are owned by private individuals who are free to use them in any industry that appeals to them. This merely means, of course, that we have private property and free enterprise. The interplay of prices and profits under market conditions of supply and demand attracts productive factors to the various industries according to the demands for the products of those industries.

How can we be sure under such circumstances that we shall get a wise distribution of the factors of production—that they will be put where they are most needed? The answer is that a capitalist society attempts to assure the proper distribution by means of competition between producers and competition between consumers. Insofar as this competition is effective, the owner of any factor of production will put that factor to the use that promises the greatest income, and that use will promise the greatest income which reflects the greatest consumer demand. If the competition is complete and effective, each unit of each productive factor will be allotted to its most important use,[2] and it will be impossible to find any combination of uses of productive factors that would fill more insistent demands. Competition and

[1] As we noted in Sec. 36 • 5, the name "capitalist system" was given to our type of economy by its opponents, and the term is not particularly descriptive of its organization and operation. All societies, whether capitalist, fascist, communist, or socialist, use capital. It is difficult to know how one can find a descriptive name that really fits. Some people have called our system an "enterprise system" and some have referred to it as the system of "rugged individualism." The term "price and profit system" perhaps describes its major features as well as any. But the name "capitalist system" has become so ingrained that we shall continue to use it, hoping that the student will understand that the use of capital is by no means its distinguishing characteristic.

[2] "Most important" in the sense that it represents the particular consumer demand which offers the most for its use. No other consumer demand will offer so much. This is not necessarily the most important demand on any other grounds, such as on the grounds of moral value.

the price and profit system working together separate the more insistent from the less insistent demands, and allot the productive factors to the production of those commodities for which consumers are willing to pay the most.

In the absence of competition, the productive factors of society are not so distributed. As we have seen, under monopolistic conditions, or even under conditions of partial monopoly, there is a tendency for production in monopolized lines to be restricted below that which would prevail if there were competition, so that fewer goods are produced in the monopolized industries than would be produced if we were to satisfy the most insistent of consumer desires. Under such conditions, it is impossible to rely on freedom of contract and freedom of enterprise to make a proper distribution of the factors of production, and the government is forced to step in to influence the situation.

In a capitalist society, the efforts of the government are spent, not in directing production and consumption (since these are left to the free choice of free producers and consumers), but in maintaining a competitive situation—in discovering and eradicating monopoly. The capitalist believes that consumers know their own wants better than they can be known by any government official, and the direction of the economy is left in the hands of consumers, who, in turn, are protected by governmental eradication or regulation of monopoly. *Monopoly and capitalism are incompatible,* and the superficial student who has gained the impression after listening to the stump speeches of "radicals" that the capitalist is one who favors "big business" and "entrenched power" has been sadly misled. The capitalist is an individualist who stresses the dignity and importance of the individual common man and who fears great concentrations of power whether they be in the hands of corporations or captains of industry or governments. Public authority is used in a capitalist society to keep power units small and to retain effective competition, rather than to direct in details the economic aspects of men's lives.

36 · 9. Achievements of Capitalism. Capitalism, as has been said, has never been tried in full. Some countries, like our own, have approached it, yet even in our own country there have been monopolistic elements that have been allowed to continue, there has been government "planning" on a greater or smaller scale, the protective tariff has diminished competition as have government programs in the fields of labor and agriculture, etc. Yet, in those countries where free enterprise, freedom of contract, the profit motive, competition, and the other characteristics of capitalism have been given the greatest prominence, there have been certain results which it seems fair to ascribe to the capitalistic nature of the economy.

In the first place, it must be admitted that the capitalist economy has given a very real and very strong *incentive to productivity.* This has been true in the cases of all the factors of production. Businessmen and laborers

alike have been stimulated to unusual efforts in production. Ingenuity and invention have been encouraged. The expectation of reward seems even more effective than fear of punishment as a stimulant. The capitalist system, insofar as it works, allows a man to follow his own selfish advantage and so arranges things that his selfish advantage will coincide with social benefit.

In the second place and partly as a result of the first, capitalism has turned out to be *tremendously productive.* The few have become fabulously rich, but their very riches have multiplied the productivity, and their savings have added to the capital of society to such an extent that even the "poor" are far better off than the average people of most noncapitalist societies. The output per man has risen far above that in other societies, and the increase is still continuing. Those who speak of capitalism as "decadent" tend to overlook both the fact that productive efficiency has continued to increase and the fact that the major difficulties in which our society found itself in the depression of the thirties seem to have been even more due to departures from true capitalism (interference with trade by the tariff and other means, monopoly elements in production, unwise government attempts to improve the situation which actually turned out to be interferences, etc.) than to the capitalistic nature of our society itself. Recurrent business cycles seem to be more a product of industrialization than of capitalism per se. There is still no indication that socialism or Russian communism will escape them.

In the third place, and again largely the result of the others, capitalism has given rise to a *higher level of living* than ever before known. The people of capitalist countries have had more to consume and have lived better, as far as the satisfaction of their material wants is concerned, than the people of any other system. To be sure, one would be foolish to assert that the satisfaction of material wants is the sole end of life, and throughout this volume emphasis has been given over and over again to the fact that other aims are important; yet it must be admitted that most people would prefer that society, other things equal, which goes furthest in satisfying their desires; and any system that makes it easier to satisfy those desires sets men free for a fuller realization of those cultural aspects of civilization which are of as great or even greater importance.

In the fourth place, capitalism has given men more *individual freedom* than has any other system yet tried. To be sure, freedom is always relative. We sometimes distinguish positive freedom and negative freedom. Negative freedom is mere absence of restraint, while positive freedom gives men power to do what they wish. If a country does not forbid flying, its citizens are free in a negative sense to take to the air; yet before the invention of the airplane this freedom never got them anywhere. The invention of the airplane, however, gave them positive freedom and made it possible for them actually to satisfy their desires.

It has been argued by many opponents of capitalism that its freedom

is more apparent than real, that it is illusory, that it is strictly a negative freedom and not a positive one. It has been argued that capitalism gives men the "freedom" to work for low wages in sweatshops under unsanitary conditions—to line the pockets of an unscrupulous boss while they starve. Yet the incontrovertible answer to such a suggestion is that the common working people of the capitalist countries have actually enjoyed a higher level of living than have other people and that the freedom has not been an empty phrase. It must also be admitted that most of us would prefer a system that does give each of us the freedom to choose his own job and to make his own destiny, realizing, of course, that we are all subject to the limitations of our own training and ability and of the type of world in which we live.

As a matter of fact, it is no matter of chance that capitalism and political democracy have tended to develop side by side, while other forms of economic organization have usually been associated with political despotism. In a political democracy, the single individual is important, and the attempt is made to abolish special powers and privileges of special groups or individuals. In a capitalist economy, it is also true that the attempt is made to distribute the power to individuals rather than to larger groups, and to put the economic system out of the reach of special interest classes (whether capital, labor, landowner, farmer, or other) and into the hands of the common man. *Democracy and capitalism both depend for their successful operation on a breakdown of power units to the point where no one unit is particularly potent or can have very much effect on the outcome.*

36·10. Faults of Capitalism. Yet no man-made institution or group of institutions can be expected to be perfect and, despite its tremendous advantages, the capitalist system surely has much that can be listed on the liability side of the ledger.

In the first place, and probably the most telling of all the counts against capitalism, is the fact that this method of organizing society has led to unforgivable *inequalities and inequities in the distribution of wealth and income.* This is a matter that we have discussed often before, especially in Chap. 4. To be sure, the defender of capitalism can truthfully and properly point out that some of the most flagrant cases of inequity can be ascribed to the lack of capitalism rather than to the failure of capitalism to work. Thus, the huge incomes have often arisen in cases where competition has not been maintained, where there has been no real freedom of enterprise, or where the government has, directly or indirectly, thrown special favors to its favorites. But this excuse, valid as it is in many cases, is by no means the whole story.

Capitalism would be expected on a priori grounds to make for inequalities in incomes reflecting inequalities in abilities, industry, and ingenuity. Most people do not object seriously to inequalities as long as they go only this

far. But a capitalism that includes inheritance, as capitalism always does for reasons already mentioned, tends to perpetuate and increase these inequalities in cases where there is no special ability, industry, or ingenuity to compensate. Neither is it a satisfactory answer to point out, as can truthfully be done, that many (perhaps it would be fair to say most) noncapitalist societies have in fact given rise to even greater inequalities in the distribution of wealth and income than those which have arisen in the capitalist countries. Those cultures which have rested on slavery or on hereditary caste systems have, for example, been even more fruitful of inequality than has our own culture. While this in incontrovertibly true, nevertheless most enthusiastic supporters of capitalism would be forced to admit that its achievement in this direction is a poor one, that we should be able to do far better, that we must somehow learn to do far better if capitalism is to achieve anything like its full possibilities. Progressive income taxation, heavy inheritance taxation, and other measures are already aimed in this direction, but some more effective and automatic method must sooner or letter be devised to meet the problem.

In the second place, it is argued that capitalism tends to build *false standards* into men's minds, that it tends to make them materialistic, that it brings with it spiritual losses which offset or largely offset any material gains. Capitalism is said to depend on and to develop individual selfishness rather than community spirit and understanding and social consciousness.

It is hard to know to what extent, if at all, these criticisms are justified. Certain it is that many (perhaps most) of the people of our society are seeking to serve their own advantage—that they are self-centered. Yet there is little reason to suppose that the situation in this regard is startlingly different in noncapitalist societies. Unfortunate as it is that our society includes a great deal of ugliness and that our people display little in the way of artistic taste or social consciousness, it must in all fairness be conceded that these facts depend more on the nature of man than on the nature of the society in which he lives. Throughout history, in every sort of society, men's efforts and attention have been concentrated primarily on economic problems—on the hard, long, and often distasteful task of wresting a living from a recalcitrant and unwilling environment. The time that the average man has had to give to the "higher" things of life has always been pitifully small. Actually, the unmatched productivity of the capitalist society has made it possible for men to give a larger proportion of their attention to these things than ever before and, if they fail to make the most of this opportunity, as some people think they do, the responsibility for their failure can surely be ascribed in part, but only in part, to the system of economic organization under which they live.

In the third place, capitalism is often accused of giving rise to many *wastes*. Human effort and precious resources are used to produce liquor and

tobacco while children need milk. The annual expenditure for chewing gum far exceeds the amount necessary to supply spectacles to all school children who need them. Trivial wants are satisfied, and millions of dollars are poured into advertising and other forms of purely competitive effort, while socially important problems are left untouched.

This argument is a very common one and is urged vigorously and with complete sincerity by people who would be shocked to have it said that they set themselves up as potential dictators over other men's lives. Yet, the essence of this argument is that the objector to the "wastes" of capitalism knows what the "right" things are for men to want, that tobacco, liquor, and chewing gum are really not the right things for men to want, and that milk and spectacles are inherently "better." Not only does the objector feel that milk and spectacles are better for him, but he is quick to decide that he knows what is good for other people better than they know themselves. The capitalist system does, insofar as it is allowed to operate, put the resources of society into the production of those things which men demand most insistently; and alteration of such a disposal of resources can logically be made only by someone who feels either that he knows men's wants better than they know them or that their wants are less important than his own objectives.

The preceding paragraph should not be interpreted as suggesting that some men are not really better judges of what is "good" or "proper" than others. Quite possibly there are criteria on the basis of which some human wants are to be preferred to others. These criteria, however, are outside the realm of science, and cannot be demonstrated on any objective ground to be valid.[1] If the user of tobacco or liquor insists that his wants are just as "good," there is no way to prove him wrong. The person who raises this particular objection to the capitalist system should be put in the position of admitting that he feels that men should not be allowed to make their own decisions on these matters and that he has some superior source of wisdom, and the matter should then be argued on that basis.

Very closely allied to this argument is a fourth objection to capitalism; namely, that *it puts the direction of the economic system in the wrong hands.* Instead of putting the direction of production into the hands of highly skilled, well-trained, competent technicians and experts, it leaves control in the hands of ignorant consumers. This is the argument often advanced by the less philosophical of engineers and physical scientists, who have failed to see the actual position of science and technology in human life, or by people who have failed to understand how a capitalist system actually operates.

A capitalist system does adopt the most efficient techniques for reaching its objectives. Any set of human wants being given, it gives preference to

[1] This point is discussed at greater length in Chap. 2.

those production methods which will yield the greatest product from given resources. Whether a given machine or organization is efficient or not is not purely (nor even primarily) a problem in physics or engineering or technology. One gasoline engine may pull a far greater load than a second, while the second succeeds merely in giving off large volumes of fumes with little power. Both use the same amount of gasoline with very different results. Which is more efficient? The question cannot be answered in mechanical terms, since the answer depends primarily on human motives and desires—on an evaluation of objectives. If the engine is wanted as a source of motive power for an automobile we may well prefer the first; if it is wanted to smoke out woodchucks, we may prefer the second. One is more efficient for one purpose, and the other for another. *We can judge efficiency only in terms of men's objectives.*

When men's objectives are given, the technician and the scientist usually know far better than anyone else how to go about reaching the objectives efficiently; but the technician and the scientist are no better qualified than anyone else to decide what the objectives should be. Neither, of course, does the economist have any special aptitude in making such decisions. The capitalist system leaves the decision of this question in the hands of the individual consumer. He is given freedom to decide for himself what objectives are important, and to make his wants known in line with his objectives by his demand for goods on the market. The capitalist system then tends to organize the most efficient means of meeting these demands, using those methods which the scientist and the technician find best adapted. Capitalism, as a matter of fact, makes an important distinction between means and ends, which is not made with sufficient care by those who criticize capitalism on this particular ground. Capitalism leaves decisions as to ends (objectives, goals) in the hands of consumers and puts decisions as to means (methods, techniques) in the hands of competent experts.

A fifth objection to capitalism, which is somewhat at variance with some of the others, is that *capitalism lends itself to control by special interest groups.* This criticism has a good deal of merit and can be illustrated over and over again from widely different segments of the economic (and political) field. Capitalism depends for its successful operation on keeping power units small. If big business interests get big enough, it is natural for them to use their power for their own advantage in seeking tariff protection, in charging exhorbitant prices, in economic imperialism, etc. These activities of big business have been widely and properly condemned. Other pressure groups, as we have seen, carry on the same sort of self-centered activities and as before, if they grow to the point where they can exert enough pressure, they become dangerous. Organized agriculture, the "silver bloc," organized labor, manufacturers' associations, chambers of commerce, and other such

organizations are inherently anticapitalistic. As long as they stay small and are kept powerless, they can present no real challenge to capitalism but as soon as they grow to the point where they exert effective pressure, they tend to defeat the major purpose of a capitalist society. Perhaps the very multiplication of such organizations is the only effective safeguard for capitalism since, if the number of such organizations is sufficiently great, no one of them can wield much power, and competition becomes competition between pressure groups rather than between individuals. There is no guarantee under such circumstances, however, that the decisions that are made will be in the interest of consumers.

A sixth objection to capitalism is that *it makes for insecurity,* that unemployment, and especially cyclical unemployment, is a unique characteristic of the capitalist economy. Of course, no society has ever offered its members complete security in the economic field. The fear of starvation is an ever-present fear in primitive societies, and hunger and privation are felt to some degree in every society. If security implies certainty that one's wants will be satisfied, then security is a figment of the imagination. On the other hand, most of those who make this criticism of capitalism presumably mean to indicate that the ordinary worker is less secure under this form of social organization than he would be under alternative forms. The business cycle, at least in its present form, is thought by many students to be peculiar to capitalism, as we noted in Chap. 22.

It is difficult, in the face of present inadequate knowledge as to the cause of business cycles, to know how to treat this objection. The people who make the statement have no way of knowing that it is true, and their opponents have no way of knowing that it is false. In the present state of human knowledge it is on approximately the same plane as an argument about the appearance of the far side of the moon. In all fairness it must be said that recurrent depressions have tended to become more severe during the past 50 to 100 years, and that during at least part of that time the capitalist system has come as near to adoption in leading commercial nations as it has ever been. On the other hand, it must be remembered that many students of business phenomena believe that the most recent of these depressions was caused more by departures from true capitalism (such as monopolistic practices, trade restrictions, etc.) than by any peculiarities of capitalism itself. This particular argument must be set down as unsettled either way, and it must be kept in mind along with it that, even in times of depression, the people of capitalist countries have lived at a higher level than have those of most other societies even at their best. Much of the suffering of capitalist depression comes, not from starvation, but from temporary inability to get things to which the people have become accustomed in time of plenty but which the people of other societies have never had a chance to enjoy. This,

to be sure, does not make the privation any less real nor any easier to endure.

Next, there are those who object to capitalism on the ground that *it leads to war and imperialism.* It is commonplace to hear people speak of wars as evidences of capitalist aggression and as being fought for the benefit of a few war profiteers and capitalists. The answer in this case is so transparently evident that one hesitates to give it. War antedates capitalism by centuries. Wars have not been confined to the people of any particular type of social organization but have been waged by capitalist, fascist, and communist nations alike. As a matter of fact, war seems to be more closely tied up with nationalism and with great power of the state than with the economic system, and most forms of economic organization tend to glorify and magnify the state far beyond anything known to capitalist societies.

Finally, there are those who object to the capitalist system because it bases the direction of economic activity, not on the importance of men's wants, but on their *ability to pay.* The rich man's ten-dollar bill spent for a nonessential has far more influence on the productive system than does the poor man's half dollar spent for badly needed food. It is not human wants or human needs that direct the capitalist market, but dollars and cents; and these dollars and cents are very unequally distributed. The man with an income of a million dollars a year carries a thousand times as much weight in the making of economic decisions as does one with a $1000 a year income.

One's attitude toward this argument will depend in large measure on his philosophy of life. The argument obviously ties back to the inequality in the distribution of wealth and income, which we have discussed several times and which was last taken up for treatment in the second and third paragraphs of this section. Insofar as the inequalities of income and wealth represent differences in contribution to society (as to some extent of course they do), few people would object to them. Under such circumstances there is some merit in allowing some consumers to have more influence on production than others. The doctrine of absolute equality in consumption is by no means demonstrably valid, especially when it is recalled that incomes serve the twofold purpose of rewarding past productivity and of giving incentive to further productivity. If incomes were satisfactorily distributed, this particular objection to capitalism would carry far less weight. In other words, this argument amounts to little more than further criticism of the distribution of wealth and income, which we have discussed before. The validity of such criticism has already been acknowledged. If critics mean to imply more by the argument than that capitalism directs production on the basis of dollars rather than on the basis of needs, then they are on considerably weaker ground.

36·11. Alternative Forms of Social Organization. Having listed very briefly some of the major advantages and disadvantages of capitalism, we

shall turn to an even briefer description of some of the alternative systems of social organization that have been proposed. Here, as with capitalism, there is the difficulty that there is little uniformity in the use of terms. Capitalism means one thing to a Republican and another to a Democrat, one thing to a farmer and another to a factory worker, one thing to the secretary of the local chamber of commerce and another thing to the president of a labor union, one thing to an American and another thing to a Russian. Similarly, such words as communism, socialism, fascism, and anarchism carry decidedly different connotations in the minds of different users. The student need not be surprised if the usage adopted here differs somewhat from that to which he has been accustomed, although we shall try as far as possible to adopt the usage that is most widely accepted in academic circles today.

Oscar Jaszi has suggested[1] that the various "reform" movements such as communism, socialism, anarchism, and collectivism share six common characteristic ideas:

1. Condemnation of the existing social and political order as unjust.
2. Advocacy of a new order consistent with moral values.
3. A belief that such a new order can be realized in practice.
4. A belief that the evils of the existing order depend on human institutions rather than on the nature of man or of his environment.
5. A program of action for establishing the new order.
6. A revolutionary will to carry out this program.

Yet the different programs differ widely among themselves in the emphasis that they give to these various points.

36·12. Socialism and Communism. Even those who call themselves socialists and communists differ sharply in the use of the terms. Today, despite variations in meaning in the past, it is common to think of them as being in most essentials similar, but differing in degree. For this reason we shall treat them together despite the fact that according to some uses the two are in opposition.

Socialism and communism agree in their belief that private ownership must give way in greater or lesser degree to collective or public ownership. The less extreme socialists would limit collectivization to the public utilities and perhaps to such other basic industries as mining. The more extreme communists would insist that all production goods—all industry—pass under public control. And to the extent that production becomes a public function, the socialist and the communist believe in government planning in place of the allegedly "automatic" consumer control of the capitalist system. The position and power of the government would be greatly increased under

[1] In his article on socialism in the "Encyclopedia of Social Sciences," vol. 14, p. 188, The Macmillan Company, New York, 1934.

either socialism or communism. Some schools of socialist thought favor an intermediate stage in which production is carried on, not by the government, but by cooperating groups of consumers, producing "for use rather than for profit." All groups agree that the profit motive must be discarded and that the state must assume a greater degree of supervision and direction of production. Pure communism involves common ownership of all goods, production goods and consumption goods alike, with the public ownership of production goods as an intermediate step.

Not only are there differences of opinion as to how far collectivism should go and how great the power of the state should become, but also there are decided differences of opinion as to how the new goal should be realized. Again, common usage tends to ascribe the more radical measures to communism and to call the more temperate measures "socialistic." Programs vary from those of Fabian socialism, which favors a slow and orderly evolutionary change through education, persuasion, and the ballot box, to those of revolutionary communism which favor the forceful overthrow of the existing system and the establishment of a dictatorship of the proletariat by violence.

Most socialists and communists would expect, if they were to organize society, to continue the use of a wage system. Some of the more extreme have urged absolute equality of wages, but for the most part they favor some differences in wages. In Russia, for example,[1] it soon became evident that equality of wages did not furnish sufficient incentive for men to put their best efforts into production, and a system of differing wages was quickly established. Yet the socialist and the communist would insist that their wage system was not merely the same old wage system found in capitalist countries. In the first place, it would not be tied in with private industry operating for a profit, so that wages would not be determined on a profit-making basis; in the second place, although wages would differ somewhat, all socialists and communists would agree in insisting that the degree of inequality would be much smaller than that of competitive capitalism. Labor unions would be allowed to exist—in fact, they would be encouraged and membership in them might even be compulsory—but they would lose the bargaining functions which are their most characteristic feature in a capitalist society, since such economic decisions as the determination

[1] It is no more fair to think of Russia as an example of pure communism than it is to think of the United States as an example of pure capitalism. The economic system of the United States between the Civil War and the first World War perhaps came closer to capitalism than any other large-scale experiment, but even it differed from pure capitalism in many ways. The present Russian system comes closer to communism than any other large-scale experiment, but it, too, differs from the pure type in a number of respects. In neither case can every good or bad feature of the country's life be ascribed to the capitalism or the communism that it is supposed to represent.

of wages would be made by the government rather than by free contract between employer and employee.

Most socialists and communists believe that capitalism bears within it the seeds of its own destruction, that capitalist society is decaying, and that its ultimate destruction is certain. They differ in their opinions as to whether it is best to wait for nature to take its course, allowing the slow concentration of riches and power on the one hand and the increasing misery of the masses on the other hand to make the wisdom of fundamental reform obvious to everyone and a matter of common consent, or whether it would be better to undergo the admitted agony and loss of violent revolution immediately in order to reach the goal sooner and to escape the alternative miseries that they foresee if things are allowed to drift. Karl Marx, perhaps the leading socialist or communist writer of the nineteenth century, and many of his followers, believe that there is a fundamental cleavage of interest between two great classes in society, the *bourgeoisie* and the *proletariat.*[1] Between these two there always has been and continues to be, according to Marxian philosophy, an inescapable class struggle which will continue unless and until the entire economic basis of society is altered. The more moderate of modern socialists reject this doctrine of class struggle and recommend cooperation with capitalist governments in a more or less opportunistic program of piecemeal reform and peaceful change; the more extreme of modern communists feel that the doctrine of class struggle is of basic and fundamental importance and that there can be no cooperation between two philosophies that are in such extreme opposition.

Many people are to some extent sympathetic with the aims of socialism and communism because it appears to them that these philosophies contain a large element of sympathy for the underprivileged and the downtrodden, that their adherents recognize the problems of the poor and propose to do something about it. Sympathy and good will and altruism are, however, no more characteristic of these philosophies than of any other. The real issue, of course, is what sort of social organization would actually come closest to reaching whatever goals we set up for ourselves, whether these be stated in terms of physical productivity, of equality of status and income, of human happiness, or of spiritual welfare. The student will be more understanding in his own study of alternative proposals if he realizes that most of the adherents of any one of the plans believe sincerely that the adoption of their proposals would make the world a better place in which to live.

[1] The *bourgeoisie* are the "exploiting" or ruling class, the "capitalists," the owners of the means of production, the employers of labor. The *proletariat* are the "oppressed" or "exploited" class, the wage earners who, having no means of production of their own, are reduced to selling their labor-power in order to live. These definitions are those of Frederich Engels in his preface and footnotes to the "Communist Manifesto" which he wrote with Karl Marx.

Opponents of socialism and communism argue that these systems would reduce those personal liberties which have been won at great sacrifices and bring about a return to an all-powerful state. They feel that past experience demonstrates the futility of assuming that such power would in practice be used for social advantage and general welfare rather than for personal and political ends—for the entrenchment and perpetuation of the power of individuals or small groups. They argue that, even if the government were honestly and completely disinterested and socially minded, there is no reason to suppose that government officials are more competent to direct the economic machinery than are consumers; and they believe that the bureaucracy of government is apt in practice to make for far greater inefficiency than that which arises under capitalism where the spur of the profit motive is an ever-present incentive to perfect and adopt more efficient methods. They foresee very great difficulties in accumulating sufficient capital for efficient production in a socialist state, although some system of taxation might be politically possible which would allow for such accumulation. They point to the fact that socialists and communists express the hope that hours of work can be materially reduced and workingmen may be given more leisure. It is their fear that such reduction of hours, especially when coupled with the absence of the profit motive, would result in a material reduction in output and a consequent reduction in the level of living. Despite its disadvantages, capitalism has succeeded in developing a system of incentives that has led men to work hard, to risk much, and to exert their best efforts in production. One of the greatest tasks of socialism is to find some acceptable and practical substitute. To some extent, social recognition and acclaim may furnish a spur to production, but the opponents of socialism doubt that it will be so effective as the recognition that appears in the pay envelope.

36·13. Anarchism. The untutored student who gets his impressions from the radio news releases or from the headlines of the tabloid press commonly thinks of socialism, communism, and anarchism as three names for the same or approximately the same thing. True enough, the terms "socialism" and "communism" have been used almost interchangeably; but anarchism, far from being a minor variant of the others, represents almost the exact antithesis. Socialism and communism magnify the position and importance of the state, relying on government action to overcome the evils of capitalism. Anarchism, on the other hand, proposes the complete abolition of government of every sort, relying on the spontaneous cooperation of men to establish justice in human relations. The anarchist does not favor anarchy in the sense of complete confusion and lawlessness. On the contrary, he believes that men are essentially good and that, without the corruption of the state, they would tend to live together in peace and harmony. Under a system of anarchy, the importance of the individual would be

greatly magnified. Law and regulation would disappear as we now know them, and each man would be a law unto himself. Private property (which, as we have seen, depends for its very existence on government recognition and enforcement) would disappear and men would cooperate freely and spontaneously in the use of the good things of life.

Such a proposal will seem to the student idealistic. It is. This is a fact in its favor rather than an argument against it. Society can be improved only by men with ideals. But, to most people, anarchy seems not only idealistic but impractical—unworkable. Certainly it would be necessary to have a long period of education before men could safely trust themselves to such a system.

Anarchism is associated in the minds of most people with extreme violence. As a matter of fact, its advocates differ among themselves as to the means by which they would reach their goal. The revolutionary anarchists have favored violence as the only possible means to their end but have urged that the overthrow of the state must not bring about the establishment of any new and comparable power in its place. It must bring about the complete abolition of all units of social power. The philosophical anarchists, on the other hand, have felt that it was wise to rely on education and nonresistance as the best means to their end. Since true anarchists object to all compulsion, it is natural that they should hesitate to establish their system by compulsory methods. In fact, anarchists are essentially pacifists.

36·14. Syndicalism. Syndicalism arose in France as a reaction to socialism and communism. It agrees with them and with anarchism in its opposition to the capitalist organization of society, but it differs sharply from socialism and communism and agrees with anarchism in the belief that the state should be dispensed with. It would substitute direct economic action for political action in the attempt to reform society. It therefore favors the general strike, sabotage, and other such direct methods in an effort to obtain the ownership and control of industry, not by the state, but by the workers themselves. It envisages an economic society made up of autonomous and self-governing industries. These independent industries would enter into various sorts of agreements with one another for mutual benefit, but they would be under no compulsion to adopt any particular sort of agreement, since there would be no superior agency to force the acceptance of such agreements.

In order to imagine what a syndicalist society would be like the student may try to picture a United States (not a government, of course, but a locality or geographical territory) within which the farms are owned by the farmers who actually operate them (no tenant farming), the railroads are owned and operated by the railway "Brotherhoods," the textile factories are in the hands of the United Textile Workers Union, the coal and iron

mines are in the hands of the United Mine Workers, etc. In a syndicalist economy the labor unions would play the major role, and a general strike of existing unions is often looked upon as the proper method for the establishment of the new order. Within each industry the union would determine all policies of employment, wages, price, output, etc.; and between industries there would be some nebulous sort of ill-defined cooperation which was not enforced by any superior power but which unified the economic program of the society as a whole. Thus, syndicalism lies in intermediate ground, sharing some of the characteristics and beliefs of socialism and communism with some of anarchism.

36 · 15. Fascism. Fascism is a curious combination of elements from the proposals for social reform listed above. It has used the title "National Socialism," but is socialistic only in its emphasis on the power of the state. It has been intensely nationalistic in viewpoint and has tended to rely on a very highly centralized political dictatorship in much the same way that socialism and communism have tended toward political absolutism. In this respect prewar Russia and prewar Germany or Italy were much alike. Fascism, however, advocates a continuation of private property and the private control of industry, with, to be sure, an increased regulation by the state for nationalistic ends. It does not rely, like capitalism, on control by individuals acting in their own interests. Rather, it relies on control by great interest groups, such as a "council of corporations." In fact, the "corporative state" of fascism was thought of as a visible evidence of the unification of the state and industry. Fascism asserts that the state is supreme and that it has unlimited sovereignty over the individual and over all aspects of human life. The state is thought of as having purposes separate from those of its citizens, and the interests of citizens are subordinated to the national plan. Fascism is nationalism at it peak, combined with a strange admixture of syndicalism which bases social organization on industrial representation. It retains the private ownership of production goods which is a hallmark of capitalism, but in contrast with capitalism it subordinates the advantage of the individual to national and racial objectives. Communists and socialists have decried it as a desperate attempt to save a failing capitalism; capitalists have been prone to point out the many essential similarities in practice between fascism and communism.

36 · 16. Conclusion. Society is constantly changing, and with the changes our own concepts change. What is called "socialistic" today is not exactly the same as what was called "socialistic" a generation ago. No society, in practice, furnishes an unqualified example of any of the ideologies that we have just been describing. Each is a mixture. The public school system of the United States combines government ownership and control, support by taxation and equality in use. It is "socialistic" in nature. Yet the differences

described in this chapter are by no means unimportant. It does make a difference to men how their society is organized.

Changes in technology have made it advantageous to carry on production on an ever larger scale. Individualism in production has given way to highly organized mass effort. Society can give up this production system only at the expense of great loss of product. Hardly anyone seriously suggests that this should be done. If it is not to be done, society is faced with the necessity of protecting itself against the great power of these production units, and of finding means to make sure that they do not become more powerful than is necessary in the interest of efficiency. This may take the form of increasing the power of the government, or of reducing the power of industrial corporations, or of some entirely different method. Certain it is that we can expect continued experimentation, and it is surely the part of wisdom for us to welcome experiments by other people and to watch their outcome with interest. The trial of new methods in Russia, Germany, or France or by the "labor government" of England gives us an opportunity to observe without risking much ourselves, and only the foolish will condemn such experiments in advance.

Yet, our own system is not so bad that we can afford to scrap it for another, sight unseen. Our system has grown slowly and has been built at great sacrifice. It has serious weaknesses, which its own best friends are the most anxious to seek out and repair; yet the high level of living and the unparalleled individual freedom that it has brought are not to be discarded lightly. It has not worked perfectly nor has its working been close to perfection, and its inequities are obvious to every serious student; yet these inequities must not be allowed to blind us to the fact that all substitutes that have been suggested carry their own difficulties with them, and that many of the economic problems under which men chafe arise, not from the type of economic organization under which we live, but from the nature of man and of his environment.

Whether we live under capitalism or under socialism, communism, fascism, syndicalism, or some other sort of *ism,* it seems likely that we shall continue to find men wanting things—yearning always for more than they have. The variety of human wants is an inherent characteristic of human beings. Similarly, communists and fascists and socialists and capitalists and syndicalists and all other sorts of *ists* live in a world of limited resources, faced with the niggardliness of nature and, as a result, find themselves unable to satisfy all the wants of their members. The variety of wants and scarcity are, as we noted in our very first chapter, characteristics of the world in which we live. Want and privation are not the monopoly of any society, nor does any society escape them.

Similarly, the student will note in retrospect that most of the laws and

principles that we have taken up for study depend on the nature of man and of his environment rather than on the organization of his society. Diminishing productivity occurs in any type of society. So does diminishing utility. Where men are left free to make choices, they tend to select those alternatives which seem to them most advantageous and to give up those things which involve the least sacrifice. Yet this is but another way of stating Gresham's Law, which is by no means confined to the field of money.

Despite these fundamental and important similarities, it is important to repeat, however, that the method of organizing a society does make a difference. It affects the lives, the happiness, and the fortunes of all those who live within it. Although it is perhaps impractically Utopian to dream of perfection, nevertheless the quest for improvement in social organization is one of the most important and promising quests that men have undertaken. The quest has already carried men far, but the goal is by no means in sight, and the very nature of the task makes it possible, perhaps probable, that there is no real end other than the quest itself.

Questions

1. List some things that men want without any regard to economic objectives.
2. Do any of the things that you listed in question 1 have indirect economic repercussions or aspects? Explain.
3. When a group of fraternities combine in a political deal in a campus election, are there any direct or indirect economic motives involved?
4. To what extent do you think men would be gregarious if they had no economic motive for living in groups?
5. How does a capitalist society stimulate production? Be specific. If we want more wheat, exactly how do we go about getting it?
6. Who are the "we" referred to in question 5?
7. How does a socialistic or a communistic society stimulate the production of more wheat if they want it? Who are the "they" referred to in the preceding sentence?
8. What do you anticipate would happen if inheritance were forbidden in the United States and if all the property of deceased persons were sold at public auction and the proceeds used for the support of the government? Note that this question assumes the continuation of private ownership, but the elimination of the power to control the disposition of one's property after death.
9. What specific injustices arise because of the inheritance of property? What specific injustices would arise if inheritances were entirely eliminated?
10. If a social system were set up that guaranteed a completely equal income to everyone regardless of productivity, would you expect the total product of society to fall? If

so, would you expect the reduction to be greatest in the case of manual laborers, skilled laborers, business managers, or professional people? Why?

11. What other incentives to production are possible in a society to take the place of the profit motive?

12. If a boy promises his chum to take him to the movies, does the promise constitute a contract? Why or why not?

13. What is the difference between the statements that "complete competition gives men what they want most" and that "complete competition gives men what they will pay most to get"? Would not the two be exactly the same in practice? If not, why not?

14. Distinguish positive and negative freedom. Give an original example of each.

15. Distinguish as well as you can the meanings of the word communism and socialism.

16. What is anarchism?

17. What is syndicalism?

18. What is fascism?

19. What is meant by bourgeoisie and proletariat?

MATHEMATICAL APPENDIX

Introduction. Economics deals with quantitative variations and their interrelationships—quantities of utility or cost or income or profit, etc. Relationships between quantities can usually be expressed most precisely and satisfactorily in mathematical terms, and it is not surprising that many of the principles of economics can be explained most easily and most precisely in mathematical form. Yet the student should remember that, in the study of economics, mathematics is a means and not an end. No principle of economics is important *because* of its mathematical treatment or its mathematical interest. Only as the mathematical expression of the principles helps the student to understand the economics involved is that mathematical expression justified in this book. Many students, lacking formal mathematical training, will find it far easier to understand what the principles are all about by wading through the long-winded explanations in the body of the text than by attempting to read the short mathematical explanations given here. The student who finds these mathematical notes confusing or not understandable should rest assured that the principles can be understood without going further. The engineer or other student with mathematical training should, however, find his work much easier if he will follow through the simple proofs and explanations given here.

Note 1 (page 50).

Vilfredo Pareto, Italian economist and sociologist, thought that in all countries and at all times the distribution of incomes could be described by the equation:

$$y = Ax^{-\alpha}$$

where y is the number of people having income as great as x or greater, where A is a constant depending on the national income of the country in question, and where α is approximately equal to 1.5.

We can illustrate this law with data describing the distribution of incomes in the United States in 1929[1] (see Table A·1).

Our original figures on incomes and recipients are in the first two columns, and the last two columns are merely the logarithms of the first two. We use these logarithms because it is evident from the formula for the

[1] Leven, Maurice, H. G. Moulton, and Clark Warburton, "America's Capacity to Consume," p. 54, Brookings Institution, 1934.

Pareto Law given above that the law becomes linear when thrown into logarithmic form, since, if

$$y = Ax^{-\alpha}$$

then obviously,

$$\log y = \log A - \alpha \log x$$

and this is linear in $\log y$ and $\log x$. If we plot the data of the last two columns of our table, we get Fig. A·1 and see at once that the Pareto Law gives a fair description of the data. For most income data, the law gives a reasonably good description if we make two exceptions to its original statement:

1. The law breaks down at the lower incomes, as one would realize

Table A·1. United States Incomes, 1929

Income in thousands of dollars (x)	No. of families receiving this income or more, thousands (y)	$\log x$	$\log y$
$1	21,575	0.0000	4.3339
2	11,120	0.3010	4.0461
3	5,928	0.4771	3.7729
4	3,488	0.6021	3.5426
5	2,256	0.6990	3.3533
7	1,183	0.8451	3.0730
10	631	1.0000	2.8000
25	160	1.3979	2.2041
50	63	1.6990	1.7993
100	24	2.0000	1.3802

must be true when he inspects the formula, since the law would show limitless numbers of income recipients at very small incomes.

2. The coefficient, α, is not always 1.5, but varies from year to year and from country to country. In the case of Fig. A·1 the straight line drawn on the diagram was fitted by eye, and has the formula

$$\log y = 4.54 - 1.59 \log x$$

or

$$y = 35{,}000x^{-1.59}$$

In this case the value of α is approximately 1.59.

The value α is thought by some to be a measure of the degree of inequality in the distribution of income, but it has never been adopted generally for that purpose. Other economists and statisticians, following Pareto, have suggested other functions as describing the distribution of incomes, but no such function, which is entirely satisfactory, has yet been proposed.

Note 2 (page 61).

We can define our terms somewhat more precisely in mathematical form as follows:

Let U be the amount of utility that one gets from a stock of goods.

Let Q be the quantity of the good—the number of units in the stock.

Then,

$$U = f(Q)$$

This we call the "utility function." But it is of such a nature that

$$f''(Q) < 0$$

This last expression is our statement of the Law of Diminishing Utility,

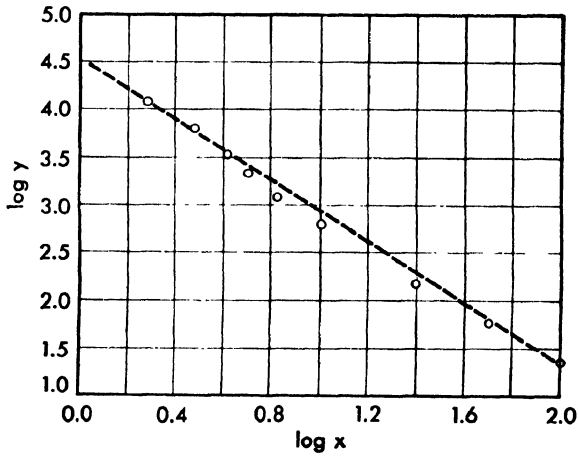


FIG. A • 1.—Illustration of Pareto's Law as applied to the distribution of incomes in the United States in 1929.

which we are able to put into verbal terms by defining marginal utility (*M.U.*) thus,

$$M.U. = \frac{dU}{dQ} = f'(Q)$$

In order to find the marginal utility from these formulas and definitions, we would have to know the actual parameters of the utility function. These we seldom know. Suppose that we did know that the utility function was

$$U = 20Q - Q^2$$

We note that the marginal utility of any stock, Q, would be

$$M.U. = \frac{dU}{dQ} = 20 - 2Q$$

We note also that:

$$\frac{d^2U}{dQ^2} = -2$$

which satisfies our requirement that the second derivative be negative, or that the marginal utility decrease as the stock increases.

Since, in this case, we actually know the utility function, we can find the amount of utility for any size of stock as in Table A·2.

Table A·2. Illustration of Diminishing Utility

Q	U	dU/dQ	$M.U.$
0	0	20	?
1	19	18	19
2	36	16	17
3	51	14	15
4	64	12	13
5	75	10	11
6	84	8	9
7	91	6	7
8	96	4	5
9	99	2	3
10	100	0	1
11	99	−2	−1

The first two columns of Table A·2 are computed from the formula $U = 20Q - Q^2$. Only positive values of Q are reasonable. We note that U rises more and more slowly, passing through a maximum at $Q = 10$.

The third column shows values of the marginal utility found by differentiating the utility function. The fourth and last column shows the values of the marginal utility found by the finite-difference method used in the body of the text. The student will note that the finite-difference method always yields somewhat larger values than does the method of infinitesimals. The reason is evident when we consider the graphic presentation. Figure A·2 represents a hypothetical utility function of the kind that we have defined. In infinitesimals we define the marginal utility of quantity OQ as the slope of the tangent CA. In finite differences, when the quantity drops from OQ to OQ', the total utility drops from QA to $Q'B$ along the secant slope BA. This finite secant slope is greater than the slope of the tangent and hence, when we define marginal utility in finite terms, we always get a slightly larger answer than when we define it in infinitesimals.

We note in our numerical example that, after we pass through the maximum point, the marginal utility is negative. For values of Q where the marginal utility is negative, we say that the good is illth. When the marginal utility is positive, we say that the good is an economic good. At the maximum point, where the marginal utility is 0, we say that the good is a free good.

The student may be bothered by having two slightly different values of marginal utility, depending on whether he uses the method of infinitesimals or that of finite differences. To be sure, if the change in Q is a small one, the slope of the secant will be nearly the same as the slope of the tangent, and the two methods will give nearly the same result. For this reason we have defined the marginal utility in the text as the difference in total utility when we subtract *one unit* from the stock. In practice, the finite-difference method probably comes closer to describing the facts than does the method of infinitesimals. Buyers cannot usually make infinitesimally small changes in the quantity purchased. They can buy 3 pounds of sugar or 2, but even with sugar they cannot buy 3.029 pounds. Surely with shoes they cannot buy 4.026 pairs, nor can they buy 1.3 pairs of sox. The consumer is forced to

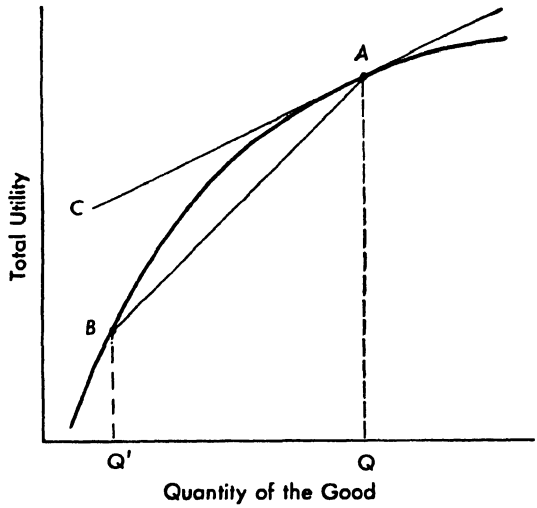


FIG. A·2.—Illustrating two approaches to the idea of marginal utility.

buy whole numbers of units. Moreover, neither the consumer nor the economics student usually knows the equation of his utility function, so he has nothing to differentiate. The only way that he can find marginal utility is to take finite differences in some hypothetical utility table such as those used in the text. Therefore, when the student is given numerical problems involving marginal utility, he will be expected to compute the marginal utilities by the finite-difference method illustrated in the last column of Table A·2 rather than by the slightly more exact method of column 3.

Note 3 (page 93).

We can define elasticity of demand more precisely in the following terms:
Let P represent the price per unit.
Let Q represent the quantity that consumers are willing to buy at price P.

Then the Law of Demand tells us that

$$Q = f(P)$$

and

$$f'(P) < 0$$

We define elasticity of demand thus,

$$\eta = \frac{dQ/Q}{dP/P} = \left(\frac{dQ}{dP}\right)\left(\frac{P}{Q}\right) = \frac{d(\log Q)}{d(\log P)}$$

Since dQ/dP is inherently negative, economists have been accustomed to taking the absolute value of η as their measure of elasticity.

Now let us see the relationship between elasticity and total consumer expenditures, discussed in Sec. 7·4. If consumers buy Q units of the product at price P, their total expenditures are obviously PQ. Our problem is, what happens to PQ as prices change?

$$d\left(\frac{PQ}{dP}\right) = Q + P\left(\frac{dQ}{dP}\right) = Q + Q\eta = Q(1 + \eta)$$

But since η is essentially negative and since economists use its absolute value rather than its actual negative value, we shall have to change the sign in the last expression and say that

$$d\left(\frac{PQ}{dP}\right) = Q(1 - |\eta|)$$

The value Q is always positive, of course. Note what happens as we change the value of η:

If $\eta = 1$, then $Q(1 - \eta) = 0$ and there is no change in consumer outlay as price changes.

If $\eta > 1$, then $Q(1 - \eta) < 0$ and consumer outlay falls as prices rise.

If $\eta < 1$, then $Q(1 - \eta) > 0$ and consumer outlay rises as prices rise.

In practice, since we do not usually know the nature of the function that connects P and Q, we have to fall back on an approximation to $(dQ/dP)(P/Q)$ by finite differences. Suppose that we take the expression dQ/Q. By finite differences, we should take the difference in Q and divide by the average value of Q, or by half the sum of the Q's. Similarly, the finite approximation to dP/P would be the difference between the prices divided by half their sum. We would thus get the approximation formula:

$$\eta = \frac{Q_1 - Q_2}{\frac{(Q_1 + Q_2)}{2}} \Big/ \frac{P_1 - P_2}{\frac{(P_1 + P_2)}{2}}$$

where Q_1 and Q_2 are the two quantities and P_1 and P_2 are the corresponding prices. If we cancel the 2's, we see that this gives us the formula used in the text, namely,

$$\eta = \left(\frac{Q_1 - Q_2}{Q_1 + Q_2}\right)\left(\frac{P_1 + P_2}{P_1 - P_2}\right)$$

Let us note that, if the total consumer outlay PQ is constant, this formula gives us unit elasticity. In this case, $P_1Q_1 = P_2Q_2$.

But, if we carry out the multiplication of our formula for elasticity we get

$$\eta = \frac{P_1Q_1 - P_1Q_2 + P_2Q_1 - P_2Q_2}{P_1Q_1 + P_1Q_2 - P_2Q_1 - P_2Q_2}$$

If $P_1Q_1 = P_2Q_2$, the first and last terms of both numerator and denominator disappear and, in this particular case where consumer outlay is constant, our formula reduces to

$$\frac{-P_1Q_2 + P_2Q_1}{P_1Q_2 - P_2Q_1} = -1$$

Changing the sign so that we have the absolute value, we see that constant consumer outlay coincides with unitary elasticity of demand

It should be evident from this formulation that the elasticity will itself usually be a function of P or Q, varying at different parts of the demand curve. The elasticity will be constant at all parts of the curve when and only when the demand function is of the general form $QP^{\eta} = k$, where η is the elasticity and k is a constant; for under these circumstances we see that

$$Q = kP^{-\eta}$$
$$\log Q = \log k - \eta \log P$$

Under these circumstances, it is evident that $d(\log Q)/d(\log P)$ is constant, which means that the elasticity of demand is constant. Or, we can see it directly by differentiating $Q = kP^{-\eta}$ with respect to P, and then evaluating our definition of elasticity, namely,

$$\eta = \left(\frac{dQ}{dP}\right)\left(\frac{P}{Q}\right)$$

which in this case $= -\eta$. Thus, if we let $QP^2 = 15{,}000$, we can work out various values of P and Q which will yield a demand curve with a constant elasticity of 2. The student will see that demand curves of constant elasticity will be curves rather than straight lines when plotted on ordinary axes. In fact, they will be curves that are concave upward, similar in this respect to those drawn in the text. It will also be evident that such demand curves will yield straight lines if plotted on logarithmic axes.

In the case of constant unit elasticity, the formula will become $QP = k$, so that we see again that with unit elasticity the total expenditure, QP, is constant. This equation describes a demand curve which is a rectangular hyperbola, often called by the economist a "constant-outlay curve." One of the characteristics of such a curve is that lines drawn through any point on it parallel to the two axes will enclose, with the two axes, exactly the same area as such lines drawn through any other point on the curve.

Linear Demand Functions. Economists often use straight-line demand

curves in their diagrams, despite the fact that they are probably uncommon or nonexistent in practice. If we write the equation of a straight-line demand curve in the form

$$Q = a(b - P) \tag{1}$$

the student will find it easy to prove that the elasticity of demand at any price, P, is

$$\eta = \frac{P}{b - P} \tag{2}$$

Such a straight-line demand curve will evidently have unit elasticity at the point where $P = b/2$. If we wish to discover the point on this straight-line demand curve that has any particular desired elasticity, η, we substitute the given elasticity in the expression

$$P = \frac{\eta b}{1 + \eta} \tag{3}$$

For example, if we wish to know where the elasticity will be equal to 3, we substitute 3 for η in equation (3) and find that this elasticity will exist at a price of $3b/4$.

If we prefer these same formulas in terms of Q rather than P, we find that the elasticity corresponding to any particular quantity is

$$\eta = \left(\frac{ab}{Q}\right) - 1 \tag{4}$$

If we want to find the quantity corresponding to any particular elasticity, we find that

$$Q = \frac{ab}{\eta + 1} \tag{5}$$

Specifically, the straight-line demand curve will have unit elasticity when the quantity is equal to $ab/2$.

Let us illustrate these peculiarities of the linear demand curve with an example. Shepherd gives a demand curve for potatoes[1] which can be described by the equation

$$Q = 470.7 - 1.01P$$

where Q is the quantity in millions of bushels and P is the price in cents per bushel. First, we have to put this in our standard form, $Q = a(b - P)$, which gives us

$$Q = 1.01(466 - P)$$

We see that a has a value of 1.01 and b a value of 466. Now equation (2) tells us the elasticity at any particular price:

$$\eta = \frac{P}{466 - P}$$

[1] SHEPHERD, GEOFFREY S., "Agricultural Price Analysis," p. 198, Iowa State College Press, Ames, Iowa, 1941.

When, for example, the price is 95 cents, the elasticity of demand is $95/(466 - 95) = 0.256$. Equation (4) tells us the elasticity at any particular quantity.

$$\eta = \left[\frac{(1.01)\ (466)}{Q}\right] - 1$$

When, for example, the quantity is 350 million bushels, the elasticity of demand is $[(1.01)\ (466)/350] - 1 = 0.345$. Where does this demand function have an elasticity of 0.3? Using equation (3) we find that $P = 0.3\ (466)/1.3 = 108$. Thus, the elasticity of 0.3 corresponds to a price of \$1.08. If we use equation (5), we find that $Q = (1.01)(466)/1.3 = 362$. Thus, the elasticity of demand is 0.3 when the quantity is 362 million bushels.

From equations (2) and (4) it is evident that straight-line demand curves not only vary in elasticity from point to point, but that always with such functions, when P grows larger and Q grows smaller, the elasticity becomes greater.

Note 4 (page 253).

Let A represent the average cost per unit and Q the quantity of output. Then AQ is the total cost. We define marginal cost as the rate of change in total cost as we vary the output. If we represent the marginal cost by M, our definition reads

$$M = \frac{d(AQ)}{dQ} = A + Q\left(\frac{dA}{dQ}\right) \qquad (1)$$

At the lowest point on the average-cost curve, $dA/dQ = 0$; consequently at this particular point $M = A$, or marginal cost equals average cost at the point of lowest average cost. It is also easy to see that, wherever expansion of output is reducing average cost, dA/dQ will be negative, and it is evident from (1) that $A > M$. On the other hand, when expansion of output is increasing the average cost, dA/dQ will be positive, and from (1) we see that $A < M$. Therefore, the marginal-cost curve lies below the average-cost curve when the latter is falling, above it when the latter is rising, and crosses the average-cost curve at the latter's lowest point.

Note 5 (page 257).

For proof of the principle that a businessman always finds it profitable to expand his output, in those cases where the selling price is fixed, until marginal cost equals the selling price:

As in Note 4, let A be the average cost per unit of output, and Q the quantity of output.

Let P be the selling price per unit.

Then the total income is PQ and the total expense is AQ. The net income, N, will be the difference.

$$N = PQ - AQ \qquad (1)$$

The rate of change of net income as quantity changes will be

$$\left(\frac{dN}{dQ}\right) = d\left(\frac{PQ}{dQ}\right) - d\left(\frac{AQ}{dQ}\right) \tag{2}$$

$$\left(\frac{dN}{dQ}\right) = P + Q\left(\frac{dP}{dQ}\right) - d\left(\frac{AQ}{dQ}\right) \tag{3}$$

To find the quantity corresponding to the maximum net income, we set this expression equal to zero, which gives us

$$P + Q\left(\frac{dP}{dQ}\right)d = \left(\frac{AQ}{dQ}\right) \tag{4}$$

But we have specified that the price is constant and does not change with changing output. Therefore,

$$\left(\frac{dP}{dQ}\right) = \text{O} \tag{5}$$

and equation (4) becomes

$$P = d\left(\frac{AQ}{dQ}\right) \tag{6}$$

Inspection of equation (1) of Note 4 shows that $d(AQ)/dQ$ is the marginal cost, M, so we can rewrite equation (6) to show that, to maximize net income when P is constant, we must have

$$P = M \tag{7}$$

or, that the selling price and the marginal cost must be equal. As the selling price is fixed and beyond the producer's control, he adjusts his output, Q, until his marginal cost equals the selling price, and he then gets the maximum profit (or the minimum loss).

Note 6 (page 265).

It is obvious from the definition that elasticity of supply is in most ways analagous to elasticity of demand. The supply function, however, is positive, while the demand function is negative. This introduces one or two differences. For example, we learned that, when demand is elastic, rising prices reduce total consumer outlays; when demand is inelastic, rising prices increase total consumer outlays; and when there is unit elasticity, the total consumer outlay is constant. Since the quantity offered by producers rises with rising price, the total value of the goods offered for sale always varies directly with the price, regardless of elasticity. Every linear supply curve that passes through the origin has a constant elasticity of unity, as we see thus:

A straight line passing through the origin can be described by an equation of the general form

$$Q = bP \tag{1}$$

where Q is the quantity offered, P is the price, and b is the slope of the supply

curve (or the reciprocal of the slope as we have drawn the curve, following the general custom of putting the price on the vertical axis). The elasticity of supply can be defined in infinitesimals as follows:

$$\epsilon = \frac{dQ/Q}{dP/P} = \left(\frac{dQ}{dP}\right)\left(\frac{P}{Q}\right) \tag{2}$$

But in equation (1) we have

$$\frac{dQ}{dP} = b \tag{3}$$

Hence, substituting in (2), we get

$$\epsilon = \frac{bP}{Q} \tag{4}$$

But from (1) this gives us

$$\epsilon = \frac{bP}{bP} = 1 \qquad \text{Q.E.D.} \tag{5}$$

This is but part of a more general rule, as follows: In cases of linear supply curves, the supply is elastic, inelastic, or of unit elasticity, depending on the point at which the linear supply curve cuts the axes. If we describe the linear supply curve as follows,

$$Q = a + bP \tag{6}$$

where a is the horizontal intercept if P is plotted on the vertical and Q on the horizontal axis, then we see that the reciprocal of the slope of the curve is

$$\frac{dQ}{dP} = b \tag{7}$$

From (2) we get

$$\epsilon = \frac{bP}{Q} \tag{8}$$

From (6) and (8) we get

$$\epsilon = \frac{bP}{a + bP} \tag{9}$$

It is obvious from (9) that, when a is negative (which means that the quantity offered for sale drops to zero at some positive price), $\epsilon > 1$ and the supply is elastic. When a is positive (which means that the quantity offered for sale is still positive even when the price has dropped to zero), $\epsilon < 1$ and the supply is inelastic. If a is equal to zero, we get the case discussed above in which the line passes through the origin, and there is unit elasticity of supply. These observations hold true, of course, only for linear supply curves. Every supply curve of constant elasticity other than those of constant unit elasticity will be curvilinear when plotted on arithmetic scales.

Note 7 (page 305).

The problem of Tables 16·1 and 16·2 can easily be set up in mathematical form and solved for all values of the variable expense. We note that the demand function given in Table 16·1 is linear. Each added dollar of price reduces the quantity taken by consumers by 10 units. Obviously, the formula for the demand is

$$Q = 200 - 10P \tag{1}$$

where Q is the quantity taken by consumers and P is the price. The total income obtained by selling Q units at price P will be PQ, or, from (1) we get

$$PQ = 200P - 10P^2 \tag{2}$$

as our expression for total income. The total variable expense will be made up of the variable expense per unit, V, multiplied by the quantity of output, Q, or VQ, and the total fixed expense will be F, a constant. The net income will be the total income, PQ, minus the total expense. Thus, we can write the expression for net income as follows, where N is the net income,

$$N = PQ - VQ - F \tag{3}$$

Since the total variable expense is VQ, we substitute the value of Q from (1) to find that the variable expense is

$$200V - 10VP \tag{4}$$

Substituting (4) and (2) in (3) we get a new expression for the net income:

$$N = 200P - 10P^2 - 200V + 10VP - F \tag{5}$$

We want to find the price at which this net income will be maximized, which means that we must differentiate (5) with respect to P and set the first differential equal to zero. Thus, the condition for maximum net income is that

$$\frac{dN}{dP} = 200 - 20P + 10V = 0 \tag{6}$$

Evidently, this is true when

$$20P = 200 + 10V \tag{7}$$

or when

$$P = 10 + \frac{V}{2} \tag{8}$$

This is our general formula for the case set up in Table 16·1. It tells us that the most profitable price will vary with the variable expense and can be found by adding half the variable expense to 10. Thus, when the variable expense is \$10, the most profitable price is \$10 + \$10/2, or \$15; when the variable expense is \$8, the most profitable price is \$10 + \$8/2, or \$14. We can find at once the answers to the various problems of Sec. 16·8.

It is interesting to note the reason why fixed costs play no part in

determining the most profitable output. The fixed cost appears in (5) but, since it is a constant, it disappears in differentiation, and we do not find it in (6), which defines the most profitable price. It is necessary to know the fixed cost if we wish to evaluate the amount of net income, as in (5), but it is not necessary to know the fixed cost if we merely want to know what price is most profitable, as in (6).

We may make our statement above somewhat more general, to cover all cases of linear demand curves, if we state our linear demand function in its general form:

$$Q = a + bP \tag{9}$$

$$N = aP + bP^2 - aV - bVP - F \tag{10}$$

$$\frac{dN}{dP} = a + 2bP - bV \tag{11}$$

The condition for maximum net income is that

$$a + 2bP - bV = 0 \tag{12}$$

or that

$$P = \frac{bV - a}{2b} \tag{13}$$

Since demand functions such as (9) are negative, b will be negative and P in (13) will be positive if a is positive, as would certainly be true for economic goods. For any particular linear demand function, the values of a and b will be given and (13) can be put in the form:

$$P = \frac{V}{2} - \frac{a}{2b} \tag{14}$$

The last term of this equation will be a constant, so it is evident that for linear demand functions the most profitable price will increase exactly with $V/2$; if we add \$10 to the variable expense, we add \$5 to the most profitable price, etc. The student can show easily that in the still more general case, where the demand function is not necessarily linear, the most profitable price will still rise when the variable expense rises and fall when the variable expense falls if we merely make the following reasonable assumptions:

The variable expense depends on the amount of the output, and, at least after diminishing productivity begins to operate, the function is positive, thus,

$$V = f(Q), \qquad f'(Q) > 0 \tag{15}$$

The quantity taken by consumers depends on the price and varies inversely with the price. This is the Law of Demand, thus,

$$Q = F(P), \qquad F'(P) < 0 \tag{16}$$

From these, of course, one sees that the variable expense is a negative function of the price:

$$V = \phi(P), \qquad \phi'(P) < 0 \tag{17}$$

Note 8 (page 311).

If Q represents the quantity of output and P represents the price per unit, then total income will be PQ. We have defined marginal revenue as the change in total income that accompanies a small change in output. If we are to be more precise, we define it as follows, letting M stand for marginal revenue:

$$M = d\frac{(PQ)}{dQ} \tag{1}$$

By the usual multiplication theorem, this can be written

$$M = P + Q\frac{dP}{dQ} \tag{2}$$

We earlier defined elasticity of demand (see Note 3) as follows:

$$\eta = \left(\frac{dQ}{dP}\right)\left(\frac{P}{Q}\right) \tag{3}$$

From (3) we see that

$$\frac{1}{\eta} = \left(\frac{dP}{dQ}\right)\left(\frac{Q}{P}\right) \tag{4}$$

Hence,

$$\frac{P}{\eta} = \left(\frac{dP}{dQ}\right)Q \tag{5}$$

and equation (2) can be written

$$M = P + \frac{P}{\eta} \tag{6}$$

Since, however, the demand function is always negative and since the economist always uses the positive or absolute value of η, we shall have to reverse the signs in (6) to give

$$M = P - \frac{P}{\eta} = P\left(1 - \frac{1}{\eta}\right) \tag{7}$$

which is the relationship given in the text.

While we are on the subject, we should point out that equation (2) tells us something of the relationship between marginal revenue and the selling price, which we mentioned on page 310. If variations in output have any effect at all on price, increases in output will reduce the price and dP/dQ will be negative. In cases of complete competition, the price will not change at all when the output is changed, and dP/dQ will be equal to zero. We can thus lay down the two rules:

1. For cases of complete competition:

$$\frac{dP}{dQ} = 0 \tag{8}$$

2. For cases in which there is any element of monopoly whatever—all cases save those of complete competition:

$$\frac{dP}{dQ} < 0$$

If we substitute these values of dP/dQ in equation (2), we find that, for the case of complete competition,

$$M = P \tag{9}$$

In cases in which there is any element of monopoly,

$$M < P \tag{10}$$

The reason for this last result is obvious when we consider the meaning of equation (2). What happens when a businessman sells one more unit of output? If he is producing under competition, his increased output does not affect the price and he sells all his former quantity at the old price plus one new unit at the old price. Hence, the increase in his income is equal to the price. But, if a monopolist sells one more unit he gets the new price, P, for that additional unit; but the fact that he has increased the output reduces his price on all the units that he was selling before. This reduction, amounting to dP/dQ, has to be taken on the whole quantity, Q, which he sells. His new income is made up of two parts: the income from the one new unit of output, which will be equal to the price P; and a reduction in the price from increasing his output, this reduction affecting all the units, Q. Therefore the marginal revenue of equation (2) consists of P and of $Q(dP/dQ)$. Since dP/dQ will always be either negative or equal to zero (the latter in cases of complete competition), it is evident from (2) that the marginal revenue will always be smaller than the selling price save in the limiting case of competition, where marginal revenue and selling price are equal.

Equation (7) shows the relationships among the three quantities: marginal revenue, price, and elasticity of demand. If any two are known, the third may be easily computed. Equation (7) is explicitly in terms of marginal revenue and can be most easily used when price and elasticity are known and one wishes to find the marginal revenue. This formula can easily be recast so as to give explicit expressions for the price or for the elasticity of demand when the other two are known. Such recasting yields the following:

$$P = M\left(\frac{\eta}{\eta - 1}\right) \tag{11}$$

$$\eta = \frac{P}{P - M} \tag{12}$$

Suppose, for example, that in a given case we are told that the selling price is \$5 and the marginal revenue is \$3. Equation (12) tells us immediately that the elasticity of demand is (5)/(5–3) = 2.5. This is an elastic demand (as we could have told anyway as soon as we noted that the marginal revenue was positive). If, on the other hand, we are told that the elasticity of the demand for a commodity is 1.5 and the marginal revenue is \$4, we see from equation (11) that the price must be 4(1.5/0.5) = \$12.

Note 9 (page 313).

For proof of the statement that maximum business profits coincide with the output where marginal revenue equals marginal cost, we proceed as follows:

Let V represent the variable cost per unit of output, and Q represent the quantity of output.

Let F be the amount of the fixed cost.

Then total variable cost will be VQ, and total cost will be $VQ + F$. If P is the selling price, then total income is PQ.

Marginal revenue was defined in equation (1) of Note 8 as

$$M = \frac{d(PQ)}{dQ} \tag{1}$$

Since marginal cost is the rate of change in total cost as output is changed, we have the marginal cost, m, as follows:

$$m = d\frac{(VQ + F)}{dQ} \tag{2}$$

Since the fixed cost is a constant, it disappears on differentiation and (2) becomes

$$m = \frac{d(VQ)}{dQ} \tag{3}$$

The net income, N, is the difference between total income and total cost, or

$$N = PQ - VQ - F \tag{4}$$

The change in net income as output changes is

$$\frac{dN}{dQ} = \frac{d(PQ)}{dQ} - \frac{d(VQ)}{dQ} \tag{5}$$

To maximize the net income, this expression must be set equal to zero, which shows us that the condition for maximum net income is

$$\frac{d(PQ)}{dQ} = \frac{d(VQ)}{dQ} \tag{6}$$

But from (1) and (3) we can restate (6) thus: The condition for maximum net income is

$$M = m \qquad \text{Q.E.D.} \tag{7}$$

The principle can be well illustrated by the simple numerical problem of Table 16·1. In this table, the linear demand function is

$$Q = 200 - 10P \tag{8}$$

Solving for P we get

$$P = 20 - \frac{Q}{10} \tag{9}$$

Total income is PQ, which is

$$PQ = 20Q - \frac{Q^2}{10} \tag{10}$$

Marginal revenue is $d(PQ)/dQ$, or

$$\text{Marg. revenue} = 20 - \frac{Q}{5} \tag{11}$$

Substituting (8) in (11), we get

$$\begin{aligned} \text{Marg. revenue} &= 20 - \frac{200 - 10P}{5} \\ &= 20 - 40 + 2P \\ &= 2P - 20 \end{aligned} \tag{12}$$

The total expense is

$$\text{Expense} = E = 10Q + F \tag{13}$$

Marginal cost is

$$\text{Marg. cost} = \frac{dE}{dQ} = 10 \tag{14}$$

If marginal cost and marginal revenue are to be equal, then from (14) and (12) we have

$$2P - 20 = 10 \tag{15}$$

$$P = 15 \tag{16}$$

This tells us that marginal revenue and marginal cost will be equal when the price is \$15, and we found when we analyzed this particular problem that the price of \$15 did actually yield the largest possible total income.

Since in Table 16·1 the demand is, from (8),

$$Q = 200 - 10P$$

the elasticity of demand will be, by definition,

$$\eta = \left(\frac{dQ}{dP}\right)\left(\frac{P}{Q}\right)$$

which, for this demand gives

$$\eta = -10\left(\frac{P}{Q}\right) \tag{17}$$

Since the economist always uses the absolute value of the elasticity of demand, we can drop the negative sign and say that

$$\eta = \frac{10P}{Q} \tag{18}$$

Now we can recast Table 16·1, computing for each price or volume of output the corresponding marginal revenue by equation (12), the corresponding marginal cost by equation (14) and the corresponding elasticity of demand by equation (18). This will give us the accompanying table, which shows a number of interesting facts, such as that the marginal revenue is

always negative when the demand is inelastic, always positive when the demand is elastic, and equal to zero when there is unit elasticity of demand; or that the most profitable volume of output, which we discovered to be at a price of $15, comes at the point where marginal revenue and marginal cost are equal; or that a monopolist always finds it profitable to contract his output until the demand for his product is elastic. To be sure, this particular problem is peculiar in that the marginal cost is constant at $10. This happened because we assumed that the variable cost was $10 per unit of output, and stayed at that amount for all outputs. Therefore, each added unit of output added $10 to the cost, and the marginal cost was $10.

Computations Based on Table 16·1

Price per unit (P)	Quantity consumers will take (Q)	Marginal revenue ($2P - 20$)	Marginal cost (10)	Elasticity of demand ($10P/Q$)
$7	130	−$6	$10	0.54
8	120	−4	10	0.67
9	110	−2	10	0.82
10	100	0	10	1.00
11	90	2	10	1.22
12	80	4	10	1 50
13	70	6	10	1.86
14	60	8	10	2.33
15	50	10	10	3.00
16	40	12	10	4.00
17	30	14	10	5 67
18	20	16	10	9.00
19	10	18	10	19.00

In practice, it would usually be true that the amount of the variable cost per unit was itself a function of the amount of output, in which case, as soon as we reach diminishing productivity, we would have positive marginal cost but not constant marginal cost.

To show the interrelationships of our formulas still further we might also have computed the marginal revenues of our table from the elasticities of demand by equation (7) of Note 8, namely,

$$\text{Marg. revenue} = P\left(1 - \frac{1}{\eta}\right)$$

If one substitutes for the marginal revenue and for η in this equation the values from equations (12) and (18), he will discover that the two methods

are identical, as they should be. The student might try making a table similar to the one from this note based on Table 16-2.

Note 10 (page 316).

The fact that demand tends to be elastic at low volumes of output and at high prices, and inelastic at large volumes of output and at low prices means, of course, that the demand curve is not so markedly concave upward as it would otherwise be. We saw in Note 3 that demand curves of constant elasticity are concave upward on ordinary arithmetic axes, but straight lines on logarithmic axes (see page 889). But the more elastic the demand, the gentler the slope; the less elastic the demand, the steeper the slope. Therefore, if the demand is most elastic toward the low volumes at the left side of the demand curve, and if it gets less elastic as we progress toward the right side of the diagram and the large volumes, an ordinary demand curve on logarithmic axes can be expected to be concave downward; and a demand curve on ordinary arithmetic axes will at least be less concave upward than otherwise would be true. It is evident that a linear demand curve on arithmetic axes will exhibit smaller and smaller elasticity as the volume of output is increased and as prices are lowered. If Q represents the quantity of the commodity and P represents the price, such a linear demand function would be represented by an equation of the general form:

$$Q = a - bP \tag{1}$$

The elasticity of this demand would be

$$\eta = \left(\frac{dQ}{dP}\right)\left(\frac{P}{Q}\right) = -\frac{bP}{Q} \tag{2}$$

As usual, we discard the minus sign and use the absolute value as our measure of elasticity, giving

$$\eta = \frac{bP}{Q} \tag{3}$$

Since b is a constant, the elasticity will change as P and Q change, varying directly with P and inversely with Q. At large volumes or at low prices, the elasticity will be low; at low volumes or at high prices, the elasticity will be large. And somewhere we shall pass through a point of unit elasticity, where $bP = Q$. We can illustrate this easily with the figures in the table accompanying Note 9. This table shows a linear demand function, and we note from the final column that the demand is elastic at low quantities and inelastic at high quantities. We note that the curve passes through a point of unitary elasticity and that this point falls at the place where $bP = Q$ if we use the absolute value of b, since the formula for this function is $Q = 200 - 10P$. Another easy way of finding the point at which a linear demand function has unit

elasticity was given in Note 3 in the comment between equations (2) and (3) under the discussion of linear demand functions.

Note 11 (page 318).

In the problems of Tables 16·4 and 16·5, the variable expense is \$2 per unit and the fixed expense is \$100. In Table 16·4 we have a linear demand function with the formula

$$Q = 640 - 80P \tag{1}$$

According to equation (14) of Note 7 (see page 895), this would give an optimum or most profitable price of \$5 per unit. But if the demand of equation (1) is broken down into its two constituents, as in Table 16·5, we find the two linear demand functions:

$$Q_a = 400 - 40P \tag{2}$$

$$Q_b = 240 - 40P \tag{3}$$

Function (1) is obviously the sum of functions (2) and (3). But, again using equation (14) of Note 7, we see that net income will be maximized for equation (2) at a price of \$6, while for equation (3) it will be maximized at a price of \$4. Since a price of \$6 will yield more net income from equation (2) than would a price of \$5 and since a price of \$4 will yield more net income from equation (3) than would a price of \$5, it is obvious that the sums of these two net incomes obtained by charging distinct prices of \$6 and \$4 will exceed the sum obtained by charging a price of \$5 in each case. Just as obviously, no matter what the demand functions are, if they differ, they may well lead to different maximizing prices. One will be maximized at one price and one at another. And the sum of these individual maxima at different prices will exceed the sum of the smaller amounts that would be received in the different markets at any single uniform price.

The problem may be even more informatively approached in another way. Equation (7) of Note 9 shows us that a monopolist maximizes his profit in any market when he charges such a price that the marginal revenue and the marginal cost are equal. If he is selling to several customers, he will try to make the marginal revenue from each customer equal to his marginal cost, so that, to maximize his net income, the marginal revenue in all markets or from all customers must be equal. But, by equation (7) of Note 8, we see that the marginal revenue in any market is equal to $P(1 - 1/\eta)$ where P is the price charged in the market and η is the elasticity of the demand in the market. If we have two different markets, and distinguish the prices and elasticities in them by the symbols P_1 and P_2 and η_1 and η_2, then, if marginal revenues are to be equal, we can write

$$P_1\left(1 - \frac{1}{\eta_1}\right) = P_2\left(1 - \frac{1}{\eta_2}\right) \tag{4}$$

or

$$\frac{P_1}{P_2} = \frac{1 - \frac{1}{\eta_2}}{1 - \frac{1}{\eta_1}} \tag{5}$$

This can be put into the form

$$\frac{P_1}{P_2} = \frac{\eta_2 - 1}{\eta_2} \Big/ \frac{\eta_1 - 1}{\eta_1} = \frac{\eta_1(\eta_2 - 1)}{\eta_2(\eta_1 - 1)} = \frac{\eta_1\eta_2 - \eta_1}{\eta_1\eta_2 - \eta_2} \tag{6}$$

In interpreting this equation, we should recall the statement of Sec. 16·12 that monopolists always find it advantageous to contract their output until the demand for their product is elastic. This means that both η_1 and η_2 can be taken as greater than unity in equation (6). But equation (6) makes it clear that if $\eta_1 > \eta_2$, then $P_1 < P_2$; while if $\eta_1 < \eta_2$, then $P_1 > P_2$. To translate this into ordinary speech, we see that, if the demand for the commodity is more elastic in market 1 than in market 2, the monopolist will find it to his advantage to charge a lower price in market 1 than in market 2; in more general terms, a discriminating monopoly will find it advantageous to charge different prices to different customers or in different markets if the demands in these different markets have different elasticities; and it will pay to charge the higher price in the market where the demand is least elastic.

Note 12 (page 327).

To prove that the marginal cost curve cuts the curve of average variable cost at the latter's lowest point:

Let V = average variable cost per unit.
Let Q = volume of output.
Let F = the total fixed cost.
Let T = the total cost.

Then,

$$T = VQ + F \tag{1}$$

But the marginal cost is the rate at which the total cost changes as we vary output, or,

$$\text{Marg. cost} = \frac{dT}{dQ} = \frac{d(VQ)}{dQ} \tag{2}$$

The fixed cost drops out because it is constant. But, by the usual theorem,

$$M = d\frac{(VQ)}{dQ} = V + Q\left(\frac{dV}{dQ}\right) \tag{3}$$

Since Q is always positive, when the variable cost curve is falling and dV/dQ is negative the marginal cost is smaller than the average variable cost, and the marginal cost curve lies below the average variable cost curve. When the variable cost curve is rising and dV/dQ is positive, the marginal cost exceeds the average variable cost, and the marginal cost curve lies above

the average variable cost curve. Marginal cost will equal average variable cost, and the curves will cross, only when dV/dQ is equal to zero. This is the condition for the minimum point on the curve of average variable costs. Therefore, the two curves cross at the point where V, the average variable cost, is a minimum.

The student who wishes to experiment with these ideas will be interested to learn that Fig. 17·1 is drawn to scale from the following formulas:

$$V = 170 - 22Q + Q^2 \tag{4}$$

Evidently total variable cost is VQ, or

$$VQ = 170Q - 22Q^2 + Q^3 \tag{5}$$

The fixed cost is taken to be \$676, so total cost according to (1) will be

$$T = 676 + 170Q - 22Q^2 + Q^3 \tag{6}$$

Marginal cost, according to (2), will be

$$M = 170 - 44Q + 3Q^2 \tag{7}$$

It is evident from (7) and (4) that

$$M = V - 22Q + 2Q^2 \tag{8}$$

To find the point of lowest average variable cost, we differentiate (4) and set it equal to zero, giving

$$2Q - 22 = 0 \tag{9}$$

This tells us that the point of lowest average variable cost occurs when Q is 11. To find the lowest average cost, we remember that the average cost, A, is

$$A = \frac{T}{Q} \tag{10}$$

But, from (6) this gives us

$$A = \frac{676}{Q} + 170 - 22Q + Q^2 \tag{11}$$

This will pass through its minimum point if its first differential with respect to Q is set equal to zero

$$\frac{-676}{Q^2} - 22 + 2Q = 0 \tag{12}$$

This turns out to yield the value

$$Q = 13 \tag{13}$$

which tells us that the minimum point on the average cost curve occurs when Q is 13. We note from (11) and (7) that, when Q is 13, A and M are both 105, so that these curves cross at the point where Q is 13. Similarly we note from (4) and (7) that, when Q is 11, A and M are both 49, so that these curves cross at the point where Q is 11.

Note 13 (page 499).

To illustrate the cobweb theorem let us define our terms as follows:

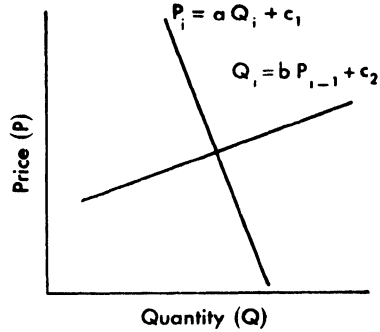
Q_i is the quantity of the commodity on the market at time i. Thus, in year 0, the quantity is Q_0, in year 5 the quantity is Q_5, etc.

P_i is the price of the commodity at time i. As above, in the year 0 the price is P_0 and in the year 5 the price is P_5.

If we illustrate with linear functions, the price curve has the general formula:

$$P_i = aQ_i + c_1 \tag{1}$$

where a and c_1 are constants for the particular price curve in question.



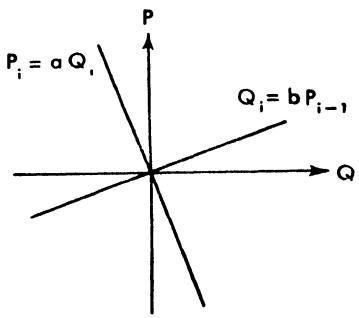


FIG. A·3.—Illustration of cobweb theorem.

The forecast curve has the general formula,

$$Q_i = bP_{i-1} + c_2 \tag{2}$$

where b and c_2 are constants for the particular forecast curve in general. It will be noted that price is dependent in (1) and independent in (2). It will also be noted that the price for any year depends on the quantity in the same year in (1); but that the quantity in any year depends on the price of the preceding year in (2). The curves are shown in the upper part of Fig. A·3.

For ease in treatment, let us shift the origin to coincide with the crossing point of the two curves, as in the lower part of the diagram. We can then drop the constant terms of our equations to get

$$P_i = aQ_i \tag{3}$$

$$Q_i = bP_{i-1} \tag{4}$$

Suppose that we start with the price, P_0, in year 0. By (3) we have

$$P_0 = aQ_0 \tag{5}$$

In year 1 the price becomes

$$P_1 = aQ_1 \tag{6}$$

But by (4),

$$Q_1 = bP_0 \tag{7}$$

From (6) and (7),

$$P_1 = abP_0 \tag{8}$$

For later years we have

$$Q_2 = bP_1 = ab^2P_0$$

$$P_2 = aQ_2 = (ab)^2P_0$$

$$Q_3 = bP_2 = a^2b^3P_0$$

$$P_3 = aQ_3 = (ab)^3P_0, \text{ etc.}$$

Obviously,

$$P_n = P_0(ab)^n \tag{9}$$

The constants a and b in equation (9) are related to the slopes of the curves in the diagram. Thus a is the slope of the price curve, and b is the reciprocal of the slope of the forecast curve. Let us take

$$B = \frac{1}{b} \tag{10}$$

Then B is the slope of the forecast curve and a the slope of the price curve, and (9) becomes

$$P_n = P_0\left(\frac{a}{B}\right)^n \tag{11}$$

From equation (11) we can find the price in any year, n, as it is related to the price in the year 0. Let us remember that the slope of the price curve, a, is negative and the slope of the forecast curve, B, is positive. Therefore the factor (a/B) in equation (11) is always negative, and as we pass from year 0 to years 1, 2, 3, 4, . . . , it is raised alternately to even and odd powers. Thus the price fluctuates from year to year, first above and then below the equilibrium price. The results can conveniently be classified into three groups:

1. Those cases where the forecast curve has a steeper slope than the price curve. This means that $|B| > |a|$. Under such circumstances $0 > a/B > -1$. As a/B is raised to higher and higher powers, P_n draws closer and closer to

the equilibrium position, alternating first above and then below but always drawing closer. For example, if $a = 1.5$ and $B = -3.0$ (which means of course that $b = -\frac{1}{3}$), we have $a/B = -\frac{1}{2}$. If we start with $P_0 = \$12$, the price in the following years will be $\$12(-\frac{1}{2})$, $\$12(-\frac{1}{2})^2$, $\$12(-\frac{1}{2})^3$, etc. This series gives rise to the prices \$12, −\$6, \$3, −\$1.5, etc. This is the converging case. Whenever the forecast curve is steeper than the price curve, the series converges on the equilibrium point, approaching it through alternations above and below.

2. Those cases where the forecast curve and the price curve are equal in slope, so that $|a| = |B|$. Under such circumstances $a/B = -1$, and we can write equation (9) in the form

$$P_n = P_0(-1)^n$$

As we raise a/B to higher and higher integral powers, the value of P_n will fluctuate back and forth, first above and then below the equilibrium point, keeping a constant distance from equilibrium. For example, if we start again with $P_0 = \$12$, the prices in the following years will be $\$12(-1)$, $\$12(-1)^2$, $\$12(-1)^3$, $\$12(-1)^4$, etc. This series gives the values \$12, −\$12, \$12, −\$12, etc.—first \$12 above and then \$12 below equilibrium alternately. This is the constant case, and it occurs whenever the forecast curve and the price curve have equal slopes.

3. Those cases where the price curve is steeper than the forecast curve, so that $|B| < |a|$. Under such circumstances $a/B < -1$. As a/B is raised to higher and higher integral powers, it becomes larger and larger in absolute value but alternates from positive to negative in sign. Therefore, P_n recedes farther and farther from the equilibrium position through a series of alternations first above and then below equilibrium. If, for example, $a = 3$ and $B = -1.5$, we have $a/B = -2$. If we start with $P_0 = \$12$, the prices in the following years will be $\$12(-2)$, $\$12(-2)^2$, $\$12(-2)^3$, etc. This series gives rise to the prices \$12, −\$24, \$48, −\$96, etc. This is the diverging case, and it occurs whenever the price curve is steeper than the forecast curve.

It may be of interest to point out that the forecast curve may have a negative slope like the price curve. The situation might occur in which a high price in any given year would bring about a low production in the following year; for example, because each producer in making his own individual plans assumes that the high price will lead other producers to expand; therefore, he decides that the market will be glutted and he should shift to some other crop. In such a case, both curves will have negative slope, so a and B are both negative and a/B is positive. Under these circumstances, there is no alternation of sign as a/B is raised to higher and higner powers—no alternation above and below the equilibrium point; yet it is still possible to have either converging or diverging cases. These are illustrated in Fig. A·4.

1. If, as in the upper part of the diagram, the forecast curve is the steeper, so $|B| > |a|$, we find that $0 < a/B < 1$. As a/B is raised to higher and higher powers, its value becomes smaller and smaller and we get convergence toward equilibrium. Thus, if, in the upper part of the diagram, we start with output Q_0, it will sell with a price at A. This will stimulate an output the next year at B, which will force the price down to C, which will stimulate an output the next year of D, and so on until equilibrium is reached.

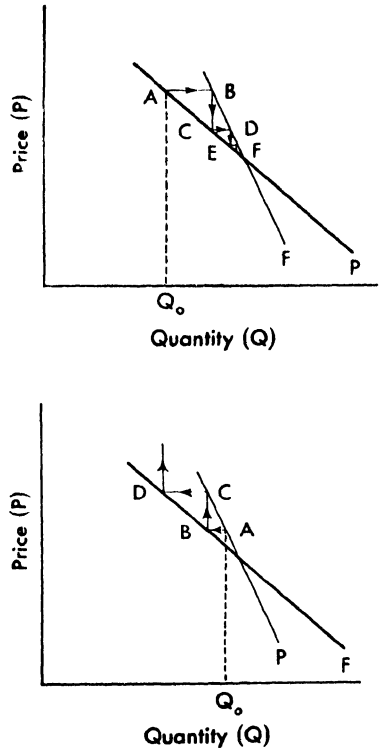


FIG. A • 4.—Converging and diverging cobweb movements.

2. If, as in the lower part of the diagram, the price curve is the steeper, so $|a| > |B|$, we find that $a/B > 1$. As a/B is raised to higher and higher powers, its value becomes larger and larger and we get divergence from equilibrium. Thus if, in the lower part of the diagram, we start with output Q_0, it will sell at price A. This will stimulate an output next year of B, which will sell at C, which will stimulate an output the next year of D, etc., drawing farther and farther from equilibrium. As before, if the forecast curve is steeper, the series converges; if the price curve is steeper, the series diverges.

If, of course, we happen to be in an equilibrium position, where P and Q are zero (since we have taken our origin at the crossing point of the lines), there is no tendency to leave the equilibrium position. This, to be sure, is why it is called an "equilibrium position." Since under these circumstances $P_0 = 0$ and $P_n = P_0(a/B)^n$ from equation (11), P_n will always stay at 0 no matter what value we give to n. But, if equilibrium is once upset by some external force, the cobweb theorem tells us that we would then

1. Tend to restore the old equilibrium by a series of converging movements, if the forecast curve is steeper than the price curve.
2. Set up a perpetual oscillation above and below the old equilibrium with a constant amplitude, if the curves have equal slope.
3. Set up a series of movements getting farther and farther from equilibrium, if the price curve is steeper than the forecast curve.[1]

Note 14 (page 627).

The problem of the localization of industry is a complex one, influenced by many factors. On the whole, however, these factors can be broken down into two major classes: differences in the cost of producing the product locally, and differences in the cost of transporting it to market. The costs of transportation depend on many things, such as the means of transportation used (oxcart, motor truck, airplane, etc.) and the availability of highways, rail lines, navigable rivers, etc. It may consequently be much cheaper to transport in some directions than in others from a given starting point.

For our purposes, however, let us consider consumers scattered over the surface of the territory in which two manufacturing companies produce exactly the same product at exactly the same cost, so that both quote exactly the same price, P, for a unit of the commodity sold at their own plants. The two firms differ, however, in transportation cost. For each firm, the cost varies directly with the distance that the goods must be transported, the distance in each case being measured from factory to consumer in a straight line. (This is equivalent to assuming that the goods are transported by helicopter rather than over highways, railroads, navigable rivers, etc. In other words, it is assumed that the cost of transportation is the same in all directions, and varies only with distance.)

It costs firm A an amount of t_a per mile to transport a unit of the commodity, while it costs firm B an amount of t_b. If a given consumer is M_a miles from firm A and M_b miles from firm B, it will cost t_aM_a to deliver a unit of the commodity from firm A and t_bM_b to deliver it from firm B. In addition to this delivery cost, the buyer will also have to pay the selling

[1] While I am indebted to such writers as Ezekiel, Buchanan, Allen, Boulding, Hicks, and Lundberg for my treatment of the cobweb theorem, I should take this occasion to pay tribute to Dr. Oscar Lange, present Polish representative at the United Nations, whose treatment of the subject has seemed to me the most general and yet the simplest with which I am familiar.

price at the factory, P. Therefore any consumer will have to pay a total amount of $P + t_aM_a$ if he buys from firm A, and a total amount of $P + t_bM_b$ if he buys from firm B. Some buyers will find it cheaper to buy from one firm and some from the other; but on the dividing line between the two territories the cost of buying from one firm will exactly equal the cost of buying from the other. On this dividing line, then,

$$P + t_aM_a = P + t_bM_b \tag{1}$$

Therefore, on the dividing line,

$$t_aM_a = t_bM_b \tag{2}$$

We can perhaps vizualize the situation best in connection with Fig. A·5, in which A and B represent the locations of the two firms, and P is a point selected at random on the dividing line between the two territories. Let us so select our axes that the straight line connecting the two firms

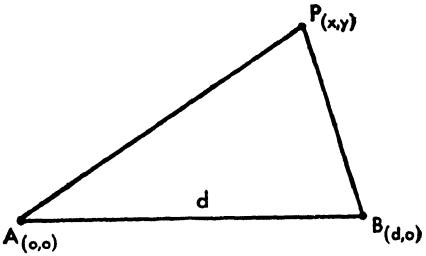


FIG. A·5.—Competing firms with equal production costs but different transportation costs.

coincides with the axis of abscissas, with the origin at firm A. If the distance between the two firms is d miles, the coordinates of the two firms will be (0,0) and (d, 0), and the coordinates of the point P which is selected at random on the line dividing the territories are (x, y).

The distance from firm B to P is M_b, and the distance from firm A to P is M_a. But, by the distance formula of analytical geometry, these distances are

$$BP = M_b = \sqrt{(x-d)^2 + y^2} \tag{3}$$

$$AP = M_a = \sqrt{x^2 + y^2} \tag{4}$$

If we substitute (3) and (4) in (2) we get

$$t_a\sqrt{x^2+y^2} = t_b\sqrt{(x-d)^2+y^2} \tag{5}$$

Squaring both sides gives

$$t_a^2(x^2 + y^2) = t_b^2[(x-d)^2 + y^2] \tag{6}$$

Expanding the right-hand member gives

$$t_a^2(x^2 + y^2) = t_b^2(x^2 - 2dx + d^2 + y^2) \tag{7}$$

Collecting terms and factoring,

$$(t_a^2 - t_b^2)(x^2 + y^2) = t_b^2(d^2 - 2dx) \tag{8}$$

$$x^2 + y^2 = \frac{t_b^2}{t_a^2 - t_b^2}(d^2 - 2dx) \tag{9}$$

Let us set

$$m = \frac{t_b^2}{t_a^2 - t_b^2} \tag{10}$$

This obviously depends solely on the transportation costs of the two firms. If we substitute this value of m in (9), we get

$$x^2 + y^2 = m(d^2 - 2dx) = md^2 - 2mdx \tag{11}$$

Transposing terms we have

$$x^2 + 2mdx + y^2 = md^2 \tag{12}$$

Adding m^2d^2 to both sides to complete the square,

$$x^2 + 2mdx + m^2d^2 + y^2 = md^2 + m^2d^2 \tag{13}$$

Factoring,

$$(x + md)^2 + y^2 = md^2(1 + m) \tag{14}$$

But this is the familiar formula for a circle with center at $(-md,0)$ and with radius $\sqrt{md^2(1 + m)}$. Therefore, we see that under the given circumstances the firm with higher transportation costs will sell only within the territory surrounded by this circle, and the firm with lower selling cost will sell everywhere outside the circle.

The situation appears for an illustrative case in Fig. A·6. Firms A and B are 80 miles apart. Firm A has shipping charges of 5 cents per mile as compared with 15 cents a mile for firm B. Thus, $t_a = 5$ and $t_b = 15$. From (10) we have

$$m = \frac{15^2}{5^2 - 15^2} = \frac{-9}{8}$$

Obviously, m will take this same value, regardless of the actual transportation costs, as long as B's cost is three times A's cost. The relative costs, not the absolute costs, are important.

Since, in our example, the firms are 80 miles apart, d is 80. We have discovered that the circle dividing the territories is centered at $(-md,O)$, or, in this case, at $-(-9/8)(80)$, or at (90,O). The radius is $\sqrt{md^2(1 + m)}$, or $(-9/8)(80^2)(1 - 9/8) = \sqrt{900} = 30$. Firm B, having the higher shipping cost, will find its market confined within the circle, while firm A will be able to undersell firm B anywhere outside the circle. Although it may surprise the student at first to note that firm A will be able to ship through point B to customers on the far side, the reason is obvious. Let us take a case. A customer is located on the direct line from A to B. He is 150 miles from A in the direction of B and therefore 70 miles beyond B. If he buys

from B, he pays the factory cost plus 15 cents per mile for 70 miles, or the factory cost plus \$10.50. If he buys from A, he pays the same factory cost plus 5 cents per mile for 150 miles, or the factory cost plus \$7.50. Obviously, it pays to buy from A even though the goods are shipped through B. The result will seem logical if the student thinks of firm B as shipping by oxcart while firm A uses motor trucks.

Let us now examine the case where there are two firms producing the same commodity, but this time their costs of production (hence their prices at the factory) differ, while their transportation costs per mile are the same. Let the factory price at firm A be P_a and the factory price at firm B be P_b. Let the cost per mile of transporting the commodity be T. Consider any consumer, chosen at random, who is M_a miles from A and M_b miles from B.

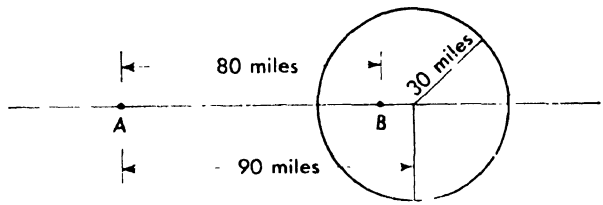


FIG. A·6.—Sales territories of firms having equal production costs but different transportation costs.

If he buys from A, he will have to pay the factory cost plus transportation, or $P_a + M_aT$. Similarly, if he buys from B, he will pay $P_b + M_bT$. On the border line between the two territories these buying costs will be identical, or

$$P_a + M_aT = P_b + M_bT \tag{15}$$

Transposing,

$$P_a - P_b = T(M_b - M_a) \tag{16}$$

Dividing both sides by T gives

$$\frac{P_a - P_b}{T} = M_b - M_a \tag{17}$$

Let us take a constant, a, such that

$$2a = \frac{P_a - P_b}{T} \tag{18}$$

Then from (17) and (18) we have

$$2a = M_b - M_a \tag{19}$$

But M_a and M_b are the distances of a point on the border line between the two trading territories from A and B, respectively. Let us show them as in Fig. A·7, where we have taken the axis of abscissas again coinciding with the line joining the two firms. We have denoted the distance between the

two firms as $2c$ and have taken the origin of coordinates halfway between them. Therefore, the coordinates of firm B are $(-c,0)$ and the coordinates

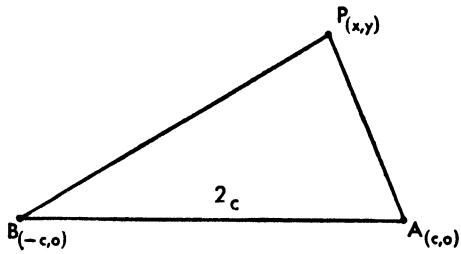


FIG. A • 7.—Competing firms with equal transportation costs but different production costs.

of firm A are $(c,0)$. The coordinates of the point on the border line are (x, y). By the ordinary distance formula,

$$AP = M_a = \sqrt{(x-c)^2 + y^2} \tag{20}$$

$$BP = M_b = \sqrt{(x+c)^2 + y^2} \tag{21}$$

Substituting these values in (19) gives

$$\sqrt{(x+c)^2 + y^2} - \sqrt{(x-c)^2 + y^2} = 2a \tag{22}$$

Transposing the right-hand radical,

$$\sqrt{(x+c)^2 + y^2} = 2a + \sqrt{(x-c)^2 + y^2} \tag{23}$$

Squaring both sides,

$$(x+c)^2 + y^2 = 4a^2 + 4a\sqrt{(x-c)^2 + y^2} + (x-c)^2 + y^2 \tag{24}$$

Expanding,

$$x^2 + 2xc + c^2 + y^2 = 4a^2 + 4a\sqrt{(x-c)^2 + y^2} + x^2 - 2xc + c^2 + y^2 \tag{25}$$

Gathering terms,

$$4xc - 4a^2 = 4a\sqrt{(x-c)^2 + y^2} \tag{26}$$

Dividing by 4,

$$xc - a^2 = a\sqrt{(x-c)^2 + y^2} \tag{27}$$

Squaring both sides,

$$x^2c^2 - 2a^2xc + a^4 = a^2[(x-c)^2 + y^2] \tag{28}$$

Expanding,

$$x^2c^2 - 2a^2xc + a^4 = a^2x^2 - 2a^2cx + a^2c^2 + a^2y^2 \tag{29}$$

Collecting terms and transposing,

$$a^4 - a^2c^2 = a^2x^2 - c^2x^2 + a^2y^2 \tag{30}$$

Factoring,

$$a^2(a^2 - c^2) = x^2(a^2 - c^2) + a^2y^2 \tag{31}$$

Dividing by $a^2(a^2 - c^2)$ gives

$$1 = \frac{x^2}{a^2} + \frac{y^2}{a^2 - c^2} \tag{32}$$

Select b in such a way that

$$b^2 = c^2 - a^2 \tag{33}$$

Substituting (33) in (32) gives

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 \tag{34}$$

But this is the equation of a hyperbola with the two firms located at the foci. Only one branch of the hyperbola is used when firm A has the higher cost of production, and the other when firm B has the higher cost.

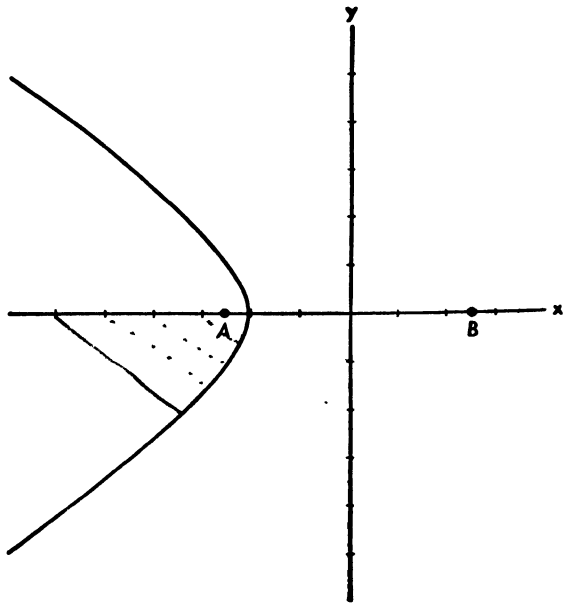


FIG. A • 8.—Sales territories of firms with equal transportation costs but different production costs.

Let us take a specific numerical example. Two firms are situated 10 miles apart. Therefore $2c = 10$ and $c = 5$. Firm A charges \$1.74 at the factory and firm B charges \$1.50. The transportation cost for either firm is 3 cents per mile. From equation (18) we discover that $a = 4$, and from equation (33) we see that $b = 3$. Substituting these values of a and b in equation (34) gives us the formula for the line dividing the territory of firm A from that of firm B. In the present case we get

$$\frac{x^2}{4^2} - \frac{y^2}{3^2} = 1$$

Substituting various values of y and solving for x, we get the following table of values:

y	0	1	2	3	4	5	6	7	8	9	10
x	4.00	4.21	4.81	5.66	6.66	7.77	8.94	10.1	11 4	12 6	13 9

These values are plotted to scale in Fig. A·8, where we see the central section of that branch of the hyperbola which separates the market areas of firms A and B. Firm B can not ship through point A this time, as it could in the preceding case. Firm A will sell throughout the territory at the left within the curve, and firm B will sell in all the territory outside.

Note 15 (page 732).

Let F be the quantity of a factor of production that is utilized.

Let Q be the quantity of output obtained.

Let P be the price per unit at which the output is sold.

Then,

$$Q = f(F), \qquad f' > 0 \tag{1}$$

For competition, output can be increased without influencing price, so

$$P = \text{constant for competition} \tag{2}$$

For monopoly, increased output can be sold only at lower price, so

$$P = \Phi(Q), \qquad \Phi' < 0 \text{ for monopoly} \tag{3}$$

Let V be the marginal value product, or value of the marginal product (whichever we prefer to call it).

By definition,

$$V = \frac{d(PQ)}{dF} \tag{4}$$

$$V = \frac{PdQ}{dF} + \frac{QdP}{dF} \tag{5}$$

But we see from (2) that, for competitive conditions, $dP/dF = 0$. That is, as the producer employs more units of the factor and increases his output, prices do not change. Hence the second term of (5) disappears for the competitive case and leaves

$$V = \frac{PdQ}{dF} \text{ for competition} \tag{6}$$

This is the same as saying that the marginal value product in the competitive case is found by multiplying the increased output by the price per unit.

In the case of monopoly, we discover from (1) and (3) that changes in the use of the factor of production will change the output of the commodity

and hence the price of the commodity. In other words, from (1) and (3) it is true that

$$P = \psi(F), \qquad \psi' < 0 \tag{7}$$

Substituting this result in (5), we see that the value of V for the monopoly situation will be less than that for the competitive situation, since dP/dF is negative. In fact, we can see that the marginal value product will be found as the difference between

1. The marginal product (dQ/dF) multiplied by the price, and
2. The reduction in price caused by the increased output (dP/dF) multiplied by the amount of the output.

Since P, dQ/dF, and Q are always positive and since dP/dF is always negative or zero (being zero for competition and negative for monopoly), the value of V will be higher for competition than for monopoly. It is this value of V that determines the maximum that the employer can offer for the use of the factor of production.

In the body of the text we have used the term "value of the marginal product" to represent the idea PdQ/dF, and the term "marginal value product" to represent the entire expression $PdQ/dF + QdP/dF$.

Note 16 (page 781).

The rate of interest quoted on any other basis can be converted to an equivalent rate of interest on an annual basis by means of the following formulas:

Let B represent the amount borrowed, R the amount repaid, I the amount of interest $= R - B$, and i the equivalent annual rate of interest. Let n be the number of years between the time of borrowing and the time of repayment.

For simple interest (no compounding), we have

$$i = \frac{R - B}{nB} \tag{1}$$

In other words, we change from the rate of interest for the entire period to the rate of interest for a year merely by dividing by the number of years. If the credit period is less than a year, n is a fraction. Thus, if one borrows for 3 months, n is $\frac{1}{4}$.

If interest is compounded annually and the credit period exceeds a year, we have

$$i = \sqrt[n]{\frac{R}{B}} - 1 \tag{2}$$

For example, a man borrows \$1000 and 10 years later pays back \$1100. The amount of interest is \$100, and the interest rate *on a ten-year basis* is 10 per cent. But the equivalent *annual* rate of interest, compounded annually is

$\sqrt[10]{1100/1000} - 1 = 0.0096$, or 0.96 per cent, as compared with the equivalent rate of 1 per cent if there had been no compounding.

If the credit period is shorter than a year, state it as a fraction, f, of the year. For example, if the credit period is 3 months, f is $\frac{1}{4}$. Then the equivalent annual rate of interest is found by the formula

$$i = \left(\frac{R}{B}\right)^{1/f} - 1 \tag{3}$$

For example, if a man borrows \$100 for a month and then repays \$102, f is 1/12, R/B is 1.02, $1/f$ is 12, and i is 0.268 or 26.8 per cent per annum.

We can apply these formulas easily in determining the relationship between bank discount and ordinary interest.[1] With a bank discount rate of 5 per cent, a bank would lend \$95 on a \$100 note and a year later would collect the full face value of \$100. Using the formula of section 33·2, namely,

$$i = \frac{R - B}{B} \tag{4}$$

which we can use without other complications since the credit period is exactly a year, we find that i is 5.25 per cent. Similarly, we find that bank discount of 6 per cent is the equivalent of 6.36 per cent interest, bank discount of 4 per cent is the equivalent of 4.16 per cent interest, etc. In fact, we can easily convert from any given rate of discount, d, to the equivalent rate of interest, i, by the formula

$$i = (1 + d)(100d) \tag{5}$$

Thus, if the discount rate is 3 per cent, so that d is 0.03, the corresponding interest rate is

$$i = (1.03)(3) = 3.09 \text{ per cent}$$

Throughout this Note interest and discount rates are treated as decimals rather than as percentages; that is, an interest rate of 3 per cent, by these formulas, would appear as 0.03.

[1] See Sec. 19·6.

NAME INDEX

A

Allen, R. G. D., 909
Arkwright, Richard, 31
Arneson, Gordon, 577
Arnold, Thurman, 597

B

Bell, Alexander Graham, 32
Bernheim, Alfred L., 134, 581, 591
Black, John D., 107, 711
Boulding, K. E., 909
Bradford, William, 19
Bresciani-Turroni, C., 466
Bryan, William Jennings, 368
Buchanan, Norman S., 586, 909

C

Chandler, F. R., 755
Chase, Salmon P., 429
Cleveland, Grover, 536
Clinton, George, 427
Comer, George P., 667
Cooke, Jay, 479
Copeland, Morris, 469
Crompton, Samuel, 31
Crowther, Samuel, 574, 648
Crum, William L., 591

D

Dallas, Alexander, 427
Davenport, H. J., 751
Davis, J. S., 711
Day, Clive, 24
Dewey, Anderson, H., 134, 135
Dewey, Davis R., 429
Dewing, A. S., 591
Douglas, Paul H., 765, 772
Duke, J. B., 596

E

Ellis, Paul W., 814, 816, 817, 820, 850-853
Engel, Ernst, 54
Engels, Friedrich, 876
Epstein, Ralph C., 591
Ezekiel, Mordecai, 909

F

Fetter, Frank A., 610
Field, J. A., 755
Fisk, James, 484
Fisher, Irving, 465, 493
Ford, Henry, 574, 649
Fulton, Robert, 32

G

George, Henry, 756
Gilbert, C. G., 195
Goethe, Johann Wolfgang von, 488
Gould, Jay, 484
Grayson, David, 174
Gresham, Sir Thomas, 364
Gumbart, Edward H., 8

H

Hamilton, Alexander, 426
Hargreaves, James, 31
Herschel, Sir William, 488
Hicks, J. R., 909
Hill, James J., 595
Hogben, Lancelot, 194
Holdsworth, John T., 384
Hoover, Herbert, 772
Houston, David F., 846
Howe, Elias, 239

J

Jackson, Andrew, 428

James, Clifford L., 577
Jaszi, Oscar, 874
Jevons, W. Stanley, 488
Jones, Eliot, 605

K

Kay, John, 31
Knight, Frank H., 6, 801, 806

L

Lange, Oscar, 500, 909
Leith, C. K., 191, 193, 194
Leven, Maurice, 42, 883
Lundberg, Erik, 909

M

McAdam, John L., 32
Madison, James, 427
Malthus, Thomas Robert, 229-231, 773
Marshall, John, 129
Marshall, L. C., 755
Marx, Karl, 238, 876
Means, Gardner C., 576
Mill, John Stuart, 760
Millis, Harry A., 651, 655, 658
Montgomery, Royal E., 651, 655, 658
Moore, Henry L., 488
Morgan, J. Pierpont, 595
Morris Robert, 426
Morse, Samuel F. B., 32
Moulton, H. G., 42, 883

N

Nourse, E. G., 711

P

Pareto, Vilfredo, 883
Pearson, F. A., 703
Pogue, J. E., 195
Polo, Marco, 352, 369
Price, Willard, 354

Q

Quaintance, H. W., 239

R

Reuter, E. R., 202
Ricardo, David, 738, 750, 756, 773
Rockefeller, John D., 596
Roosevelt, Franklin Delano, 383, 772
Stephenson, George, 32

S

Seton, Ernest Thompson, 353
Shepherd, Goeffrey S., 890
Smith, Adam, 33, 831
Smith, George, 379
Smith, W. R., 607, 608
Snyder, Carl, 49
Stephson, George, 32

T

Thompson, C. W., 607, 608
Thompson, Warren S., 204, 210, 219, 220, 231
Thorp, Willard, 582, 591
Trevelyan, G. M., 23

V

Veblen, Thorstein, 17*n*.

W

Walker, Francis A., 220
Warburton, Clark, 42, 883
Warren, George F., 703
Washington. George, 426
Watkins, Myron W., 586
Watt, James, 32
Weissman, Rudolph L., 582, 583
Weld, L. D. H., 118
Welsh, Edward C., 577
Whelpton, P. K., 219, 220
White, E. D., 597
Wilson, Woodrow, 432, 537, 598
Working, Holbrook, 627
Woytinsky, Viadimir, 776
Wright, Chester W., 628, 687, 755

SUBJECT INDEX

A

Ability to pay as basis for taxation, 836
Acceptance of draft, 398
Ad valorem tariffs, 524
Advertising, and consumer ignorance, 120
  effect of, on consumers, 119
  and elasticity of demand, 119
  importance of, 118
AFL, (*see* American Federation of Labor)
Agents of production, 177
Agricultural Adjustment Act of 1933, 710
Agricultural Marketing Act of 1937, 712
Agriculture, a competitive business, 691
  economic problems of, 686-715
  farm legislation for, 709-715
    appraisal of, 712
  fixed costs in, 696
  labor supply of, 697
  net incomes in, 713
  parity concept in, 714
  peculiarities of, 690
  position in American economy, 686
  price peculiarities in, 695
  products of, elasticity of demand for, 692
    elasticity of supply of, 695
    export market for, 697
    foreign trade in, 697
    supply of, 693
  a small-scale business, 690
  tenancy in, 700
Aldrich Commission, 432
American Federation of Labor, 639
American Tobacco Company, 596
Amortization, definition of, 393
Amplitude of business cycles, 478
Anarchism, definition of, 877
Antitrust legislation, 594-601
Arbitration of industrial disputes, 662
Art compared with science and religion, 11
Assessment rate, relationship to tax rate, 839
Assets, definition of, 145
Average costs, 249
  relationship to marginal costs, 252
    (*See also* Costs)
Average cost curves, shape of, 251
Average output, definition of, 180
Average revenue, under competition, 341
  under monopoly conditions, 342
Average revenue curves, shape of, 340

B

Balance of payments, international, 557-560
Balance of trade, 507
  favorable, under mercantilism, 28
  unfavorable, 508
Balance sheet, 145
  of commercial bank, 402
Bank clearings, 407
Bank deposits, guarantee of, 452
Bank failures, in business cycle, 480
  and size of banks, 448
  and size of community, 449-450
Bank notes, 375, 430
  legal tender power of, 378
Bank of North America, 426
Bank reserves, 410
  decentralization of, 431
  minimum requirements for, in the United States, 412
Barter, 273
  disadvantages of, 273
Beneficiary, definition of, 573
"Big Five" English banks, 431
Bill of exchange, definition of, 395, 544
Bill of lading, definition of, 397
Bimetallism, 362
  commercial ratio under, 363, 366
  legal ratio under, 362
  market ratio under, 363

Bimetallism, mint ratio under, 362
Birth rate, 207
Black list, in hiring procedure, 651
Bland-Allison Act, 381
Board of Governors of the Federal Reserve System, 437
Bonds, 141
  kinds of, 142
Book value of stock, 151
Bourgeoisie, definition of, 876*n*.
Boycott, 659
  secondary, 659
Branch banking, 431, 450.
Budget, government, 825
  and property tax, 839
Business combination, 569
  efficiency of, 585
  horizontal, 574
  vertical, 574
Business cycle, 474-500
  and capitalism, 872
  causes of, 488
  characteristics of, summary of, 487
  and the Federal Reserve System, 442
  historical background of, 477
  long, 477
  remedies for, 490
  sequential nature of, 478
  stages of, definitions of, 483

C

Call loans, 393
Call rate of interest, 784
Capital, and concentration of economic power, 244
  definition of, 236
  and economic forecasting, 241
  productivity of, 238
  and saving, 242
  sources of, 237
Capital goods, 59
Capitalism, achievements of, 866
  and the business cycle, 872
  and competition, 864
  and democracy, 868
  faults of, 868
  and freedom of contract, 863
  and freedom of enterprise, 863
  as incentive to production, 866
Capitalism, and inequality, 868
  and inheritance, 862
  and insecurity, 872
  and level of living, 867
  and materialism, 869
  and private property, 861
  and the profit motive, 863
  rise of, 33
  and special interests, 871
  and war, 873
  wastes under, 869
  weaknesses of, 868
Capitalization, of a corporation, 144
  of income, 611
  of future income, 787
  for price-setting purposes, 613
  of rent, 753
  of taxes, 841
Central reserve cities, 436
Chain banking, 450
Charter, corporate, 131
Checks, advantages of, 417
  endorsement of, 418
CIO (*see* Congress of Industrial Organizations)
Class prices, 316
  (*See also* Discrimination)
Class struggle, 876
Clayton Act, 598
Clearing house, definition of, 408
Clearings, in the Federal Reserve System, 438
  in international exchange, 551
Closed shop, 645
Closed union, 647
Cobweb theorem, for production cycles, 495-500
  mathematical treatment of, 905
Coinage, 358
  free, 360
Collective bargaining, 644
Combinations, business (*see* Business combinations)
Commercial banks, 393
  balance sheet of, 402, 406
  functions of, 423
  as investors, 422
  operation of, 402
  profits by lending, 403
Commercial credit, 392
Commercial paper, 415
  description of, 377

Commercial ratio under bimetallism, 363
in the United States, 366
Commodity discrimination in shipping rates, 619
"Commodity dollar," definition of, 493
Common stock, 137
Communism (*see* Socialism and Communism)
Company unions, 638, 654
Comparative advantage, law of, 169
and the tariff, 533
Compensated dollar, definition of, 493
Compensatory action under bimetallism, 365
Competing goods, 96
Competition and capitalism, 864
and elasticity of demand, 281
nature of, 278
and price setting, 282
Complementary goods, 96
Compulsory arbitration, 662
Compulsory investigation, 663
Concentration (*see* Inequality of wealth and income; Monopoly)
Conciliation, 661
Confidence, and the banking business, 453
Congress of Industrial Organizations, 641
Consolidation, definition of, 574
Constant-outlay curve, description of, 889
Consumer credit, 392
expense of, 786
Consumer ignorance, 108
Consumers' boycott, 659
Consumers' surplus, 75
Consumption goods, 59
Contract and capitalism, 863
definition of, 864
Contraction, stage of business cycle, 483
Cooperation in production, how obtained, 860
Copyright monopoly, 297
Corporations, 129
advantages and disadvantages of, 133
claims against, 143
control of, 143
directors of, 136
importance of, 133, 580-582
Corporative state of fascism, 879
"Corrected" dollars, rule for, 471
Correspondent banks, 438
Cost curves, 249
average variable, 326
Cost curves, interrelationships of, 903
shape of, mathematical treatment of, 891
Costs, joint (*see* Joint costs)
and monopoly price, 300
not cause of price, 347
overhead, 335
and shares in distribution, 720
sources of, 248
sunk, 267
(*See also* Average costs, Fixed costs, Marginal costs, Variable costs)
Country banks, 436
Craft guilds, 26
Craft unions, 638
Credit, 275
bases of, 420
conversion of, 421
kinds of, 392, 393
Credit money, 372
"Crime of '73," 367
Crises, business, 478
Cross elasticity of demand, 96
for farm products, 692-3
formula for, 97
Cumulative preferred stock, 137
Currency, elastic, 376
inelastic, 376
managed, 386
Cyclical movements, 475
Cyclical unemployment, 674

D

Dartmouth College case, 129
Death rate, 210
Debentures, 142
Decentralization of bank reserves, 431
Deflation, 463
Delectus personae, 127
in the corporation, 130
in the joint stock company, 129
Demand, 69
changes in, how shown, 82
definition of, 81
elasticity of (*see* Elasticity of demand)
factors changing, 81
individual, 76
instantaneous character of, 84
law of, 77
dependence of, on diminishing utility, 77
Demand, market, 78
relationship to price change, 82

Demand curves, 79
linear, 889
Demand deposits, 401
Demand for factors of production, 720
derived, 721
Demand loans, 393
Demand schedules, 80
Democracy and capitalism, 868
Deposits, definition of, 401
guarantee of, 451
kinds of, 401
why preferred to cash, 417
Depression, stage in business cycles, 484-485
Differentiation of products, 286
Diminishing productivity, law of, 182
and population growth, 232
Diminishing returns, law of, 267
Diminishing utility, law of, 62
evidence for, 63
mathematical treatment of, 885
of money, 64
results of, 69
Direct appropriation stage in economic development, 17
Direct sending privilege of banks, 439*n*.
Direct strike, 655
Direct taxes, 846
Directorates, interlocking, 569
Directors corporate, 136
Discount, bank, 399
pure, 398
Discount rate, and reserve ratio, 416
Discriminating monopoly, 316, 318, 903
Discrimination in railroad rates, 619
by commodities, 619
by localities, 621
by persons, 622
Distribuendum, definition of, 718
(*See also* Social dividend)
Distribution, 718-805
and costs of production, 720
definition of, 718
and ethics, 722
and marginal productivity, 721
in a noncompetitive society, 729
personal and functional, 719
and productivity, 719
relationship to taxation, 855
Diurnal cycles, 475
Dividends, corporate, 154
stock, 156
Division of labor, chronological, 164
geographical, 167
Domestic system in economic development, 27
Drafts, 395, 397
Dumping on foreign markets, 521, 531
definition of, 335
relationship to fixed costs, 334
Duopoly, definition of, 295
Duopsony, definition of, 296

E

Earnings, compared with wage rates, 762
Economic goods, 58, 66
Economic man, 73
Economic organization, goal of, 858
Economics, definition of, 9, 858
goal of, 858
laws of, 13
and making of choices, 865
place of, 12
as a science, 14
Elastic currency, 376
Elastic demand, definition of, 90
Elasticity of demand, and advertising, 119
competition, 281
constant, 101
mathematical treatment of, 889
definition of, 89, 888
and expenditures, 90, 888
consumer, 90
mathematical treatment of, 888
factors affecting, 93
for farm products, 692, 699
formula for, 89
graphic computation of, 103
graphic representation of, 101
importance of, 95
when supply changes, 289
interpretation of, 92
for labor, 771
with linear demand functions, 901
and marginal revenue, 311, 896
mathematical treatment of, 887
measurement of, 88
difficulties in, 98
and monopoly, 294, 300
when output changes, 300
when prices change, 299
and price discrimination, 318, 903

Elasticity of demand, and public utilities, 606
stable, 101
and substitutes for commodities, 94
and tariffs, 523
unitary, 90
variations of, along the demand curve, 100, 315
Elasticity of income tax, 846
Elasticity of supply, 265
compared with elasticity of demand, 266
computation of, 265
for farm products, 695
formula for, 265
importance of, 289
with linear supply curves, 892
mathematical treatment of, 892
Eminent domain, right of, 606
Endorsement, 395, 418
Engel's laws, 53
Equation of exchange, 465
illustration of, 469
sizes of items in, 468
Equivalent interest rate (*see* Interest)
Ethics and distribution, 722
Excess reserves, 444
Exchange, 272-561
and evaluation, 276
in isolation, 272
nature of, 272
and private property, 272
and scarcity, 273
and specialization, 272
Exchange, equation of (*see* Equation of exchange)
Exclusive franchises, 606
Expansion, phase of business cycle, 483
Expectation of life, 212
Export point, gold, 550
Extensive margin, and rent, 740
External causes of business cycle, 488
Extractive industries, 161
temporary character of, 162

F

Factors of production, 176
demand for, 720
fixed and variable, 179
proportionality of, 178
supply of, 728
Fair, medieval, 25
Farm problem, 702-709
Fascism, 879
Favorable balance of trade, 508
Federal Deposit Insurance Corporation, 452
Federal Farm Board, 710
Federal Reserve Agent, 437
Federal Reserve bank notes, 376
Federal Reserve Banks, directors of, 436
reserves behind deposits, 438
Federal Reserve districts, 433
Federal Reserve notes, 376
Federal Reserve System, 432
Board of Governors of, 437
and the business cycle, 442
clearings by, 438
Federal Advisory Council of, 446
and foreign central banks, 446
functions of, 437
interdistrict settlement fund of, 440
member banks of, 434
Federal Trade Commission, 115, 598
Federation of Labor, 639
Fees, definition of, 829
Fiat money, 373
legal tender power of, 374
First Law of the Market, 286
Fiscal monopolies, 297
Fixed costs, 250
in agriculture, 696
in monopoly, 894
and profit, 260
temporary neglect of, 331, 333
Fixed expenses (*see* Fixed costs)
Fixed factors, 179
Food and Drug administration, 112
Forecast curve, in cobweb analysis, 497
Foreign exchange, 543-557
effect on gold balances, 547
regulation of, 556
Foreign trade (*see* International trade)
Forest resources, 198
Free goods, 58, 66
Free market, 277
Free silver movement, 382
Free trade, the case for, 533
Freedom under capitalism, 863, 867
"Full-bodied" money, 368
Functional distribution, 719
Future values and interest rate, 786

G

Gains by trade, 172, 174

General price level (*see* Price level)
General strike, 655
Genetic industries, definition of, 161
Gentleman's agreement, 570
Geography of prices, 623
Gilds, medieval, 26
Goal of economic organization, 858
Gold certificates, 371
Gold export point, 550
Gold import point, 551
Gold points, 550
Gold standard, advantages of, 385
  disadvantages of, 386
Gold Standard Act, 382
Goods, classified by use, 59
  definition of, 57
  kinds of, 58, 59
Government, growing cost of, 822
  relation of income and expenditure of, 811
  as a spending agency, 810
Government expenditures, 810-826
  contrasted with private expenditures, 824
  purposes of, federal, 812
    local, 815
    state, 815
  volume of, 816
Government income, 828-856
  sources of, 828
  volume of, 852
Grand list of taxing unit, 838
Greenbacks, 373, 429
Gresham's law, 362, 364
  experience in operation, 365
Guarantee of bank deposits, 451

H

Handicraft stage in economic development, 24
Historical "long run," definition of, 338-340
Holding company, 570
Horizontal combination, 574
Human resources (*see* Labor)

I

Illth, definition of, 58, 66
Immigration, economic effects of, 227
  regulation of, 220, 225
  United States experience with, 217
  and wages, 774
Immigration legislation in U. S., 222
Immigration quotas, 224
Import point, gold, 551
Incidence of taxes, 847
Income, distribution of, in the United States, 41
  division between labor and property, 50
  Engel's laws of, 53
  kinds of, 761
  money, 40
  national (*see* National income)
  psychic, 41
  real, 40
  sources of, 248
    (*See also* Inequality of wealth and income)
Income account of a corporation, 152
Income taxes, 843
  elasticity of, 846
  purpose of, 845
  in the United States, 844
Inconvertible paper monetary standard, 385
Independent treasury system, 428
Index numbers, 457
  base period for, 458
  data for, in United States, 461
  use of, in "correcting" prices, 471
  interpretation of, 459
  weighting of, 459
Indirect taxes, 846
Individual proprietorship, 124
  advantages and disadvantages of, 125, 132
Industrial production in the United States, index of, 481, 482
Industrial Revolution, 29
  causes of, 30
  inventions during, 30
Industrial unions, 638
Industries, kinds of, 161
  localization of, 168, 189, 909
Inelastic currency, 376
Inelastic demand, definition of, 90
Inequality of wealth and income, under capitalism, 868
  causes of, 45
  desirability of, 47
  existence of, 44
  and industrial disputes, 663
  methods of reducing, 48
  and saving, 794
  universality of, 48
Infant industries, and the tariff, 530
Inflation, 463

Inheritance and capitalism, 862
Injunctions, 651
Inputs, physical and economic, 248
Insecurity under capitalism, 872
Insolvency, 149
Installment credit, cost of, 786
Institution, definition of, 18
"Insurance" of bank deposits, 452
Integration of industry, 575
Intensive margin and rent, 745
Interest, 779-802
  on bank loans, 398
  cause of, 781
  definition of, 779
  form of, 780
  kinds of, 783
  rate of, 781
    call, 784
    determination of, 798
    effective, 787
    equivalent, mathematical treatment of, 916
    market, 783, 798
    natural, 798
Interlocking directorates, 569
Intermediate credit, 393
Intermediate goods, 59
Internal causes of the business cycle, 488
International trade, 505-540
  in agricultural products, 697
  contrasted with domestic trade, 510
  importance of, 511
  interferences with, 519
  invisible items of, 508
  leading countries in, 517
  price adjustments to, 560
  problems in measuring, 513
  purposes of, 505
  trends in, 514
  United States in, 518
  visible items of, 509
Interstate Commerce Commission, 631-632
Inventions during Industrial Revolution, 30
Invisible items of international trade, 508
Irredeemable paper monetary standard, 385

J

Joint costs, defined, 335
  example of, 336
  prices under conditions of, 337
Joint stock companies, 128
  advantages and disadvantages of, 133
Jurisdictional strikes, 656

K

Knights of Labor, 638

L

Labor, 202-233
  definition of, 202
  demand for, 767, 771
  nature of, 763
  quality of, 203
  quantity of, 206
  supply of, 764
    (*See also* Organized labor; Labor unions; Unions)
Labor unions, 638
  and marginal productivity of labor, 770
    (*See also* Unions)
Laborsaving machinery, opposition of unions to, 649
Laissez faire, policy of, 33
Land, 186-200
  differences in quality of, 739
  economic properties of, 188
  quantity of, 187
  supply of, 739
  utilization of, in the United States, 189, 190
Land values and rent, 753
Laws, economic, 13
    of comparative advantage, 169
    of demand, 77
    of diminishing productivity, 182
    of diminishing returns, 267
    of diminishing utility, 62, 63, 885
    Gresham's, 364
    Malthusian, 229
    of the market, 286
    Pareto's, 883
    of rent, 750
    Ricardian, 750
    of satiety, 66
    of supply, 264
    of supply and demand, 290
    of variety of human wants, 7
  kinds of, 12
  natural, 13
  statute, 13

Legal ratio under bimetallism, 362
Legal tender, definition of, 360
present situation of, in the United States, 361
Level of living, under capitalism, 867
definition of, 52
Liabilities, definition of, 145
Life tables, 214
Limited partnership, 132
Loanable funds, demand for, 789
supply of, 791
Local discrimination in railroad rates, 621
Localization of industry, factors determining, 168
and land values, 189
mathematical treatment of, 909
Lockout, definition of, 653
Long cycles in business, 477
Long run, definition of, 338-340
Long-term credit, 393
Lorenz chart of income distribution, 45, 795

M

Machinery, laborsaving, opposition of unions to, 649
Malthusian doctrine, 229
Managed currency, 386
Manor, in medieval England, 20
Marginal costs, definition of, 251
directions for computing, 252
equals marginal revenue for maximum profit, 898
mathematical treatment of, 891
relationship of, to average cost, 252
to selling price, 257, 891
Marginal land, 742
Marginal output (*see* Marginal product)
Marginal product, 181
value of, mathematical treatment of, 915
Marginal productivity, and distribution, 721
and wages, 768
Marginal revenue, definition of, 309
equality with marginal cost for maximum profit, 898
mathematical treatment of, 896
and maximum profit, 313, 898
relationship of, to competitive price, 309
to elasticity of demand, 311, 896
to marginal cost, 313
to monopoly price, 310
Marginal utility, definition of, 61
by finite-difference method, 886, 887
by infinitesimal-difference method, 886, 887
mathematical treatment of, 885
negative, 65
Marginal value product, mathematical treatment of, 915
Market, clearing the, 284
free, 277
law of, 286
Market areas, relationship of, to production cost, 912
to transportation cost, 909
Market demand, 78
Market rate of interest, 798
Market ratio in bimetallism, 363
Market value of corporate stock, 150
Materialism and capitalism, 869
Mediation in labor disputes, 662
Medium of exchange, definition of, 275
(*See also* Money)
Member banks of Federal Reserve system, 434
Mercantilism, 27
Merchant gilds, medieval, 26
Mergers, 574
Migration, 216
(*See also* Immigration)
Military necessity, and tariffs, 529
Mineral resources, of United States, 191
Minimum wage legislation, 683
Mint ratio in bimetallism, 362
Monetary Commission, 432
Money, 352-389
classification of, 368
commodities used as, 352
definition of, 275, 352
functions of, 387
paper, 369
private, 356, 357, 379
qualities of materials in, 353
representative, 371
standard, 359
United States system of, 384
value of, quantity theory of, 467
Money income, 40
Monopoly, 566-601
advantages of, 583
classes of, 297
costs in, fixed, 894
variable, 895
of demand, 296

Monopoly, efficiency of, 585
and elasticity of demand, 294, 318
forced on employers by unions, 667
growth of, in the United States, 577
incompatibility with capitalism, 866
indications of existence of, 575
mathematical treatment of, 894
natural, 297
nature of, 294, 566
and organized labor, 636, 667
periods of growth of, 582
prices under, discrimination in, 316, 318, 902
formula for, 894
power to fix, 567
stability of, 575-577
and public interest, 296
public regulation of, 593
and public utilities, 604
regulation of, by taxation, 306
restriction of output by, 346
and the tariff, 534
undesirable features of, 567
and use of productive factors, 568
Monopsony, definition of, 296

N

National bank notes, 375, 430
redemption fund for, 376
National banks, 394, 429
National income, 718
(*See also* Social dividend)
National Labor Relations Act, 664
National origins, basis for immigration quotas, 223
National Socialism, 879
Natural law, 13
Natural monopolies, 297
Natural rate of interest, 798
Natural reproduction rate, 215
Natural resources (*see* Land)
Needs and wants, 4
Negligence and workman's compensation, 672
Negotiable instruments, definition of, 396
Net worth of business, 149
Niggardliness of nature, theory of, 8
Noncompeting groups of workers, 766
No-rent land, 742
Normal price, 324
competitive, 327, 332
definition of, 325
Normal price, noncompetitive, 344
relationships of, summary of, 346
Normal unemployment, 674
Norris-LaGuardia Act, 658
Northern Securities Company, 595

O

Old-age pensions and insurance, 677
Oligopoly, definition of, 295
Oligopsony, definition of, 296
One-name paper, definition of, 419
Open-market Committee, 444
Open-market operations, 444
Open shop, 645
Organized labor, 636-668
aims of, 642
character of, 636
federations of, 639
historical background of, 637
and monopoly, 636, 667
and politics, 660
violence and sabotage by, 659
(*See also* Unions)
Outputs, graphic determination of, 257
physical vs. economic, 248
Overcapitalization, 151
Overdrafts, 406
Overhead costs, 335
of railroads, 617
Owners' equity, 148
status of, as a liability, 149

P

Panics, as stage of the business cycle, 478
Paper money, 369
Par of exchange, 543
Par value of corporate stock, 149
Pareto's law, mathematical treatment of, 883
Parity concept in agriculture, 714
Participating preferred stock, 138
Partners, silent, 128
sleeping, 128
Partnership, 126
advantages and disadvantages of, 132
limited, 127
Pastoral stage in economic development, 17
Patents, 297
Pawnbrokers, charges of, 786

Period of the cycle, 478
Personal discrimination in railroad rates, 622
Personal distribution, 719
Picketing, 657
Pittman Act, 383
Planned economy and the business cycle, 494
Planning in consumption, how obtained, 860
Political activities of organized labor, 660
Pool, definition of, 569
Population, checks to, 230, 773
  desirable amount of, 228
  and diminishing productivity, 238
  quality of, 203
  rate of increase of, 215
  of the United States, 205
Poverty, 52
Preferential shop, 646
Preferred stock, 137
  cumulative, 137
  participating, 138
Present values and interest rate, 786
Preventive checks to population growth, 773
Price, competitive, effect of changing supply on, 288
    how set, 284
Price curve in cobweb analysis, 497
Price discrimination, and elasticity of demand, 903
  in monopolistic situations, 902
  by railroads, 619, 623
Price geography, 623
Price lags, 462
Price level, 457
  changes in, causes of, 464
    effects of, 461
    history of, 461, 462
  movements of, and real income, 470
  and value of money, 460
Price, monopoly, 294
  affected by changes in cost, 303
  discriminating, 316, 902
    and elasticity of demand, 318, 903
  factors influencing, 298
  how set, 343
  illustrated, 301
  set where demand is elastic, 314
  and taxation policies, 306
  and rent of land, 751
Primary boycott, 659
Primary industries, 161
Private money, 356, 357, 379
Producer credit, 392
Product differentiation, 286, 566
Production, 124-268
  definition of, 159
  factors of, 176
  industrial (*see* Industrial production.)
Production credit, 392
Production goods, 59
Productivity, as a claim to ownership, 726
  and distribution, 719
Profits, 802-805
  affected by cost, fixed, 261
    variable, 260
  affected by selling price, 254, 262
  the aim of management, 247
  an arbitrary accounting concept, 804
  nature of, 802
    temporary, under competition, 329, 803
  situation of, under monopoly, 804
Profit-and-loss account, 153
Profit motive under capitalism, 863
Profit sharing in industry, 663
Progressive taxes, 833
Proletariat, 876
Promissory notes, 395, 396
Property, 18
  and capitalism, 861
  and exchange, 272
  justification of, 723-725
  payments to, 723
  productivity of, as claim to ownership, 726
  as source of income, 50
Property taxes, 838
Proportional taxes, 833
Proprietorship, 124
Prosperity as phase of business cycle, 483, 484
Protective tariff, 522
  the case against, 533
  the case for, 529
    (*See also* Tariffs)
Psychic income, 41
Public debt, 819
Public expenditures (*see* Government expenditures)
Public revenue (*see* Government income)
Public utilities, 604-634
  changes in the concept of, 607
  character of, 604
  demand for the products of, 606
  monopoly and, 604

Public utilities, rate making for, 610
regulation of, 608
Public welfare monopolies, 297
Purchase, 274
Putting-out system, definition of, 27
Pyramiding, of bank reserves, 431
of corporate control, 571

Q

Quantity theory of the value of money, 467
Quasi rent, 756

R

Railroad transportation, 616-630
and costs, 616, 617, 627
position among carriers, 629
rates for, discrimination in, 619
commodity, 619
local, 621
personal, 622
regulation of, 630
Random movements in business, 475
Rate of exchange, 545
in absence of gold standard, 554
factors affecting, 549
Rate making for public utilities, 610
Real income, 40
affected by price movements, 470
national, and foreign trade, 509
Recession, phase in business cycle, 483, 485
Reciprocal trade agreements, 529
Recognition of labor unions, 645
Recovery, phase in business cycle, 483, 486
Redemption fund for National Bank notes, 376
Rediscount, 415
under Federal Reserve System, 432
Rediscount rate, regulation of, 442
"Reform" movements, characteristics of, 874
Regressive taxes, 833
Relative prices, for index numbers, 458
Religion compared with art and science, 11
Rent, economic, 738-758
cash, 743
contract, 752
definition of, 738
explicit, 753
implicit, 753
and land values, 753
margins of, combined, 747
Rent, extensive, 740
intensive, 745
money, 743
and prices, 751
quasi, 756
Ricardian law of, 750
share, 742
of urban land, 754
Representative money, 371
Reserve cities, 436
Reserve ratio, 413
methods of controlling, 413, 443
Reserves, bank, (*see* Bank reserves.)
Restriction of output, by labor unions, 647
by monopoly, 346
Revenue, average, 341, 342
marginal, 309, 313, 896, 898
Revenue tariffs, 522
Revival, phase of business cycle, 483
Ricardian law of rent, 750
Risk, assumption of, by workers, 672
Robinson-Patman Act, 599
Roundabout process of production, 240
Rule of reason, 597
"Run" on a bank, 454

S

Sabotage in labor disputes, 659
Safety of the banking system, 446
Salaries (*see* Wages)
Satiety, law of, 66
Saving, and inequality, 794
as a source of capital, 242
uses of, 797
volume of, 792
Savings banks, 394
Scarcity, 7
causes of, 8
and exchange, 272
Science, compared with art and religion, 10
economics as a, 14
Scientific management and labor disputes, 664
Seasonal movements, 475
Seasonal unemployment, 674
Secondary boycott, 659
Secondary industries, 161
Secular trend, definition of, 474
Self-liquidating credit, 392
Services as production, 160
Settlor, definition of, 573

Share rent, 742
Sherman Anti-trust Act, 594
Sherman Silver Purchase Act, 381
  repeal of, 382
Shifting of taxes, 846-850
  and elasticity of demand, 847
  and generality of the tax, 849
Short cycles, 477
"Short run," definition of, 338-340
Short-term credit, 393
Silver, nationalization of, 383
  under the New Deal, 383
  in United States monetary system, 380
Silver certificates, 371
Single tax, 756
Sit-down strikes, 657
Size of business, growth in United States, 578-580
Small loan companies, charges of, 786
Smith, Adam, influence of, 33
Smyth v. Ames rule, 618-619
Social dividend, 718
  size of, 732
Social security, 670-683
Socialism and Communism, 874-877
  disadvantages of, 877
  and labor unions, 875
  and the wage system, 875
Soil Conservation Act, 711
Span of life, 213
Special assessments, definition of, 829
Specialization (*see* Division of labor)
Specific tariffs, 524
Stability of prices and monopoly, 575-577
Standard money (*see* Money, standard)
Standard of living, 52
Standard Oil Company, 596
State banks, 374, 428
  notes of, 430
Statute law, 13
Stay-in strikes, 657
Stock, corporate, 135
    market value of, 150
  common, 137
  preferred, 137
  values of, 149
Strikes, 655
Strikebreakers, use of, 653
Submarginal land, 744
Subsidiary coins, 368
Substitutes, and cross elasticity, 96
Substitutes, and elasticity of demand, 94
Sumptuary legislation, definition of, 109
  forms of, 112
  historical changes in, 110
  unpopularity of, 111
  in the United States, 111
Sumptuary monopolies, 297
Sunk costs, 267
Supply, 263
  factors affecting, 264
  individual, 264
  of labor, 764
  of land, 739
  law of, 264
  market, 264
  of productive factors, 728
    (*See also* Elasticity of supply)
Supply and demand, law of, 290
Supply curve, 264, 892
Supply schedule, 264
Surplus, corporate, 149
Sympathetic strike, 655
Syndicalism, definition of, 878

T

"Take-home pay," 763
Tariff Commission, 528
Tariff policy, desirable, 538
  effects of, 524
  and elasticity of demand, 523
  how made, 527
  history of, in the United States, 535
  kinds of, 522, 524
Tax base, 832
Tax rate, 839
  and rate of assessment, 839
Taxation, limits to, 855
  of monopolies, 306
  nonrevenue aspects of, 855
  relationship to distribution, 855
Taxes, and ability to pay, 836
  burden of, 842
  canons of, 831
  capitalization of, 841
  definition of, 829
  direct and indirect, 846
  incidence of, 847
  income, 843
  objectives of, 830
  property, 838

Taxes, proportional, progressive, and regressive, 833
on real estate, 842
shifting of, 846-850
single, 756
used by the Federal government, 850
used by local units of government, 852
used by state governments, 851
withholding, 845
Technological unemployment, 674
Teleology in economics, 859
Tenancy on farms, 700
Time deposits, 401
Time preference, 237, 782
effect on change in interest rate, 800
Time problems in economics, 338
Trade, gains by, 172, 174
Trade acceptance, 398
Trade agreements, reciprocal, 529
Trade union, 638
(*See also* Unions)
Transportation, 614
cost of, effect on markets, 909
and price geography, 623
railroad, cost of, 627
importance of, 629
Transportation Act of 1920, 633
Trend, secular (*see* Secular trend)
Triangular trade, 552
Trust companies, 394
Trustee, 573
Trusts, 572
Two-name paper, definition of, 419

U

Ultra vires acts, definition of, 131
Unassignable costs, 335
of railroads, 617
Undivided profits, 148
Unearned increment, 755
Unemployed person, definition of, 674
Unemployment, kinds of, 674
Unemployment compensation, 674
Unfavorable balance of trade, 508
Unions, closed, 647
kinds of, 638
opposition to laborsaving machinery by, 649
recognition of, 645
restriction of output by, 647
structure of, 638
(*See also* Organized labor)
Unitary elasticity of demand, 90
United States Bank, 426, 427
United States notes, 373
Util, definition of, 60
Utility, 57
definition of, 57
diminishing, law of, 62
kinds of, 159
marginal, 61
numerical illustration of, 60
subjective nature of, 57

V

Variable costs, 250
average, 326
in monopoly, 316, 895
and profit, 260
Variable expense (*see* Variable costs.)
Variable factors of production, 179
Variety of human wants, 6, 7
law of, 7
Velocity of turnover, figures for the United States, 468
and public confidence, 467
Vertical combinations in business, 574, 587-589
Violence and labor disputes, 659
Visible items of international trade, 509
Voluntary arbitration of labor disputes, 662

W

Wages, 760-777
importance of, among distributive shares, 775
and labor unions, 770
long-run level of, 772
of management, 805
and marginal productivity, 768
in a noncompetitive market, 770
rates of, 762
and salaries, 762
Wages-fund theory, 773
Wagner Act (*see* National Labor Relations Act)
Wants, 2
change of, through time, 5
conformity of, 3
differences in intensity of, 5
individuality of, 2
and needs, 4

Wants, source of, 4
  variety of, 6
Water resources, 195
Watered stock, 151, 152
"Wealth of Nations," influence of, 32
Webb-Pomerene Act, 600
Weights for index numbers, 459
Wheeler-Lea Act, 116
Wildcat banks, 429
Withholding taxes, 845
Workers' boycott, 659
Workmen's compensation, 672

Y

Yellow-dog contract, 652